GENOMES, EVOLUTION, AND CULTURE

Past, Present, and Future of Humankind

Rene J. Herrera | Ralph Garcia-Bertrand | Francisco M. Salzano

WILEY Blackwell

Genomes, Evolution and Culture

Past, Present and Future of Humankind

We dedicate this book to our families

(RJH)
Esther Martinez (Tata)
Diane
Giselle and Daniel

(RG-B)
Lydia Evans (Mom)
Dianne
Jacob, Zachary, and Daniel

(FMS)
Thereza
Felipe, Renato
Grandchildren and great-grandchild

As well as

To all those who contributed to construct humankind history, either as researchers or through the generous donation of time and biological material for this enterprise. With gratitude to the American Museum of Natural History, New York City, for the awe and inspiration provided over the years.

Genomes, Evolution, and Culture

Past, Present, and Future of Humankind

Rene J. Herrera
Department of Molecular Biology, Colorado College, USA

Ralph Garcia-Bertrand
Department of Molecular Biology, Colorado College, USA

Francisco M. Salzano
Emeritus, Department of Genetics, Universidade Federal do Rio Grande do Sul, Brazil

WILEY Blackwell

This edition first published 2016 © 2016 by John Wiley & Sons Ltd

Registered office: John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial offices: 9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK
111 River Street, Hoboken, NJ 07030-5774, USA

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

The right of the author to be identified as the author of this work has been asserted in accordance with the UK Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty: While the publisher and author(s) have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. It is sold on the understanding that the publisher is not engaged in rendering professional services and neither the publisher nor the author shall be liable for damages arising herefrom. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Names: Herrera, Rene J., author. | Garcia-Bertrand, Ralph, author. | Salzano, Francisco M., author.
Title: Genomes, evolution, and culture : past, present, and future of humankind / Rene J. Herrera, Ralph Garcia-Bertrand, Francisco M. Salzano.
Description: Chichester, West Sussex ; Hoboken, NJ : John Wiley & Sons, Inc., 2016. | Includes bibliographical references and index.
Identifiers: LCCN 2015048491 (print) | LCCN 2015049134 (ebook) | ISBN 9781118876404 (cloth) | ISBN 9781118876381 (pdf) | ISBN 9781118876398 (epub)
Subjects: | MESH: Genome, Human | Biological Evolution | Culture
Classification: LCC QH447 (print) | LCC QH447 (ebook) | NLM QU 460 | DDC 611/.0181663–dc 3
LC record available at http://lccn.loc.gov/2015048491

A catalogue record for this book is available from the British Library.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Set in 8.5/12pt, MeridienLTStd-Roman by Thomson Digital, Noida, India.

1 2016

Contents

Preface, ix

1 The history of human evolutionary genetics, 1

World views, 1

Science and philosophy, 1

The biology of mankind: anatomy and physiology in a historical context (up to the 16th century), 3

Beginnings of the present scientific model, 4

Biological evolution and genetic foundations: the brilliant quartet, 5

Nineteenth century: cytology, embryology, and reproduction, 9

Twentieth century, the century of genetics, 10

The synthetic theory of evolution, 11

Bacterial and molecular genetics, 11

Parallel developments: paleoanthropology, 12

Technical and methodological developments, 13

Conclusions, 14

Review questions and exercises, 14

References, 14

2 The human genome: structure, function, and variation, 16

Science, politics, and ethics, 16

Structural aspects, 17

Normal and abnormal phenotype distribution, 19

Function, 21

Sex chromosomes, 23

Paleogenomics, 25

Variability: mtDNA, 28

Nuclear variability, 28

Exomes and proteomes, 30

Selection or drift? History, 31

Selection or drift? Methods, 32

Selection or drift? Analyses, 34

Nervous system and culture, 37

Conclusions, 37

Review questions and exercises, 37

References, 38

3 Population structure, 42

DNA-based marker systems, 42

SNPs, STRs, and Indels as DNA markers, 45

Population genetic tools for analyzing population structure, 49

Forces affecting population dynamics, structure, and evolution, 55

Applications of population genetics, 60

From populations to races and species, 62

Review questions and exercises, 64

References, 65

4 Genetic variability, 68

On the nature of variability, 68

Mechanisms responsible for generating genetic variability, 69

Randomness of mutations, 71

Inheritance and environment, 75

Selection works on the phenotype, 76

The impact of selection, 77

Cultural expressions as markers of ancestry, 79

Congruency among marker systems, 80

Does junk DNA exist?, 80

How genetic diversity is studied?, 82

Epigenetic diversity, 84

Review questions and exercises, 85

References, 86

5 Gene and genomic dynamics, 87

Molecular evidence for punctuated equilibrium and gradualism, 88

Next-generation sequencing, 90

Genetic variation, 92

Variation, population structure, and effective population size, 94

Recombination and its effect on variation, 94

Linkage equilibrium and disequilibrium, 95

Forces leading to linkage disequilibrium, 98

Linkage disequilibrium and SNP haplotypes, 99

Linkage disequilibrium in humans, 100

Genome structural variations, 102

CNV classifications and formation mechanisms, 102

Methods used to detect CNVs, 103

CNVs associated with human phenotypes, 105

CNVs and evolution, 106

CNV in primates, 107

Chromosome rearrangements and selfish genetic elements, 108

Transposable elements, 109

Population dynamics of transposable elements, 110

Transposons in human evolution, 111

Selfish genetic elements in evolution, 113

Genome-wide association studies, 113

Concerns over the effective use of GWAS, 117

Conclusions, 117

Review questions and exercises, 118

References, 119

6 Human origins and early diasporas, 124

The *on switch* to humanity, 124

Early hominins, 128

Emerging themes and variations in hominin evolution, 130

The first hominin migrants, 134

The emergence of modern humans, 135

The Saharan pump, 135

Early migrations, 136

Neanderthals prevailed, 136

Review questions and exercises, 138

References, 139

7 Culture, 141

Concept, 141

Origin and development, 141

Factors that could condition cultural evolution, 142

Biology–culture interaction, 145

Language, 146

Domestication, 148

Art, 151

Free will, morality, and religion, 153

Conclusions, 155

Review questions and exercises, 155

References, 155

8 Health and disease, 158

Hopes and reality, 158

Concept of health and methods of study, 158

Darwinian medicine, 160

Parent–offspring conflict, 161

Pathogen history, 161

Evolution of infectious diseases, 162

DNA damage, mutagenesis, and teratogenesis, 163

What is better, more or less gene product?, 165

Genetic manipulation of animals to study health and disease, 166

Reproductive fitness and health, 167

Consanguinity, 168

Violence, 168

Cancer, 171

Degenerative diseases, 172

Ecogenetics, pharmacogenetics, and pharmacogenomics, 174
Detection of genetic diseases, 176
Genetic counseling, 176
Treatment, 178
Conclusions, 179
Review questions and exercises, 179
References, 180

9 Recent human evolution: an integrative approach, 182
Recent human evolution, 182
Out of Africa, 184
Back to Africa, 188
Beyond Arabia, 190
The Asian agricultural revolution and the Austronesian expansion, 194
Evidence from plants and animals, 198
Contacts between South America and Polynesia, 200
Review questions and exercises, 202
References, 203

10 Bioethics: consequences and implications of genetic technology on human evolution, 206
Social and biological evolution, 207
Overview of ethics and philosophical influences on Western ethics, 207
Evolution of ethics and morality, 212
The history and beginning of modern-day bioethics, 213
Reproductive technologies and the new eugenics: unnatural selection?, 216
Enhancement through IVF, PGD, and CRISPR, 219
Ethical issues associated with medical technology, 220
Gene therapy, 220
Stem cell therapy, 222
Biosimilars, 222
Genetic privacy, 222
Genetic testing, 223
DNA profiling, 224
Conclusions, 227
Review questions and exercises, 227
References, 228

11 Future of human evolution, 230
Gene and culture coevolution, 231
Life expectancy and population growth: past, present, and future, 231
Mutation rates and future evolution, 232
The evolution of new genes, 234
Climate change, 234
Diet, 235
Sex selection, 236
Artificial selection, 236
Transhumanism and artificial intelligence, 237
Conclusions, 238
Review questions and exercises, 239
References, 239

Appendix, 241
Index, 249

Preface

At the end of 2012, one of us (RJH), concerned about the lack of a book on human evolution that would include not only genetics and evolution, but also other areas of knowledge that could influence this process (cultural anthropology, linguistics, demography, and other disciplines of the so-called humanities), decided to write a piece of work that would contemplate all of these aspects. Contacts with RG-B and FMS resulted in the prompt acceptance of a specific collaboration.

We then prepared a specific proposal that was, after proper consideration, accepted by Wiley-Blackwell in December of the following year (2013). In the ensuing two and a half years, we worked in close contact by e-mail, on the manuscript, followed by a face-to-face meeting in Porto Alegre, Brazil, in March 2015. The result is now presented for the appreciation of the readers.

The book comprises 11 chapters, distributed as follows: Chapter 1: history; Chapter 2: basic structural aspects; Chapters 3–5: population structure, variability, and its dynamics; Chapter 6: early migrations; Chapter 7: culture; Chapter 8: health and disease; Chapter 9: recent human evolution; Chapter 10: bioethical aspects; and Chapter 11: the future. There was a determined attempt to provide a holistic approach to the subject. The readers will decide whether we have succeeded or not. It was a pleasure to write this book, and we hope that its contents will reflect this state of spirit.

It is a pleasure to acknowledge the help we had, in the preparation of some chapters, from Drs. Jason Somarelli (Duke University, Durham, NC), Robert Lowery (Indian River State College, Fort Pierce, FL), and Marion Hourdequin (Colorado College, Colorado Springs, CO). RG-B is grateful for the monetary support from Colorado College to travel to Brazil, and the support of his colleagues in the Molecular Biology Department at Colorado College. FMS is grateful to the inspiring environment of the Genetics Department, Biosciences Institute, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil, and especially to Dr. Maria Cátira Bortolini, a long-time colleague in the studies of human molecular evolution. Our research was financed by Colorado College, Colorado Springs, CO, a Howard Hughes Medical Institute Undergraduate Biological Sciences Program, a grant from the Freeman Foundation, in the United States, and by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), as well as Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) in Brazil.

Miami, FL — Rene J. Herrera
Colorado Springs, CO — Ralph Garcia-Bertrand
Porto Alegre, Brazil — Francisco M. Salzano

CHAPTER 1

The history of human evolutionary genetics

The conflation and confusion of functions, of aims and criteria, is the normal, original condition of mankind.

—*Ernest Gellner* [1]

SUMMARY

We start by reviewing humanity's world views, and the relationship between science and philosophy. Afterward, a selected overview of what happened in the biological sciences during the 16th to the 20th century is presented, with special emphasis, in the 20th century, on the synthetic theory of evolution, as well as on bacterial and molecular genetics, emphasizing technical and methodological developments. Molecular evolution is opening now new horizons for the understanding of human history, and details of our present knowledge will be given in the following chapters.

World views

We are a naturally curious species. Since we crossed the pre-human threshold, therefore, we developed theories about ourselves and the world in general. These theories can be classified into three world views as follows: (a) magical; (b) metaphysical; and (c) scientific.

The *magical* world view was established at the beginning of our history, from a prelogical mentality that would not distinguish between wishes and the external world. There was a belief that through prayer or the influence of supernatural gods one could influence the course of events, such as the occurrence of rain or success in hunting or gathering. Cause and effect could not be clearly distinguished, and daily life was characterized by inexplicable events, which could only be understood by creating a mythology as general as the natural world itself. Fire, whose manipulation can be regarded as one of our first technological applications, was identified as a divine entity. There was no need for a coherent relationship between facts on the basis of previous knowledge. The observations were influenced by beliefs.

Around the 7th century before Christ (BC), there was a substantial change in the history of humanity, with attempts to explain the world by a set of rational premises and not by revealed or empirical evidence. This separation of the knowledge of the individual and of the surrounding environment characterized the *metaphysical* view of the world.

The *scientific* view, on the other hand, is based on the application of the scientific method, which is defined by the basic tenet of the cause–effect relationship. From the perspective of the scientific world view, the detailed analysis of part of reality can lead to an explanation as how one event results from the other. This perspective is basically materialistic, with no need for supernatural explanations.

Science and philosophy

So far, so good, but how do we separate science from philosophy? Mayr [2] defined science as "a body of facts (knowledge) and the concepts that permit explaining these facts." Philosophy, on the other hand, is translated literally from Greek as "the love of wisdom," and a dictionary definition describes it as "the general science of beings, principles, and causes."

Mayr [2,3] was skeptical about the importance of philosophy for science, and in his 1982 book he

Genomes, Evolution, and Culture: Past, Present, and Future of Humankind, First Edition.
Rene J. Herrera, Ralph Garcia-Bertrand, and Francisco M. Salzano.
© 2016 John Wiley & Sons, Ltd. Published 2016 by John Wiley & Sons, Ltd.

expressed doubts whether philosophy would have made any contribution to science after 1800. In particular, he indicated three philosophical concepts that would not be applicable to biology. (a) *Essentialism (typology)*: the world would consist of a limited number of sharply delimited and unchanging essences. This concept is unable to explain the vast organic variation present in our planet. (b) *Determinism*: everything would be rigidly conditioned by the structures of things. This view ignores stochastic and random processes that can lead to unpredictable evolutionary events. (c) *Reductionism*: the explanation of a system could be made if the system had been reduced to its smallest components. The concept of *emergence* is strictly related to reductionism and is characterized by three properties: (a) a genuine novelty is produced; (b) the characteristics of this novelty are qualitatively unlike anything that existed before; and (c) this novelty was unpredictable. Despite dissenting views of several scholars, it is now clear that evolutionary emergence occurs and that it is an empirical phenomenon without any metaphysical foundation.

The relationship between science and philosophy, however, should be explored, since both try to explain life and the universe, and this was aptly done by Bitsakis [4]. His main propositions are summarized in Box 1.1. To completely understand them, it is necessary to recall that *epistemology* deals with methods and grounds of knowledge, considering their limits and validity; therefore, it investigates the formation, status, classification, and development of the sciences, internal and external factors influencing it, and so on. On the other hand, *ontology* is the branch of knowledge that investigates the nature, essential properties, and relations of being. Propositions have to be tested through our senses and scientific instruments; they are derivative phenomena reflecting objective entities and processes. The real world

Box 1.1 Basis for a realistic and evolutionary epistemology.

1. There is an objective world accessible through the senses and scientific instruments.
2. Sense data are derivative phenomena that reflect objective entities and processes.
3. Natural laws are not conventions. They are the transcription, in human language and mathematical formulas, of objective relations, processes, and entities. They are *a posteriori* propositions.
4. Scientific propositions are subject to empirical testing.
5. Observational and experimental data are decisive for scientific research. Many times, from observational data we arrive at a scientific hypothesis, and the knowledge of the essential structures and mechanisms may remove the elements of uncertainty present due to this empiricist view.
6. Empiricism is a simplistic epistemology. Science does not recognize the dichotomy between phenomenon and essence, and a phenomenon both manifests and conceals deep structures and relations.
7. Simplistic, statistical empiricism leads to agnosticism. Theories are tested intersubjectively, since an objective criterion is not possible, and scientific laws are relative, and can be changed as knowledge progresses.
8. Scientific truth is not absolute, but we can successively reach objective truths.
9. The question of the truth or falsity of a proposition cannot be exhaustively answered by the criteria of empiricism. In the same way, philosophical propositions are not formulated independently of a number of ideological, social, and political factors. However, they can be tested using scientific data.
10. Sciences emerge and develop as theoretical appropriations of the laws of the objective world. Formalist epistemologies have stressed the importance of the laws of the objective world, but they also have stressed the importance of either internal or external factors, being unable to understand their dialectic unity. Scientific revolutions do not mean formal negation of the older proposition, but a dialectic transcending of alternative visions of reality.
11. Science includes, but at the same time produces, ideology.
12. Dogmatic metaphysics is historically obsolete. This, however, does not mean that philosophy as a whole is also dead. There is no science without ontological and epistemological presuppositions.
13. Science is structured with concepts, while philosophy deals with object categories. However, there is need for a mediation between the two.
14. Philosophy does not produce specific knowledge, but produces knowledge at the ontological and epistemological levels.
15. Epistemology has its own object and methods, and their relationship with scientific knowledge should be explored.

Source: Reference 4.

can only be assessed through successive approximations to this reality.

The content of Box 1.1 can be summarized by the following six main points: (a) there exists an objective world that can be investigated through the senses and scientific instruments; (b) scientific propositions can be empirically tested; (c) in epistemological terms, a distinction should be made between the phenomena seen and their essence; (d) scientific truth is not absolute, but we can successively reach objective truths; (e) dialectics is important; scientific revolutions do not mean formal negation of the older proposition, but a dialectic transcending of alternative views of reality; and (f) neither science, nor philosophy is free of ideological influences.

The biology of mankind: anatomy and physiology in a historical context (up to the 16th century)

How did the knowledge of the anatomy and physiology of mankind evolve with time? Box 1.2 lists 20 key persons

Box 1.2 Selected list of main contributors to the anatomical and physiological knowledge of mankind, 500 BC to the 16th century.

Historical/geographical periods and names	Time	Contributions
Ancient Greece		
Alcmaeon of Croton	ca. 500 BC	Anatomy, brain, mouth, and ear dissections.
Empedocles	504–433 BC	Humoral theory of disease.
Hippocrates	460–375 BC	Considered the Father of Medicine. Treatment of individuals rather than diseases.
Aristotle	384–322 BC	Comparative anatomy and physiology.
Alexandria School		
Herophilus	325 BC	Anatomy of the nervous system, study of pulse and lung rhythms.
Erasistratus	ca. 280 BC	Anatomy, physiology of arteries and veins.
Rome		
Aulus Cornelius	1st century AD	Author of the most famous Latin compilation of medical works.
Dioscorides	54–68 AD	Identification and description of about 600 plants with medicinal value.
Galen	131–200 AD	*On Anatomical Preparations*, the standard medical text for about 1400 years.
Middle Ages		
Rhazes	852–925 AD	He is credited with 237 medical books, including *Continens Liber*, widely used.
Avicenna	980–1037 AD	His book *The Canon of Medicine* was used for five centuries in European universities.
Averroes (Ibn Rushd)	1126–1198	Physician, condemned by the church due to his materialistic and pantheist opinions.
Theodoric of Lucca	1205–1298	Wound treatment.
Renaissance		
Guy de Chauliac	1300–1370	Surgery of cancerous tissues and ulcers. Traction as a treatment for fractures.
Leonardo da Vinci	1452–1519	Detailed anatomical studies and artistic reproduction of the human body.
Michelangelo Buonarroti	1475–1564	Detailed anatomical studies and artistic reproduction of the human body.
Andreas Vesalius	1514–1564	The Father of Modern Anatomy.
Ambroise Paré	1517–1590	Ligature of blood vessels to stop bleeding, introduction of artificial limbs.
Gabriele Falloppio	1523–1562	*Observationes Anatomicae*, detailed description of the female reproductive organs.
William Harvey	1578–1627	*de Motu Cordis*, the first book to clearly explain blood circulation. Description of the developmental stages of human embryos.

Sources: References 5 and 6.

who contributed in a significant way to this knowledge. It is broken down into five chronological periods that are further classified on the basis of geography.

From the perspective of the occidental tradition, the intellectual history of humankind started around 600 BC in Greece. The four paradigmatic persons listed as living in ancient Greece were important not only for their contribution to anatomy and physiology in general, but also in relation specifically to applications for medical practice. Alcmaeon described the human optic nerves, established the distinction between arterial and venous blood, identified the trachea, and assigned the brain as the center of reasoning in humans. Empedocles is best known by his humoral theory of disease. In this theory, the four elements fire, water, earth, and air were associated with four body humors, blood, black bile, phlegm, and yellow bile. Empedocles proposed that good health would result if there were a balance among these humors, and different types of personality and health problems would occur if an imbalance existed. Although the theory was wrong, its emphasis on internal factors in the causation of diseases was laudable. He was also concerned with other biological questions, such as the origin of living beings. Hippocrates is considered the Father of Medicine, and he especially stressed the need to consider the individual patient rather than the disease, an underlying principle of the modern goal of individualized medicine. Finally, Aristotle had an important contribution to modern thinking by his empirical application to the problems considered, leading to what is today known as the scientific method. He applied this method to develop a formal classification of animals, separating, for instance, vertebrates from invertebrates.

After the conquest of Greece by Macedonia, the world cultural–scientific center was transferred first to Alexandria and then to Rome. The Alexandrian, Herophilus, has been reported as having dissected not less than 600 human bodies. He located the brain as the center of the nervous system and the seat of intelligence, and his studies of pulse and lung rhythms were important for the discovery of blood circulation. Erasistratus closely examined the passage of the blood through the veins and into the arteries and was also responsible for a series of studies related to the digestive system.

With the passage of the center of power to Rome, the emphasis of the studies turned to practical applications. Aulus Cornelius, Dioscorides, and Galen, however, should be remembered, especially Galen, whose book *On Anatomical Preparations* was used as a medical text for not less than 1400 years, probably being the longest textbook in print in history.

The Middle Ages can be characterized as a period of extreme religiosity, with an unfavorable climate for science and open inquiry. During this time, however, four persons who contributed significantly to our understanding of anatomy and medicine are listed in Box 1.2. Two of them (Rhazes and Avicenna) lived mostly in Persia, practicing medicine and contributing to the medical literature, while Averroes and Theodoric of Lucca lived, respectively, in Spain and Italy. Averroes, also a physician, developed materialistic visions of the world, and due to them was banished by the church, although a few years before his death the banishment was terminated. Theodoric of Lucca devised many procedures that were important in surgery.

Human history is characterized by several cycles of authoritarianism that, however, do not last forever. It seems that human nature considers freedom as an essential characteristic for an appropriate living. The Middle Age ecclesiastic repression, therefore, could not last forever, and it gave space to a splendid development of art, literature, and science, the Renaissance. Seven paradigmatic figures are listed in Box 1.2, and Leonardo da Vinci and Michelangelo Buonarroti are well-known personalities who do not need comments. Guy de Chauliac, Ambroise Paré, Gabriele Falloppio, and especially Andreas Vesalius excelled in different aspects of anatomy and surgery, while William Harvey, of course, was the first scholar to have a clear and accurate picture of blood circulation.

Beginnings of the present scientific model

Selected key persons responsible for the development of biology in the 17th and 18th centuries are presented in Box 1.3. Three of them (Malpighi, van Leeuwenhoek, and Hooke) were mainly microscopists, responsible for the improvement and use of this very important research tool in that period. Hooke was the first to describe and name a cell, while Malpighi gave a detailed description of the capillaries and van Leeuwenhoek of the human sperm. Regnier de Graaf, on the other hand, furnished a description of the ovulation process, indicating the role of follicles in the ovary.

Box 1.3 Key persons responsible for the development of biology in the 17th and 18th centuries.

Names	Time	Contributions
Marcello Malpighi	1628–1694	One of the first to use the microscope, he made a complete description of the capillaries.
Antonie van Leeuwenhoek	1632–1723	Extensive and detailed microscopic observations, including of human sperm.
Robert Hooke	1635–1703	Responsible for the improvement of microscopy, and was the first to use the word cell.
Regnier de Graaf	1641–1673	Description of the ovulation process, indicating the follicle's role in the ovary.
Pierre L.M. de Maupertuis	1698–1759	Investigations about inheritance, negation of creationism.
Georges-Louis Leclerc, Comte de Buffon	1707–1788	Contributed to the discussion about evolutionism and is considered the Father of Biogeography.
Thomas R. Malthus	1766–1834	His book *An Essay on the Principle of Population* greatly influenced the thinking of both Alfred R. Wallace and Charles Darwin.
Georges L. Cuvier	1769–1832	He made important contributions to comparative anatomy and paleontology.
Etienne Geoffroy Saint-Hilaire	1772–1844	Favorable to evolution. Morphologist, contributed in an important way to the homology principle.

Sources: References 3, 5, and 6.

The five other selected persons were mainly concerned with genetics and evolution. Maupertuis investigated the area of biological inheritance, and clearly opposed creationism, Buffon, Cuvier, and Saint-Hilaire were concerned with different aspects of organismal variability and its distribution, while Malthus, with his demographic studies, decisively influenced Alfred R. Wallace and Charles Darwin in their thinking about the evolutionary process, as detailed in the next section.

Biological evolution and genetic foundations: the brilliant quartet

The history in the fields of evolution of genetics is dominated in the 19th century by four paradigmatic persons, listed in the order of their births: Jean-Baptiste Lamarck, Charles Darwin, Gregor Mendel, and Alfred R. Wallace. Information about the life histories of each of these influential scientists is given in Boxes 1.4–1.7.

Mayr [3] considers Lamarck one of the most difficult persons to evaluate in the history of science due to the failure of his critics of separating his ideas on evolutionary changes *per se* and Lamarck's attempts to explain the physiological and genetic mechanisms responsible for them. Two common errors are that he postulated a direct induction of new characters by the environment, and attributed a nonmechanistic explanation to volition. Neither Lamarck's strongest critics (Darwin was one of them, classifying his main work as "trash") nor his most extreme followers (like the French, who delayed the generalized acceptance of Darwinism in their country for at least 75 years) were strictly correct.

Box 1.4 presents some of the main events of Lamarck's life. In marked contrast with Darwin, he never escaped poverty. He was unhappy in his four marriages and died blind, without due recognition for his merits. However, despite these adversities, he was able to contribute in a significant way to the science of his time.

Lamarck proposed that evolutionary change took place via two factors. The first would be an intrinsic property of the living being, which would make possible the acquisition of always higher perfection and complexity. The second would be the ability of the living being to react to special environmental conditions. The second proposal involved the principle of use and disuse; the continuous utilization of an organ would lead to its development, and the lack of use to its deterioration. These changes would then be transmitted to their descendants (inheritance of acquired characteristics).

Developments in the 20th century clearly indicated that this type of inheritance does not exist (although

Box 1.4 Selected aspects of the life of Jean-Baptiste Pierre Antoine de Monet, Chevalier de Lamarck.

Year	Event
1744	Birth at Picardy, north of France, the youngest in a sibship of 11.
1760	His father dies, leaving the family in poverty.
1761–1763	Enrollment in the French army, fighting in the Seven Years' War. Wounded, returns to Paris for treatment, but never totally recovers from a lymphatic tissue lesion.
1764–1787	Lives in Paris from a small pension and works part-time in a bank. In his free time starts to work in botany. Makes acquaintance with Antoine-Laurent de Jussieu and writes a book in four volumes about the French vegetation (1778). Becomes the tutor of Comte de Buffon's son. Travels to several European countries.
1786	Buffon indicates him as assistant in the Department of Botany, Paris Natural Museum.
1793	Becomes Professor of Invertebrate Zoology in the above-mentioned museum, turning his interest to extinct and extant mollusks.
1800	Presents his theory of evolution for the first time in his *Discours d'ouverture* to students.
1802	Proposes the term *biology* for the study of living organisms. It was also him who for the first time used the term *species* in its modern meaning.
1809	Publishes his most important book, *Philosophie Zoologique*, with a detailed description of his theory.
1815–1822	Publishes *Histoire Naturelle des Animaux sans Vertèbres*, in seven volumes.
1822–1829	In the last years of his life becomes blind and, although he had been married four times, he is only assisted by his two sisters. He died in 1829, poor and without due recognition for his merits.

Sources: References 3 and 5–7.

recent developments in epigenetics suggest a parental or an environmental influence, in some cases). Yet, despite the fact that Lamarck seemingly missed the mark with his theory, why is he still considered important for the history of science? First, because he was the first consistent evolutionist, discarding the hypothesis of a static world for that of a dynamic, ever-changing world. In addition, his emphasis on the importance of behavior, environment, and adaptation should be stressed. Other positive factors of his theory were the following: (a) his acceptance of only mechanistic factors for the phenomena considered; (b) his emphasis on the Earth's old age and in the gradual nature of evolution; and (c) his courage to include the human species in the evolutionary chain. He also contributed in a significant way to the knowledge of the French flora and the classification of invertebrates.

There is quite possibly no other scientist whose life and work has been as intricately examined and interpreted as that of Charles Darwin. This is due to not only the impact caused by his theory (it is said that the world would never be the same after the publication, in 1859, of *The Origin of Species*), but also the fact that he was an obsessive and incredibly methodical writer. His diaries informed everything that would happen to him, in both personal and professional terms, and his correspondence with people all over the world amounts to about 14 thousand letters.

Some of the main events related to Darwin's life are listed in Box 1.5. He had 73 years of life intensively dedicated to his family (he had not less than 10 children) and to science. Rich, he never needed an employment for his living. In the 6 years of his voyage around the world, he obtained a massive knowledge about the planet's geology, flora, and fauna. His dedication to science should also be stressed, in spite of the health problems that he had during a significant period of his life.

The theory of natural selection developed by Darwin had a long gestation. The first sketch was made between 1842 and 1844, but the book presenting it was published in 1859 only under the pressure that Alfred R. Wallace had independently arrived at the same idea.

One of the weaknesses of Darwin's theory, which he himself recognized, was the ignorance at the time of the laws that determined the biological inheritance in living organisms. Yet, the fundamentals of these laws would be clearly delineated in 1866, 7 years after the publication of *The Origin of Species*, by Gregor Mendel. Interestingly, Gregor Mendel had sent a copy of his remarkable article to Darwin, who either never read it or was unaware of its importance in framing Darwin's own hypotheses.

Box 1.5 Selected aspects of the life of Charles Robert Darwin.

Year	Event
1807	Birth at Shrewsbury, west of England.
1825–1831	Studies in Edinburgh (medicine, up to 1827) and Cambridge (theology).
1831–1836	Voyage around the world in the ship Beagle.
1836	Return to London and marriage with his cousin Emma Wedgwood, with whom he had 10 children. Only seven, however, survived to adulthood.
1839	Publication of the book *Journal of Researches into the Natural History and Geology of Countries Visited by H.M.S. Beagle*. Admitted to the Royal Society.
1842–1844	Change of residence, from London to Down, First sketch of the theory of natural selection.
1858	Receives a letter from Alfred R. Wallace in which he presents the independent elaboration of the theory of natural selection. Joint communication of the two to the Linnean Society on July 1. Voyage to the Wight Island and beginning of the elaboration of the book that would be his masterpiece.
1859	Publication, at the age of 50 years, of his masterpiece, *The Origin of Species*. The first edition, of 1500 copies, was all sold in 1 day. Five other editions, published between 1860 and 1887, have been produced under his supervision.
1868	Publication of *The Variation of Animals and Plants Under Domestication*, in which he presents his theory of pangenesis, completely wrong.
1871	Publication of *The Descent of Man*, in which he applies the concepts of natural selection and sexual selection to the evolution of the human species.
1872	Publication of *The Expression of the Emotions in Man and Animals*, in which he considers different aspects of the human and animal behavior.
1882	Dies and, in spite of the resistance of the church and of conservative persons, is buried in the Westminster Abbey, together with other distinguished members of the kingdom.

Sources: References 8–11.

Box 1.6 gives some selected aspects of Gregor Mendel's life. There are doubts as to whether he adopted the ecclesiastic career due to vocation or was pressed by poverty or health problems. In any case, throughout his life, Mendel demonstrated his remarkable proficiency to his chosen career, leading to his election as Abbey and the concomitant task of administering the Saint Thomas Monastery for 16 years, until his death. Throughout his life, however, he always showed a keen interest for science, performing experiments at the monastery's garden. His seminal work in peas, which established the basis for all genetic research, was published, as indicated above, in 1866, but it was largely ignored by the scientific world.

Why was Mendel's work ignored? Perhaps the scientific world at the time was not yet prepared to understand its real importance. Only after a series of discoveries and analyses that were performed in the following three decades would it become possible to relate his laws with concrete cytological and reproductive events. However, there is no doubt, also, that the fact that he lived far from the more important centers of biological research, and that he was not affiliated with any scientific institution may have also contributed to his lack of recognition.

The fourth paradigmatic person deserving special mention in the 19th century is Alfred R. Wallace. Selected information about him is given in Box 1.7. Similar to Darwin, he never had a position with a scientific institution; however, contrary to Darwin, Wallace was not as wealthy. Wallace made his living and paid the expenses of his travels by selling specimens, giving lectures, and writing books and popular articles. Unlike Darwin, who developed his theory about natural selection over many years of observation, Wallace arrived at the conclusion about the evolutionary importance of natural selection in a single flash of insight. This happened when he was confined to bed due to an attack of yellow fever, under Malthus' influence (cf. Box 1.3). By the evening of that day, he had prepared a rough outline of the idea, and sent a letter to Darwin 2 days later.

In addition to Wallace's contributions to evolutionary principles, Wallace was also well recognized as a leader in

Box 1.6 Selected aspects of the life of Gregor Mendel.

Year	Event
1822	Birth in Heinzendorf, Austrian–Hungarian Empire, now Hynčice, Czech Republic.
1839	His father has an accident in active service that unables him to work, leaving the family in financial difficulties.
1840	Finishes his basic studies and enters the Philosophical Institute at Olomouc University, to become a priest.
1843	Starts his apprenticeship at the Saint Thomas Monastery in Brünn (now Brno).
1844–1847	Theological and agricultural studies in Brno's Episcopal Seminar and Philosophical Institute, respectively. Ordainment.
1849	Adjunct Instructor at Znaim.
1851–1853	Studies at the University of Vienna.
1854	Substitute Teacher at Brno's Royal School.
1857	Beginning of the research on peas and beans.
1861	Associates with Brno's Society of Naturalists.
1862	Touristic trip to Paris and London.
1864	Finishes the research with peas.
1865	Presentation of his seminal work *Versuche über Pflanzen-Hybriden* in Brno's Society of Naturalists. This work established the basis for all genetics.
1866	Publication of the work in *Verhandlungen des naturforschenden Vereines in Brünn* (Vol. 4, pp. 3–47).
1868	Elected Abbey.
1870	Publication of the work on *Hieracium*.
1874	Questions the government about the taxes that the Monastery should pay.
1876	Becomes the Vice Director of Moravia's Loan Bank.
1881	Director of the same bank. First symptoms of Bright's disease.
1884	Dies due to uremia caused by the indicated disease.

Source: Reference 12.

Box 1.7 Selected aspects of the life of Alfred Russel Wallace.

Year	Event
1823	Birth in Llanbadoc, Monmouthshire, Wales.
1837	Apprenticeship in surveying, in partnership with his brother, William.
1844	Takes a job as School Master in Leicester.
1848	Together with Henry Walter Bates (1825–1892) he sailed from England to South America in a collecting trip. The following year a younger brother, Herbert, joined them, but he died 1 year later with yellow fever.
1852	Return to England, but his ship burned in the way and he lost most of the specimens and notes he had collected.
1854–1862	Expedition to the Malay Archipelago. Noting differences between the eastern and western regions, he devised a line between Borneo and Celebes, and between Bali and Lombok, now known as Wallace's Line.
1866	Marriage with Annie Mitten and definitive settlement in London.
1870	Publication of *Contributions to the Theory of Natural Selection*.
1876	Publication of *Geographical Distribution of Animals*, a landmark in biogeography, in which he divides the world into six regions, recognized up to the present.
1903	Publication of *Man's Place in the Universe*. In this and the 1870 book, he sets human evolution apart from natural selection and biology, developing a mystical approach.
1905	Publication of an autobiography, *My Life*.
1913	Dies in London.

Sources: References 5 and 6.

the field of biogeography, and some of the regions he identified as important in animal distribution are still recognized today.

Nineteenth century: cytology, embryology, and reproduction

Twenty-four of the persons who, besides Lamarck, Darwin, Mendel, and Wallace, contributed in a significant way for the development of cytology, embryology, reproduction, and related subjects in the 19th century, are listed in Box 1.8. Their contributions could be classified, somewhat artificially, as follows: (a) biochemistry: Nägeli, Miescher, and Altman; (b) cytology: Schleiden, Schwann, Virchow, Balbiani, Flemming, Strasburger, van Beneden, Wilson, and Boverí; (c) embryology/reproduction: Purkinje, von Baer, Kölliker, Hertwig, Roux, and Driesch; (d) inheritance: Galton and Weismann; and (e) evolution: Spencer, Huxley, Haeckel, and de Vries.

Mayr [3] identified the period from 1830 to 1860 as one of the most exciting in the history of biology, in large measure due to these researchers and the abovementioned evolutionists. He considered that much of this burst of knowledge was due to the increasing professionalization of science, the improvement of the microscope, and the rapid development of chemistry. However, even with these cultural and technological developments, the genius of some of the indicated persons, no doubt, was also of decisive importance.

Box 1.8 Key persons responsible for the development of biology in the 19th century.

Names	Time	Contributions
J.E. Purkinje	1787–1869	Discovery of the germinal vesicle in bird's eggs (1825). Introduction of the term protoplasm (1839).
Karl E. van Baer	1792–1876	First accurate description of the human egg (1827). His 1828 book *Entwicklungsgeschichte der Thiere* was the standard embryology text for many years.
Matthias J. Schleiden	1804–1881	Together with T. Schwann was responsible for the cellular theory, according to which all living organisms are composed of cells (1838–1839).
Theodor Schwann	1810–1882	Co-responsible, with M.J. Schleiden, for the cellular theory.
Albrecht Kölliker	1817–1905	Applied the cellular theory to embryology and histology. In 1841 demonstrated that spermatozoids were sexual cells originated in the testicles.
Carl von Nägeli	1817–1891	Developed a series of chemical tests in plants, but did not understand Mendel's work and in 1884 presented a completely wrong theory about biological inheritance.
Herbert Spencer	1820–1903	Philosopher, he created the "survival of the fittest" expression, and extended it to the social sciences.
Rudolf Virchow	1821–1902	Extended the cellular theory to pathology (1858). Three years before established the principle that new cells could only appear from preexisting ones.
Francis Galton	1822–1911	One of the founders of biometry and of the statistical study of variation. Application of the twin method for the investigation of human behavior (1875).
E.G. Balbiani	1825–1899	Described mitosis in one protozoan in 1861, and in 1881 the giant polytenic chromosomes of *Chironomus*.
T.H. Huxley	1825–1895	Important studies in comparative anatomy. In 1868 concluded that *Archaeopteryx* should be an intermediate between reptiles and birds. Had an important role in the defense of Darwinism.
Ernst Haeckel	1834–1919	In his 1866 book *Generelle Morphologie der Organismen* developed the concept that ontogeny recapitulates phylogeny. While not correct, this principle generated a series of important studies in comparative embryology. In the same book, he created the term *ecology*.
August Weismann	1834–1914	Important theoretician, he emphasized in 1883 the distinction between somatic and germ cells. In 1885, he postulated the continuity of the germplasm and, in 1887, the need for a periodic reduction of the chromosome number in sexual organisms.
Walther Flemming	1843–1915	Studied mitosis in detail, creating terms used up to the present (*chromatin*, *prophase*, *metaphase*, *anaphase*, *telophase*). In 1882 described amphibian's lampbrush chromosomes.

(*Continued*)

Names	Time	Contributions
Friedrich Miescher	1844–1895	In 1871 described a technique for the isolation of nuclei and one substance that he called nuclein, from which Richard Altman extracted ribonucleic acid, the genetic material.
Eduard Strasburger	1844–1912	Analyzed in detail cell division in plants, and in 1879 demonstrated that a nucleus could only be formed by another nucleus.
Edouard van Beneden	1845–1910	Described, in *Ascaris*, in 1883, chromosome reduction in meiosis and its reestablishment after fertilization.
Hugo de Vries	1848–1935	Extensive crossings in plants, rediscovery of Mendel's work. According to him, mutations would be the main factor in evolution.
Oscar Hertwig	1849–1922	Described the fertilization process in sea urchin in 1875. His book *Cell and Tissues*, published in 1893, was very well received at the time.
Wilhelm Roux	1850–1924	Pioneer in the area of experimental embryology, suggested in 1883 that the units of inheritance would be carried in the chromosomes.
Richard Altman	1852–1901	Isolation of the nucleic acid in 1889. Description of the mitochondria in 1890.
Edmund R. Wilson	1856–1939	Studies in cytology and embryology. His book *The Cell in Development and Inheritance*, published in 1896, was a landmark for the investigation in these areas.
Theodor Boverí	1862–1915	Described the mechanism of the formation of the mitotic spindle in 1888. At the beginning of the 20th century, together with W.S. Sutton, postulated that the chromosomes would be the carriers of the units of inheritance.
Hans Driesch	1867–1941	Studies in experimental embryology. Publication of *Analytische Theorie der organischen Entwicklung*, where he generalized his studies.

Sources: References 3, 5, 6, 13, and 14.

Twentieth century, the century of genetics

Due to the enormous impact that genetics had, not only in the scientific world, but also in the everyday life of common people, the 20th century can be deservedly considered the century of genetics. Carlson [14] divided the history of genetics in the 20th century in a more or less symmetrical way in two periods: (a) the period of classical genetics (1900–1953) and (b) the period of molecular genetics (1953–1999).

The century starts with the rediscovery of Mendel's law by a trinity of important individuals, the Dutch Hugo de Vries (1848–1935), the German Carl Correns (1864–1933), and the Austrian Erich von Tschermak (1871–1962). Of the three, the scientifically most important was undoubtedly Hugo de Vries. de Vries maintained that he had found Mendel's ratios before reading Mendel's paper (probably about 1896), but in his March 14, 1900 article in the German journal *Berichte der Deutschen Botanischen Gesellschaft,* in which he reports 3:1 ratios in 19 plant genera, he did acknowledge Mendel's prior contributions to this idea. Carl Correns submitted a paper on April 24, 1900, to the same journal that published de Vries' findings, clearly acknowledging Mendel's work in its title. His experiments were performed in peas and maize. Finally, Tschermak, 1 month and 9 days later (June 2, 1900) submitted also to the *Berichte* a paper describing his results in peas, also confirming Mendel's ratios.

The acceptance of Mendelism, however, was not without controversy, and intense debates raged between the so-called Mendelians, who were represented by William Bateson (1861–1926), and the biometrists, such as Karl Pearson (1857–1936) and Walter F.R. Weldon (1860–1906). Interestingly, the data that definitely established the Mendelian bases of heredity came from the other side of the ocean, largely deriving from the "fly room" of Columbia University in New York. Sturtevant [13] affirmed that with T.H. Morgan's et al. classical book in 1915, and C.B. Bridges' article of 1916, an important period in the history of genetics was closed. No more doubts existed about the chromosomal theory of inheritance, opening the room for a successful fusion between genetics and evolution, which is detailed in the next section.

The synthetic theory of evolution

The fusion between Darwinism and Mendelism occurred in two very fertile decades of the 20th century, between 1930 and 1950, through the synthetic theory of evolution. Box 1.9 lists the 11 key books that furnished the fundamentals of the theory. Fisher, Wright, and Haldane elaborated the mathematical statistics bases, and Dobzhansky's (Figure 1.1) book is considered by many to be the main work that provided the fusion between these bases and the empirical studies. Dobzhansky apparently used Darwin's book as a model (as suggested by the title of his book), but in contrast to Darwin, who needed 17 years between the formulation of his theory and the publication of the book to document that theory, Dobzhansky wrote his classic in just 4 months! The extension of the theory to zoology and systematics was done by Ford, Mayr, and Rensch; to paleontology by Simpson; to cytogenetics by White; and to botany by Stebbins. Naming of the theory as *synthetic* was done by Huxley, who included embryology in its framework and developed general principles.

Bacterial and molecular genetics

In the middle of the century, another parallel revolution would occur in another area. Attention was turned to bacteriology, and as a result, a series of brilliant experiments established the importance of bacteria and bacteriophages for genetic analyses. Results of Oswald T. Avery (1877–1955), Colin M. MacLeod (1909–1972), and Maclyn McCarty (1911–2005) in 1944, and of Alfred D. Hershey (1908–1997) in 1952 made it clear that deoxyribonucleic acid (DNA) was the genetic material underlying these principles.

Figure 1.1 Theodosius Dobzhansky doing field work with Brazilian and Chilean colleagues in 1956. From left to right: Antonio R. Cordeiro, Francisco M. Salzano (both Brazilians), Danko Brncic (Chilean), and Luiz Glock (Brazilian). (*Source*: F.M. Salzano, personal collection.)

The molecular genetics era started in 1953 with the elegant DNA model devised by James D. Watson (born in 1928) and Francis H.C. Crick (1916–2004), to which Rosalind E. Franklin (1921–1958) and Maurice H.F. Wilkins (1916–2004) significantly contributed. The structure had been discovered, but it was necessary to know how it functioned, and it was Crick again who conceived the need for an intermediate in the road from DNA to protein, messenger ribonucleic acid (mRNA),

Box 1.9 The 11 key books that furnished the fundamentals of the synthetic theory of evolution.

Names	Time	Title of the book	Year
Ronald Fisher	1890–1962	*The Genetical Theory of Natural Selection*	1930
Sewall Wright	1889–1988	*Evolution in Mendelian Populations*	1931
Edmund E. Ford	1901–1988	*Mendelism and Evolution*	1931
John B.S. Haldane	1892–1964	*The Causes of Evolution*	1932
Theodosius Dobzhansky	1900–1975	*Genetics and the Origin of Species*	1937
Julian S. Huxley	1887–1975	*Evolution: The Modern Synthesis*	1942
Ernst Mayr	1904–2005	*Systematics and the Origin of Species*	1942
George G. Simpson	1902–1984	*Tempo and Mode in Evolution*	1944
Michael J.D. White	1910–1983	*Animal Cytology and Evolution*	1945
Bernhard Rensch	1900–1990	*Neuere Probleme der Abstammungslehre*	1947
G. Ledyard Stebbins	1906–2000	*Variation and Evolution in Plants*	1950

Sources: References 3, 15, and 16.

and with three other colleagues [among them the also famous Sydney Brenner (born in 1927)] identified the nature of the genetic code.

Parallel developments: paleoanthropology

The field of paleoanthropology could only develop after it was realized that the biblical version of the creation narrative should not be interpreted at its face value, and that our species had an extreme antiquity. In the 19th century, for instance, the noted anatomist Georges Cuvier (see Box 1.3) asserted, wrongly, of course, that there were no human fossils.

This view started to change only in the late 1850s, when British geologists became convinced that the stone tools found in association with human remains indicated that our species had an antiquity that could be assigned to late geological periods.

The history of paleoanthropology can be conveniently traced to 1856, with the discovery of unique human remains in the Feldhofer Grotte located in the Neander Valley (Tal, in German). Hermann Schaaffhausen (1816–1893), an anatomist, introduced the "Neandertaler" to science, raising a controversy that remains to this day, whether what has been called *Homo neanderthalensis* is or is not an archaic hominin that is truly unique as compared to our own species.

Another landmark in the history of human origins was the book by Charles Darwin *The Descent of Man*, published in 1871. His views were essentially correct, as evaluated presently. He postulated that the human species originated in Africa and suggested an adaptive scenario that included a change from the trees to the open plains, where bipedalism was adopted as a means of locomotion. This adaptation onto land freed the humans hands for tool making, which subsequently stimulated the development of intelligence [17].

The next chapter in the development of our understanding of human origins can be dated to 1894, with the description by Eugène Dubois (1858–1940) of fossils that he named as *Pithecanthropus*, now classified as *Homo erectus*. His examination of the skullcap and a femur from these remains led him to postulate that *Pithecanthropus* was relatively small brained, but was capable of walking bipedally, placing him in a position that was intermediary between humans and apes.

While these discoveries and advances were critical, the main events of the 20th century began in 1924, when a cranium of what was previously thought to be a baboon fossil was discovered 10 km southwest of Taung in South Africa. Raymond A. Dart (1893–1988), an anatomist at the University of Witwatersrand in Johannesburg, verified that the cranium had an unprecedented blend of pongid and hominid traits. Convinced of the evolutionary significance of the material, he assigned the Taung specimen to a new genus and species, *Australopithecus africanus*. Robert Broom (1866–1951), a medical doctor and paleontologist, agreed with Dart's evaluation and became an energetic advocate of the new entity. He also found an entirely different, robust form of australopithecine, which he named *Paranthropus robustus*.

The discoveries of paradigmatic fossil remains continued through the examination of material recovered from 1959 to 1963, which were described by Louis S.B. Leakey, Phillip V. Tobias (Figure 1.2), and John R. Napier in 1964 as a new species, *Homo habilis*. Evidence on the teeth, cranial bones, endocranial casts, and hand and foot bones indicated its distinction from everything previously known, justifying the new taxonomic unit.

All the interpretation of the above-mentioned findings was subjected to intense controversy, and Tobias [18] compared the 20-year delay in the acceptance of *Homo habilis* to the 35-year delay of acceptance of *Australopithecus africanus* and proposed that both were premature discoveries in the sense that they could not be connected in simple logical steps to canonical or generally accepted knowledge. He suggested that the delay in the acceptance

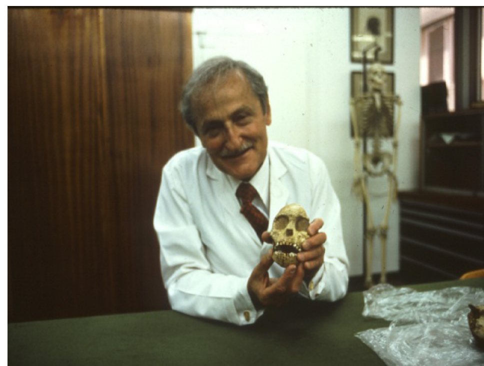


Figure 1.2 Phillip V. Tobias, a key figure in paleoanthropology and a dedicated fighter for civil rights in South Africa. (Reproduced with kind permission of Jeffrey McKee.)

of these two entities could only be understood as a sustained resistance to a change in preexisting concepts.

More recent developments led to a flurry of new taxonomic entities, but Cela-Conde and Ayala [19] and Wood [20] maintained the existence of only five pre-*sapiens* hominins: (a) the Miocene forms (exemplified in *Ardipithecus ramidus*, proposed in 1995); (b) archaic (examples: *Australopithecus afarensis* and *Australopithecus africanus*); (c) archaic megadontic (*Paranthropus robustus*); (d) transitional (*Homo habilis*); and (e) pre-modern *Homo* (*Homo erectus*).

Technical and methodological developments

Which factors lead to scientific development? This question has been frequently asked by historians of science, and the answers can be broadly classified into internal and external factors. Among the latter, we could mention (a) favorable political and socioeconomic conditions and (b) technological progress. Yet, the presence of paradigmatic persons in a certain place and at a given time is also undoubtedly important in relation to a large number of advances. Kuhn [21] differentiated what was classified as "normal" contributions to scientific revolutions, discoveries that opened entirely new horizons in a given subject, and for these events to happen three things are important: personal competence, inspiration, and the development of new analytical tools.

Population variability has been scientifically studied since the 18th century, with various methods and techniques that have steadily improved over time, especially from the 20th century onward. Box 1.10 lists the main laboratory and analytical tools that have been used both currently and in the recent past.

In the beginning, analyses of population variability had to rely only on morphological traits, which were classified using both simple, qualitative, visual inspections and quantitative, manual devices (anthropometric instruments). A revolution occurred in the study of these characteristics through the use of computerized morphometry, which through standardized photographs can infer differences in three dimensions.

At the beginning of the 20th century and subsequently thereafter, a different level of analysis came to the fore with the development of immunological methods, which relied on antigen–antibody reactions (either agglutination or lysis) and were extensively used for the investigation of the blood group and human leukocyte antigen (HLA) systems. Concomitantly, improved methods of cell culture became extremely useful for the maintenance and study of specific cell types.

Box 1.10 Main laboratory and analytic methods for the study of population variability.

1. Laboratory
    1.1. Morphological
        1.1.1. Qualitative visual inspection
        1.1.2. Quantitative manual devices
        1.1.3. Computerized morphometry
    1.2. Immunological
        1.2.1. Blood groups
        1.2.2. Histocompatibility leukocyte antigens (HLA)
    1.3. Cellular
        1.3.1. Cell culture
    1.4. Biochemical
        1.4.1. Chromatography
        1.4.2. Electrophoresis
    1.5. Molecular
        1.5.1. Restriction endonucleases
        1.5.2. Cloning
        1.5.3. Sequencing
        1.5.4. Polymerase chain reaction (PCR)
2. Analysis
    2.1. Calculating machines
    2.2. Bioinformatics
        2.2.1. Data banks
        2.2.2. Electronic programs
        2.2.3. Research networks

Biochemical techniques were also developed to study variability mainly in the middle of the 20th century, as exemplified by chromatography and electrophoresis. In particular, starch gel electrophoresis was extensively used for the investigation of intra- and interpopulation variability of human groups.

More recent developments led to the direct analysis of our genetic material (DNA). One truly remarkable and innovative development was the PCR technique, an easy-to-perform and cheap method, which led to its extensive use throughout the world. At present, sophisticated sequencing techniques allow the investigation of whole DNA regions containing thousands of nucleotide pairs.

All of these laboratory tests would not be useful if appropriate methods of data analysis were not available.

One of the earlier developments was the construction of calculating machines, which were useful in the first half of the 20th century. They were progressively replaced, however, by bioinformatic programs that could handle an enormous amount of information and have grown to process larger and larger data sets. The Internet has also revolutionized all aspects of scientific dissemination and data storage. The most recent developments were the establishment of data banks, which provide information to the entire research community, as well as the formation of consortia, made of large numbers of investigators, with the objective of solving problems that would be very difficult to tackle in other ways.

Conclusions

In this chapter, we have traveled a long way in time, considering first the magical views at the dawn of our evolutionary past and the relationships between science and philosophy. Afterward, we reviewed 2.5 thousand years of enquiry about our biological nature, covering important figures from ancient Greece to the present day. A large number of persons who could be envisaged as main contributors to the understanding of this subject have been named. The evolutionary process can be viewed as central for the knowledge of ourselves and all the biological world. Starting with the seminal contribution of Charles Darwin, successive approaches were developed, culminating with the molecular revolution, which is opening unforeseen perspectives for the elucidation of evolutionary mechanisms at the molecular, cellular, organismal, and population levels. The pace of knowledge acquisition is increasing steadily, opening entirely new areas of research. The result can only lead to a better understanding of our present condition, and a better basis for predicting our future.

Review questions and exercises

1. Do you agree with Gellner's sentence, quoted at the beginning of this chapter?

2. Why is it important to have a scientific view of the world? Are magical and mystical interpretations of the world still prevailing today?

3. Some argue that science is influenced by ideology. If this is the case, what measures could be taken to free it from this influence?

4. How would you classify the development of knowledge in human anatomy and physiology from 500 BC to the 16th century? Was the rate of development adequate?

5. Compare this rate with that which occurred during the 17th to the 19th century.

6. Which factors influenced the differential recognition of the works of Lamarck, Darwin, Mendel, and Wallace?

7. Of the 24 persons who significantly contributed to the development of biology in the 19th century, choose 5 who you consider were the most important, giving reasons for your choice.

8. Why is it considered that the 20th century was the century of genetics?

9. Explain why the currently accepted theory of evolution is named synthetic.

10. Relate the paleoanthropological and molecular approaches to human evolution, indicating both their strengths and weaknesses.

11. Compare the importance for the achievement of a given scientific result with the different stages performed to obtain it: field work, laboratory determinations, analysis, and publication. How would you rate the importance of each of them?

References

1 Gellner E. 1988. *Plough, Sword and Book: The Structure of Human History*. Collins Harvell, London.
2 Mayr E. 2004. *What Makes Biology Unique? Considerations on the Autonomy of a Scientific Discipline*. Cambridge University Press, Cambridge.
3 Mayr E. 1982. *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*. Harvard University Press, Cambridge, MA.
4 Bitsakis E. 1987–1988. For an evolutionary epistemology. *Sci Soc* **51**:389–413.

5 Gardner EJ. 1965. *History of Biology*, 2nd edition. Burgess, Minneapolis, MN.
6 Serafini A. 1993. *The Epic History of Biology*. Plenum Press, New York.
7 Gouyon P-H, Henry J-P, Arnould J. 2002. *Gene Avatars: The Neo-Darwinian Theory of Evolution*. Kluwer Academic/Plenum Press, New York.
8 Ruse M. 1987. *Taking Darwin Seriously: A Naturalistic Approach to Philosophy*. Blackwell, Oxford.
9 Desmond A, Moore J. 1991. *Darwin: The Life of a Tormented Evolutionist*. W.W. Norton & Company, New York.
10 Rose MR. 1998. *Darwin's Spectre: Evolutionary Biology in the Modern World*. Princeton University Press, Princeton, NJ.
11 Shanahan T. 2004. *The Evolution of Darwinism: Selection, Adaptation, and Progress in Evolutionary Biology*. Cambridge University Press, Cambridge.
12 Bowler PJ. 1989. *The Mendelian Revolution: The Emergence of Hereditarian Concepts in Modern Science and Society*. Johns Hopkins University, Baltimore, MD.
13 Sturtevant AH. 1965. *A History of Genetics*. Harper & Row, New York.
14 Carlson EA. 2004. *Mendel's Legacy: The Origin of Classical Genetics*. Cold Spring Harbor Press, Cold Spring Harbor, NY.
15 Provine WB. 1971. *The Origins of Theoretical Population Genetics*. University of Chicago Press, Chicago, IL.
16 Mayr E, Provine WB. 1980. *The Evolutionary Synthesis: Perspectives on the Unification of Biology*. Harvard University Press, Cambridge, MA.
17 Darwin C. 1871. *The Descent of Man*. John Murray, London.
18 Tobias PV. 1997. In: Spencer F (ed.), *History of Physical Anthropology: An Encyclopedia*. Garland Science, New York, pp 495–499.
19 Cela-Conde CJ, Ayala FJ. 2007. *Human Evolution: Trials from the Past*. Oxford University Press, Oxford.
20 Wood B. 2010. Reconstructing human evolution: achievements, challenges and opportunities. *Proc Natl Acad Sci USA* **107** (Suppl 2):8902–8909.
21 Kuhn T. 1962. *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, IL.

CHAPTER 2

The human genome: structure, function, and variation

Humans are much more than simply the product of a genome, but in a sense we are, both collectively and individually, defined within the genome.

—*Carina Dennis, Richard Gallagher, and Philip Campbell* [1]

SUMMARY

We are now in the era of genomics, and this chapter provides an overview of the human genome. It starts by considering the events that led to its first draft, and then structural aspects, the resulting normal and abnormal phenotypes, and details of the genotype–phenotype construction follow. Sex chromosomes are also exemplified as a fascinating example of evolutionary change. Paleogenomics is reviewed, as well as selected aspects of mitochondrial and autosome variability. Two main hypotheses about the factors influencing human genome variability, selection or genetic drift, are considered in some detail, and the chapter closes with a consideration of some specific aspects of our nervous system and culture.

Science, politics, and ethics

The history of the events that led to the publication of the first human genome draft in 2001 is vividly presented by Sulston and Ferry [2], and a list of the main events that took place before and after this historic date is given in Box 2.1. The idea, proposed by a few researchers in the early 1980s, was highly criticized by the scientific community, but after it was approved by the U.S. National Academy of Sciences (NAS) a whole institutional network involving the United States, France, the United Kingdom, and Canada was organized to implement it. A nongovernmental organization (Human Genome Organization, or HUGO) was created to coordinate research interactions, with the adhesion of Germany, China, and Japan.

The Human Genome Project was officially launched in 1990, with a target completion in 2005. A controversy then arose, as to whether the sequences obtained should or should not be patented. Unhappy with NIH's policy favoring patenting, James D. Watson, the Project's first Director, left the position. However, with the replacement of the NIH Director who had made the first patent requests, the institution withdrew all patent requests.

The work was developing according to plans when in 1998 one of NIH researchers, Craig Venter, announced that he would start a private company (Celera Genomics) to do the sequencing in 3 years. The news was received as a bomb at the annual Cold Spring Harbor meeting, organized for the discussion of the Project's progress during the year. Watson, irritated with Venter's attempt to own the human sequences, compared the event with Germany's invasion of Poland in 1939, and asked Francis Collins, the new HGP Director, whether he would act as Chamberlain or Churchill in the events that culminated with the Second World War.

Far from being intimidated by Venter's actions, the members of the Public Consortium accelerated the research and finally, in 2000, in a solemn ceremony, U.S. president Bill Clinton and UK's Prime Minister Tony Blair announced the end of the elaboration of the genome's first draft. According to Sulston and Ferry [2],

Genomes, Evolution, and Culture: Past, Present, and Future of Humankind, First Edition.
Rene J. Herrera, Ralph Garcia-Bertrand, and Francisco M. Salzano.
© 2016 John Wiley & Sons, Ltd. Published 2016 by John Wiley & Sons, Ltd.

Box 2.1 Some of the main events related to the history of the Human Genome Project (HGP).

Date	Event
1985	Workshop in Santa Cruz, USA, organized by Robert Sinsheimer, presents the idea.
1986	Charles DeLisi, from U.S. Department of Energy (DoE), organizes another planning workshop. These plans are discussed in the Cold Spring Harbor Symposium of that year.
1988	A National Academy of Sciences committee, presided by Bruce Alberts, approved the plans. The Program is launched as a joint initiative of DoE and the National Institutes of Health (NIH). James D. Watson is indicated to head the Program by NIH Director James Wyngaarden. Similar initiatives occur in France, the United Kingdom, and Japan. The Human Genome Organization is founded in Switzerland, with Walter F. Bodmer as president; Germany and China join the program.
1990	The Human Genome Project is officially launched, with a target of being finished in 2005. James D. Watson, disagreeing with the new NIH Director, Bernadine Healey, who decided to patent the DNA sequences, leaves the Project's direction.
1993	Francis Collins replaces Watson.
1994	Bernadine Healy is substituted in the NIH directorship by Harold Varmus, who takes back all the previous patenting applications.
1998	The news that Craig Venter would found a private company with plans of sequencing all the human genome in 3 years is received at the Cold Spring Harbor meeting. A race begins between the public and private institutions to see who would finish the study first.
2000	The U.S. president Bill Clinton and UK's Prime Minister Tony Blair solemnly announced to a collective press conference the end of the human genome first draft.
2001	*Nature*'s February 15 and *Science*'s February 16 editions, respectively, publish the Public Consortium and Celera Genomics human genome versions.
2002–2003	Discussion about the methods employed by the Public Consortium (clone-by-clone shotgun, or CCS) and by Celera (whole-genome shotgun, or WGS).
2006	End of the revision process of each of the human chromosome sequences, published in *Nature* between 1999 and 2006.

Sources: References 2–7.

the date of July 26 was chosen only because it was one of the few that were free in the two statesmen's agendas.

The Public Consortium draft genome was published in the February 15, 2001 *Nature* issue with Celera's following 1 day later in *Science*. However, discussion about the methods used by the two groups (CCS and WGS) continued in the two ensuing years. This included Waterston et al. [3] accusing the private group of utilizing the data openly available from the Public Consortium to generate their data. Current human genomic sequencing efforts use a mixed strategy, with the use of WGS first, and subsequent refining by CCS.

The original human genome sequence is being revised continuously, and a methodical revision process, chromosome by chromosome, was finished in 2006.

In the enthusiasm of the results, presentations by both scientists and politicians, as well as the media in general, exaggerated the significance of the findings. President Clinton asserted that "today we are learning the language with which God created life," and Collins said that "today we celebrate the revelation of human life's first draft." These metaphors may do more harm than good to science. As stressed by Weigmann [8], the genome carries information developed through the evolutionary process that is physiologically translated in the cells. To call the genome "life's book" implies that it was written with a purpose, to be read by humans. This confusion erroneously interprets the scientist's role.

Structural aspects

Table 2.1 provides a general view of the characteristics of our 22 autosome and sexual chromosome pairs. The total genome size was estimated as comprising 2 billion and 800 million base pairs (2.8 gigapairs or Gb). Among the

Table 2.1 Selected characteristics of the DNA of our 23 pairs of chromosomes.

Chromosome (group and number)		Genome size (Mb)	% of total	% G + C	% repeats	Number of genes	Number of pseudogenes	Gene density (Mb^{-1})
A	1	222.8	7.8	41.0	48.0	3131	991	14.2
	2	237.5	8.3	40.2	NA	1346	1239	7.0
	3	194.6	6.8	NA	NA	1585	122	8.8
B	4	187.2	6.6	38.2	NA	796	778	5.4
	5	177.7	6.2	39.5	46.3	923	577	NA
	6	167.3	5.9	40.0	43.9	1557	633	9.2
	7	154.8	5.4	41.0	45.0	1150	941	7.5
	8	142.6	5.0	39.2	44.5	793	301	5.6
C	9	117.8	4.1	41.4	46.1	1149	426	10.5
	10	131.6	4.6	41.6	43.7	1357	430	10.4
	11	131.1	4.6	41.6	48.0	1524	765	11.6
	12	130.3	4.6	NA	NA	1435	93	11.0
D	13	95.6	3.4	38.5	42.3	633	296	6.5
	14	88.3	3.1	40.9	46.2	1050	393	10.0
	15	81.3	2.8	NA	NA	695	250	NA
E	16	78.9	2.8	44.7	47.8	880	341	11.2
	17	77.8	2.7	45.5	45.3	1266	274	16.2
	18	74.7	2.6	39.8	43.5	337	171	4.4
F	19	55.8	2.0	48.0	55.0	1461	321	26.0
	20	59.5	2.1	44.1	42.0	727	168	12.2
G	21	34.2	1.2	40.8	40.1	225	59	6.7
	22	34.8	1.2	47.8	41.9	545	134	16.3
X		150.4	5.3	39.0	56.0	1098	700	7.1
Y		24.9	0.9	NA	NA	78	NA	NA
Total or average		2851.5	100.0	41.6	45.9	1073	452	10.4

Adapted from Dhand 2006 [7]. The numbers given have more recently been updated, but the basic relationships remain. NA: not available, at least in the referred publication.

autosomes, the largest chromosome is no. 2, with 237.5 megabases (Mb), 8.3% of the total, and the smallest is no. 21 (34.2 Mb; 1.2% of total). The largest chromosome is therefore almost seven times (6.9×) the size of the smallest chromosome. As for the sexual pair, chromosome X, traditionally classified within group C, presents an intermediate size within the group (150.4 Mb; 5.3% of total), being much larger (14×) than the Y chromosome, which is the smallest of all chromosomes (24.9 Mb; 0.9% of total).

The average G + C percentage in our species (41.6%) is not very different from those of plants and other animals, but there is a certain variation in relation to chromosome G–C content, from 38% (chromosome 4) to 48% (chromosome 19); 14 of the chromosomes have values between 40 and 48%. Repeat frequencies are similar (from 40%, chromosome 21, to 56%, chromosome X; 16 of the chromosomes show values between 40 and 49%).

Chromosome 1 has the largest number of genes (3131) and the highest gene density (14.2/Mb). At the other extreme, chromosome 21 shows only 225 genes and a gene density of 6.7/Mb, one of the lowest of the genome. Distribution of the number of genes does not follow the classical group or chromosome number classifications, based on gross phenotypic sizes, in a strict way. The relationship between numbers of genes and pseudogenes is highly variable. Chromosome 4 presents the highest number of pseudogenes compared with genes (778 versus 796; 98%), while this percentage is only 6%

(93 versus 1435) in chromosome 12. No relationship between these variables and gene density could be found.

Normal and abnormal phenotype distribution

The genome structure described above contains sections responsible for our morphology and condition, healthy or abnormal. Box 2.2 presents a selected list of normal and pathological conditions determined by the 22 autosomes and 2 sex chromosomes. Among the normal traits listed, we can find (a) 26 enzymes, (b) 11 immune system proteins, (c) 10 blood groups, (d) 2 coagulation factors, and (e) 32 other proteins, in a total of 81 substances or groups of substances.

A total of 77 pathological conditions or group of conditions are also listed in Box 2.2. They are

Box 2.2 Location of selected normal and abnormal conditions in the human genome.

Chromosome (group and number)		Normal characteristics	Pathological conditions
A	1	Duffy blood group (related to the malaria caused by *Plasmodium vivax*)	Gaucher disease Parkinson disease Several types of Charcot–Marie–Tooth disease
	2	Alkaline phosphatase Protein C Immunoglobulin kappa light chain Zinc finger domain	Iris coloboma Aniridia 1 Ehlers–Danlos syndrome
	3	Transferrin Ceruloplasmin Somatostatin	Spinocerebellar ataxia 7 Myotonic dystrophy 2 Xeroderma pigmentosum complementation group C von Hippel–Lindau disease
B	4	Phosphoglucomutase 2 Albumin Group-specific component (protein that links to Vitamin D) MN blood group	Huntington disease Wolf–Hirschhorn syndrome One form of muscular dystrophy
	5	Interleukins 3, 4, 5, and 13 Coagulation factor XII (Hageman factor)	Spinocerebellar atrophy Treacher Collins mandibulofacial dysostosis
C	6	Major histocompatibility complex Lp(a) apolipoprotein P blood group	SCA1 spinocerebellar ataxia Hemochromatosis Congenital adrenal hyperplasia (21-hydroxylase deficiency)
	7	Interferon beta-2 Erythropoietin Paraoxonase H1, H2A, and H2B histones	Cystic fibrosis Osteogenesis imperfecta Ehlers–Danlos syndrome type VIIA2 Williams–Beuren syndrome
	8	Carbonic anhydrases I, II, III Thyroglobulin Fibronectin Defensin	Microcephaly Plasminogen activator deficiency Epidermolysis bullosa
	9	Interferon alpha, beta Aldehyde dehydrogenase Component complement 5 ABO blood group	Philadelphia chromosome Chorea-acanthocytosis Hereditary myopathy with inclusion bodies
	10	Fibroblast growth factor 2 Dehydrogenase glutamate Oxaloacetate glutamate Soluble transaminase	Phosphatase and tensin homolog (PTEN, mutated in cancer) Hemolytic anemia due to hexokinase deficiency Metachromatic leukodystrophy

(*Continued*)

Chromosome (group and number)		Normal characteristics	Pathological conditions
	11	Hemoglobin complex (beta, delta, gamma) Apolipoprotein complex (A-I, C-III, A-IV)	Beckwith–Wiedemann syndrome Intermittent acute porphyria Congenital glaucoma
	12	Glyceraldehyde-3-phosphate dehydrogenase CD4 (AIDS virus receptor) Peptidase B	von Willebrand disease Hemolytic anemia due to triosephosphate isomerase deficiency Phenylketonuria
D	13	Coagulation factors VII and X Immunoglobulin E Collagen IV (alpha 1 and 2 chains)	Retinoblastoma 1 Dubin–Johnson syndrome Wilson disease
	14	Alpha/delta T-cell receptor Heavy chain immunoglobulin complex Beta-spectrin	Niemann–Pick disease Usher syndrome Nucleoside phosphorylase deficiency Spherocytosis
	15	Beta-2 microglobulin receptor Mitochondrial isocitrate dehydrogenase Mannose phosphate isomerase	Prader–Willi and Angelman syndromes Tay–Sachs disease Isovaleric acidemia
E	16	Alpha-globin gene complex Glyoxalase II Pseudocholinesterase 2 Haptoglobin	Kidney polycystic disease Tyrosinemia type II Urolithiasis
	17	Myosin heavy chain complex Galactokinase Thymidine kinase-1 Acid alpha-glucosidase	Miller–Dieker lissencephaly NF1 neurofibromatosis Smith–Magenis syndrome Charcot–Marie–Tooth disease type 1A
	18	Thymidylate synthase Myelin basic protein Peptidase A	Edwards syndrome Colon and rectum carcinoma Methemoglobinemia Erythropoietic protoporphyria
F	19	KIR (killer cell immunoglobulin receptors) Zinc finger transcription protein factor Lewis blood group Lutheran blood group H substance secretor	Familial hypercholesterolemia Hirschsprung disease Mannosidase Hemolytic anemia due to glucose phosphate isomerase
	20	Prion protein Adenosine deaminase Inosine triphosphatase A	Creutzfeldt–Jakob disease Severe combined immunodeficiency Alagille syndrome
	21	Amyloid beta A4 precursor protein Cystathionine beta-synthase Superoxide dismutase 1, soluble	Down syndrome Amyotrophic lateral sclerosis Homocystinuria
	22	Lambda immunoglobulin light chain complex Myoglobin P blood group	Hurler syndrome Cat eye syndrome Velocardiofacial/DiGeorge syndrome
X		Xg blood group Color vision pigments Alpha-galactosidase A	Hemophilia Duchenne muscular dystrophy Lesch–Nyhan syndrome Glucose-6-phosphate dehydrogenase deficiency
Y		Testis-specific protein, Y-encoded	DAZ, deleted in azoospermia

Sources: References 7 and 9.

(a) 17 syndromes, (b) 15 nervous system and/or muscular diseases, (c) 13 inborn errors of metabolism, (d) 9 hematological problems, (e) 4 conditions related to repair errors in DNA and cancer, (f) 4 conditions related to vision, and (g) 15 other diverse pathologies. Each chromosome represents a segregation unit that should be appropriately considered for a global evaluation of our genome.

Gregersen [10], reviewing two articles published in the same issue of *Science*, indicated how a genomic road map for complex human diseases could be constructed. The papers discussed indicated how networks of genetic regulation could be built for genes of the human innate system. As stressed, cell type and disease-associated environmental conditions are critical for the appropriate understanding of such relationships.

Function

What is the road between the discovery of the structures found through the genomic drafts and the subsequent efforts to find their connections to normal and disease phenotypes? The human genome is an elegant, but cryptic source of information, and the Encyclopedia of DNA Elements (ENCODE) Project was organized to clarify the process that exists between DNA and protein, mapping the biochemically active regions.

At the outset, however, we have to define what constitutes biological function, and what sets the boundaries between functional and nonfunctional elements. Three approaches can be envisaged, which are the genetic, evolutionary, and biochemical approaches (outlined in Box 2.3). All three present specific advantages and disadvantages. Thus, the genetic method has the advantage of providing experimental models of investigation, but it may miss elements responsible for phenotypes that occur only rarely in certain cells or environments. Loss-of-function tests are complicated by functional redundancy, leading to situations in which deletions of large segments have no discernible organismal effect.

The evolutionary approach is powerful, using comparative genomics as a tool. It has the advantage of not requiring prior knowledge of what a DNA region does, or when it is used. Major limitations include the following: (a) the need for accurate multispecies sequence alignments; (b) the fact that transcription factor binding sequences are short and highly degenerate; and (c) it requires sufficient phylogenetic distance.

The biochemical approach, employed by the Encyclopedia of DNA Elements, reveals the processes involved at a given genomic site in a determined cell type and activity state. However, biochemical signatures are a consequence, rather than a cause of function. Interactions between elements in the DNA regions and histone

Box 2.3 The three approaches employed to identify functional elements in the human genome.

1. Genetic

Relies on sequence changes to establish the relevance of a given DNA segment.

1.1. Mutations, either natural or artificially induced, to ascertain their phenotypic effects.

1.2. Transfection, using reporter assays in cell lines or embryos.

1.3. Loss-of-function tests in animal models.

2. Evolutionary

Through comparative genomics, this method investigates genomic regions that are conserved along the evolutionary process, with the simple assumption that functionally important regions should be maintained, but also looking for accelerated evolution across species or within a given lineage for regions that can lead to reproductive or survival advantages. This approach was successful in the recognition of protein-coding regions, structural RNAs, gene regulatory regions, regulatory motifs, and specific regulatory elements.

3. Biochemical

It is specific for cell type, condition, and molecular process. This approach defines major classes of functional noncoding elements, such as promoters, enhancers, silencers, insulators, and noncoding RNA genes such as microRNAs, structural RNAs, and regulatory RNAs. These factors are associated with distinctive chromatin structures that display signature patterns of histone modifications, DNA methylation, DNase accessibility, and transcription factor occupancy.

Source: Reference 11.

modifications may lead to difficulties in clear identification of the functional units.

What fraction of the human genome is functional? This question has been hotly debated. The ENCODE Project maintains the existence of not less than 4 million gene switches or myriad elements, responsible for the biochemical functions of 80% of our genome. They would include the many intergenic loci of unknown, but suspected controlling influence upon genetic expression. This view was strongly criticized by Doolittle [12], who argued that this number was not in agreement with the so-called *C*-value paradox. The *C*-value is a measure of the haploid nuclear DNA content, and the paradox exists because this value correlates poorly with organismal complexity. Thus, humans have 1000 times more DNA compared with simple bacteria, but lungfishes have at least 30 times more DNA than humans. Doolittle examined the ENCODE definition of a functional element, FE (a discrete genome segment that encodes a defined product or displays a reproducible biochemical signature) and presented the following two alternatives: (a) FEs would reflect organismal complexity regardless of *C*-values or (b) FEs would increase in number with the *C*-value. In the first case, this would imply huge differences in functional density, and very few organisms that could match our extraordinary cognitive capacities. If the second alternative is chosen, it would mean that lungfishes would have 300 times more functional units than ourselves, which certainly is not true. He suggested that function may be regarded as a diffusible quality. The problem, as outlined, remains open to further inquiry.

Methylation (addition of a methyl group) on cytosine–phosphate–guanine (CpG) sites has been identified with inactive or functional states. Methylation is of crucial importance for cell differentiation in both normal and abnormal conditions. Ziller et al. [13] charted this process along the whole human genome (42 whole-genome sequence data sets, 30 cell and tissue types). Most cell types, except germ cells and pre-implantation embryos, display relatively stable DNA methylation patterns, with 70–80% of all CpGs being methylated. They observed, however, dynamic regulation of the process in 22% of autosomal CpGs within a normal developmental context, most of which are distal to transcription sites, colocalizing also with enhancer regions. In addition, these dynamic regions are more susceptible to point mutations that are functionally consequential, and may lead to diseases. An extreme case is colon cancer, which showed global hypomethylation (10–15% less methylation).

It is probable that transcription factor-mediated regulation is the main determinant of the spatial–temporal gene expression, but the precise expression of a given gene is further regulated or fine-tuned by post-transcriptional and post-translational mechanisms [14]. More than 1500 microRNAs (miRNAs) have been annotated in our species, and at least one-third of all our genes are predicted to be miRNA targets. Basically, miRNAs are negative regulators. Interspecies comparisons indicated a high degree of evolutionary conservation, and at the intraspecific level there is a substantial reduction in nucleotide diversity in miRNAs regions compared with their flanking, other genic, or intronic counterparts. This trend was most marked in the first 14 nucleotides of mature miRNA forms, where no variants were detected. Both cross- and within-species approaches clearly revealed phenotypic consequences to the variation that was found. In the first case, species-specific physiological traits may appear, while at the intraspecific level phenotypic diversity may arise.

Evolutionary constraints are also found in the miRNA target genes, although single-nucleotide polymorphisms (SNPs) are not rare there. Around 400 human-specific mutations have been detected that might change regulatory interactions present in other species.

What makes us human? This question was asked by Sholtis and Noonan [15], who considered the role of gene regulation in the road to humanity. In addition to protein-coding genes, changes in gene expression and regulation undoubtedly played a role. Genome-wide *in vivo* and *in vitro* screens have identified thousands of distant-acting, *cis*-regulatory elements and global identification of enhancer regions strongly supports a modular regulatory architecture for many developmental genes, including master regulators (genes expressed at the beginning of a developmental lineage, regulating many downstream genes). Comparative genomics identified 992 human accelerated conserved noncoding sequences (HACNSs), and verified that these sequences are overrepresented near genes related to neuronal cell adhesion. These human accelerated regions (HARs) included transcribed and non-transcribed sequences, and HAR1, the most accelerated element, is a noncoding RNA expressed in the developing brain.

One of these studies compared the human prodynorphin (*PDYN*) promoter with orthologous sequences in nonhuman primates. *PDYN* encodes several opioid receptor ligands involved in perception and memory regulation, and the comparisons indicated that positive selection favored five human-specific substitutions in a 68 bp tandem repeat in the *PDYN* promoter.

HACNS1 is one of the most fast evolving of these regulatory elements. Although it is most conserved elsewhere, it accumulated 16 human-specific substitutions, 13 of which are clustered in a region of 81 bp. The hypothesis that this region could influence limb development was successfully tested in transgenic mouse embryos, which clearly showed that human HACNS1 acted as an enhancer of gene expression in their developing anterior limbs.

Doubts were expressed whether these substitutions had resulted from biased gene conversion (errors in DNA repair in meiosis, not random, that would lead to three copies of one allele and one of the other during recombination in meiotic prophase), and not by positive selection. Hünemeier et al. [16] tested this question using data from populations scattered over all continents and two prehistoric individuals. The absence of variability in all of them argued in favor of past positive and present conservative selection, a result that would agree with the importance of this region for *Homo*-specific characteristics as important as opposable thumbs, manual dexterity, and bipedal walking.

A technique that is promising for the investigation of DNA changes is high-throughput time-lapse imaging in live cells with DNA breaks in defined chromosomal locations marked by binding sites for fluorescent reporter proteins. Rocha et al. [17] reviewed results using this technique in double-strand breaks that can lead to chromosomal translocations. They found that the two ends of the same break, during the break-partner search, move together and separate only after a translocation completion. This technique can investigate the three-dimensional organization of the genome, examining factors that control legitimate or illegitimate DNA recombination.

Box 2.4 A schematic outline of mammal and avian sex chromosome history.

History	Time (million years ago)
1. Original situation. One pair of homologous chromosomes	>300
2. Three independent sex chromosome originations	
2.1. Original placentals and marsupials (monotremes)	181
2.2. Avians	175
2.3. Placental mammals	137
3. Rate of change	
3.1. Rapid	137–26
3.2. Stability	25
4. Factors that influenced the change	
4.1. Recombination suppression	
4.2. Purifying selection	
4.3. Maybe positive selection	
4.4. Dosage compensation (X chromosome inactivation)	

Sources: References 18–22.

Sex chromosomes

The human sex chromosomes are a fascinating example of evolutionary change. The first feature that impresses in an X–Y chromosome comparison is the size difference. The Y chromosome is 2.6 smaller than the X chromosome, and has 10× fewer genes (only 3% of its ancestral genes survived versus 98% in the X chromosome). Since both originated from a single pair of autosomes of the same size, this change is pronounced. How did this happen? Box 2.4 gives a schematic outline of mammalian and avian sex chromosome history. The original situation, which existed more than 300 million years ago (YA), gave rise to three sex chromosome originations in monotremes, avians, and placental mammals, which surprisingly occurred relatively close chronologically, 181, 175, and 137 million YA, respectively. It should be pointed out, however, that for avians these chromosomes developed in their saurischian reptile ancestors. Avians (and some reptiles) carry the reverse sex chromosome constitution compared with mammals (the heterogametic sex is female, ZW instead of XY). The monotreme lineage shows strong differences after split from their sister lineage, the placental mammals. For instance, *Platypus* has not less than five X and five Y chromosomes! Comparison of diverse mammalian Y chromosomes has shown that the rate of Y chromosome decay was rapid at first (from

137 to 26 million YA), but subsequently stabilized for 25 million years. Therefore, there is no danger of an imminent loss of the human Y chromosome!

What factors influenced these chromosomal changes? The classical explanation, based on animal models, was that recombination suppression between the proto-X and proto-Y chromosomes was the main factor. While random errors in gametogenesis and fertilization could have given rise to some of these chromosomal developments, negative selection in the Y chromosome seems highly probable, since harmful alleles are more easily eliminated in hemizygosis rather than in homo- or heterozygosis. On the other hand, positive selection is more likely to have occurred in the Y genes that were maintained.

Another factor that undoubtedly played a role in the process is dosage compensation. In the original autosomes that gave rise to both sex chromosomes, all genes would be expressed in two copies, in both males and females. With the loss of gene content and expression in human males, inactivation evolved to silence one copy of the gene in females (inactivation of one X chromosome), thus establishing an equilibrium of gene expression in the two sexes for genes that were maintained in both chromosomes. When this inactivation is imperfect or the region escaped inactivation, sexually antagonistic mutations may evolve, beneficial in one sex, but detrimental in the other. The classical example is Turner syndrome (impaired development due to X chromosome monosomy).

The Y chromosome's role in sex determination, on the other hand, evolved through the acquisition of new functions in spermatogenesis or development, by temporal and/or spatial expression shifts. The ampliconic region (formed by segments of the genome that make multiple copies after exposure of the organism to a compound that inhibits the function of a gene in the segment), which is rich in repetitive and palindromic sequences, probably avoided genetic decay through intrachromosomal gene conversion among members of a multicopy gene family.

Further insights into these processes can be obtained through a closer look at the constitution of the Y chromosome (Box 2.5). Three regions can be distinguished in its male-specific chromatin: (a) X-transposed; (b) X-degenerate; and (c) ampliconic. The latter is the largest (10.2 Mb), with seven blocks distributed in both arms, 60 codifying genes expressed mainly or exclusively in testicles, and 13.3 transcription units per Mb. The smallest constituent is the X-translocated (only 3.4 Mb, two genes, and 0.6 transcription units per Mb), while the X-degenerate is intermediary in these three characteristics, notably in size.

Evolutionary strata in the X chromosome were identified by higher values of synonymous site divergence between X and Y chromosomes, due to unique

Box 2.5 Characteristics of the three classes of sequences that occur in the male-specific portion of the human Y chromosome.

Characteristics	Regions		
	X-transposed	**X-degenerate**	**Ampliconic**
Distinctive aspects	99% X-identity	Homologous to X genes	High similarity with male-specific sequences
Evolutionary origin	Simple X-transposition	Relics of earlier autosomes	Obtained from various sources and afterward amplified
Distribution	Two blocks in Yp	Eight blocks in Yp and Yq	Seven blocks in Yp and Yq
Size (Mb)	3.4	8.6	10.2
Number of codifying genes	2	16, the majority with extended expression	60, in 9 families, expression mainly or exclusively in testicles
Number of noncodifying transcription units	0	4	74
Number of transcription units per Mb	0.6	2.2	13.3

Source: Reference 23. The numbers given may have been more recently updated, but the basic relationships remain. See also Reference 19.

mutations accumulated on their genes in older strata. Further studies along the entire X chromosome revealed a gradient of smaller number of synonymous mutations from its short arm (where the newest strata are located) to the end of the long arm (with oldest strata). The trend, however, showed no obvious boundaries.

Highly punctuated patterns of population structure are present in the X chromosome; 87% of X-linked HapMap SNPs within the top 1% of interpopulation variability (F_{ST} values) cluster into five distinct loci [24]. In addition, comparisons on size, genes per chromosome, associations and loci in genome-wide association data banks, between autosomes, X chromosome, and Y chromosome, verified that the X chromosome shows less associations than those expected by its size.

Comparison between the X chromosomes of humans and mice showed that in the nonampliconic region there is 95% interspecies similarity, which is surprising because the ampliconic genes were independently acquired and showed differences involving functions in male gametogenesis [25].

Paleogenomics

The leading investigator in this field is presently Svante Pääbo, a Swedish molecular biologist now working at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany. He recently published a book [26] describing aspects of his personal and scientific life, from 13 to 58 years of age. The book describes the problems and achievements related to this area of research in detail, vividly indicating frustrations and accomplishments. The book is a precious contribution to the history of science.

Starting in the mid-1980s, researchers realized that DNA could be successfully extracted from fossil material, thereby providing direct information about the evolutionary past. Focus was first concentrated on mitochondrial DNA (mtDNA), due to its abundance in cells, and, therefore, its easier retrieval. The first studies included human material that was thousands of years old, but in the enthusiasm of the time, by the 1990s reports appeared giving putative information from animal and plant fossils that were millions of years old (ironically named *antediluvian DNA* by Tomas Lindahl). By the mid-1990s, however, critical review showed that ancient DNA had many degradation and contamination problems, and that all these early results should be revised.

The extraction and degradation problems that should have been considered were the following: (a) not all ancient remains contained DNA that could be extracted, and even if it was there generally it was in minute amounts; new extraction techniques would have to be developed; (b) due to degradation, only small pieces, 100–200 nucleotides long, could be retrieved; (c) the polymerase reaction could sometime stitch short pieces of DNA together, forming artifactual hybrid molecules; and (d) contamination with microbial and human modern DNA was an unavoidable problem, depending on the sites' condition, such as temperature and depositional history, as well as postexcavation handling and source tissue.

A long history of extremely detailed and time-consuming attempts to solve these problems ensued. Box 2.6 presents some of the key developments. Great precautions against contamination and degradation were established, with the use of strictly sterile rooms, equipments, reagents, and garments, and duplication of findings in different laboratories was required. In addition, progress in high-throughput sequencing techniques codeveloped with new laboratory procedures that could be followed, including new methods of DNA capture, artificial adaptors for DNA fragment ends, improved enrichment procedures, and more stringent sequence-read quality controls. The sequencing results themselves could then be used to check for contamination, employing individual fragments overlapping informative differences between old and new DNAs. Repair of ancient postmortem lesions was also employed using uracil DNA glycosylase. These procedures led the researchers to presently evaluate the degree of contamination of their material to 1% or less.

The information obtained in the last two decades (Box 2.6) can be regarded as truly exceptional. Starting in the 1990s with only a few hundred nucleotides of mtDNA firmly determined, this work progressed quickly to the elucidation of the complete genome of this organelle, and in the first decade of the 2000s the full nuclear genomic sequences of several paleohominins were to be established. They included both Neanderthals and Denisovans, the latter a mysterious hominin for which just a small finger bone, and afterward a molar tooth, had been discovered in a southern

Box 2.6 Key developments in paleogenomic techniques and results.

Developments	References
1. Techniques	
1.1. Polymerase chain reaction (PCR) plus bacterial cloning	Several, 1980s
1.2. Precautions against contamination and degradation (special isolated clean rooms with strict sterilization procedures; independent duplication of findings in different laboratories)	Several, 1990s
1.3. 454 Life Sciences high-throughput sequencing	Green et al. [27]
1.4. High-throughput sequencing with array-based sequence capture, bead-based enrichment, and other approaches, Roche GS or Illumina	Briggs et al. [28]
1.5. New targeted enrichment capture	Carpenter et al. [29]
1.6. Postmortem degradation score	Skoglund et al. [30]
1.7. Methylation maps using cytosine deamination	Gokhman et al. [31]
2. Results	
2.1. 300 nucleotides, mtDNA, Tyrolean Ice Man	Handt et al. [32]
2.2. 370 nucleotides, mtDNA, Neanderthal-type specimen	Krings et al. [33]
2.3. mtDNA, Mezmaiskaya Cave, Neanderthal	Ovchinnikov et al. [34]
2.4. mtDNA, three Neanderthals	Krings et al. [35]
2.5. mtDNA, four Neanderthals	Serre et al. [36]
2.6. First nuclear data, 60 kpb, direct cloning, Neanderthal	Noonan et al. [37]
2.7. Nuclear sequences, 1 million bp (but 40% contamination), Neanderthal	Green et al. [38]
2.8. Complete mtDNA sequence, Neanderthal	Green et al. [39]
2.9. Complete DNA sequence, Paleo-Eskimo, Saqqaq Culture	Gilbert et al. [40]
2.10. Targeted retrieval and analysis of five Neanderthal mtDNAs	Briggs et al. [28]
2.11. 79% of the diploid genome, Paleo-Eskimo	Rasmussen et al. [41]
2.12. Draft sequence, Neanderthal genome	Green et al. [42]
2.13. Complete mtDNA sequence, Denisova	Krause et al. [43]
2.14. Denisovan genome	Reich et al. [44]
2.15. Aboriginal Australian genome	Rasmussen et al. [45]
2.16. Sima de los Huesos hominin, mtDNA	Meyer et al. [46]
2.17. Tianyuan Cave hominin genome	Fu et al. [47]
2.18. Clovis boy genome	Rasmussen et al. [48]
2.19. Complete exomes, two Neanderthals from Spain and Croatia	Castellano et al. [49]
2.20. Altai Neanderthal genome	Prüfer et al. [50]
2.21. DNA methylation maps, Neanderthal and Denisova	Gokhman et al. [31]

Siberian site. Full nuclear genomes of fossils from China, Australia, the Arctic, and America were also obtained, giving insights into evolutionary events of the Paleolithic, and the complete mtDNA of a 400,000 YA femur from Sima de los Huesos, Spain, has been sequenced, and surprisingly was more closely related to the Denisovans, who lived thousands of kilometers away and hundreds of years later, than to nearby Neanderthals. It is therefore possible that the Sima de los Huesos hominins represented a founder population that once lived all over Eurasia, but other hypotheses could also be made.

Shapiro and Hofreiter [51] provided a review of paleogenomics in general, while Prüfer et al. [50] furnished a panorama of the possible dynamics of gene flow that occurred between Neanderthals, Denisovans, modern humans, and another unknown hominin in the Late Pleistocene (Figure 2.1). Of course, the number of sampled individuals was small, the pace of research in this area is high, and therefore the picture may change soon. However, it seems that hominin groups met and had offspring on many occasions in the Late Pleistocene. Although low, gene flow from Neanderthals to Denisovans and Eurasian individuals,

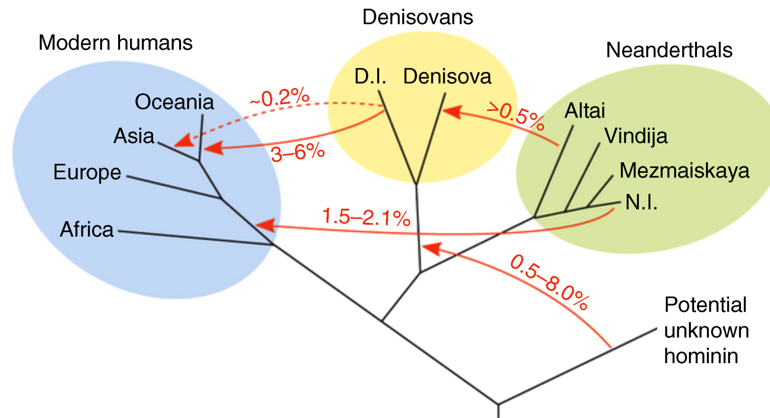


Figure 2.1 Geography of Late Pleistocene hominin admixture. The dashed line indicates uncertainty whether Denisovan gene flow into modern human in mainland Asia occurred directly or via Oceania. D.I.: introgressing Denisovans; N.I.: introgressing Neanderthals. (Reproduced from Prüfer et al. 2014 [50] with permission of Macmillan Publishers Limited.) (See the Color Plates section.)

from Denisovans to Oceanians, and from an unknown hominin to Denisovans seems to have occurred.

The question of the extent of admixture that occurred between Neanderthals and modern humans in general has been considered for many years now. An ingenious method of evaluation is to use patterns of linkage disequilibrium from many evolutionary independent loci to identify putatively archaic haplotypes and afterward running simulations to determine whether they are consistent with such an event. This approach has the advantage of assessing such admixture in populations with no fossil representatives, and it has consistently shown evidence for ancient admixture in many human populations. The numbers obtained, however, vary widely, and they are sensitive to assumptions regarding human demographic history [52,53]. Actually, Herrera et al. [54] and Lowery et al. [55] even questioned whether admixture between these two lineages had occurred, suggesting that the best explanation for the similarities found would be ancestral common polymorphisms.

Sankararaman et al. [56] used a mixed approach, combining data from the Altai Neanderthal genome with those of the 1000 Genomes Project Phase 1, as well as other sequences. They verified that genes that are more highly expressed in testes than in any other tissue show a notable lack of Neanderthal ancestry, which concurs with the fivefold reduction of this ancestry in the X chromosome. A possible explanation for these findings would be decreased fertility of Neanderthal/modern human hybrids.

Another possibility for haplotype analysis is to simply survey modern populations for those haplotypes that carry derived alleles shared with Neanderthals but absent in sub-Saharan Africans, of putatively young age (<130,000 YA; after modern humans first attempted to leave Africa). An X-linked haplotype that fulfills these characteristics is an 8 kb intronic segment of the dystrophin gene, *dys44*. Out of 12 major *dys44* haplotypes, B006 presented the indicated characteristics, and Yotova et al. [57] found that it was widely spread worldwide (9% overall), indicating, due to this result, a very early admixture between the expanding African migrants and Neanderthals.

Exome analyses in three Neanderthals [49] found that genes involved in skeletal morphology have changed more in the lineage leading to Neanderthals than in the ancestral lineage common to archaic and modern humans, while genes involved in behavior and pigmentation have changed more in the modern human lineage.

Gokhman et al. [31], on the other hand, used the natural process of cytosine deamination, in which unmethylated cytosines decay with time to uracils and methylated cytosines decay to thymines, to construct DNA methylated maps of Neanderthals and Denisovans. They found around 2000 differentially methylated regions compared with present-day humans. Specifically, they found substantial methylation changes in

the HOXD cluster, which may explain anatomical differences between archaic and present-day humans. These results are of obvious importance, since they may lead to approaches exploring regulatory changes, in a model of burst-like evolution where multiple genes may change their activities after a single evolutionary event.

Variability: mtDNA

A total of 5140 human mtDNA genomes were considered for an analysis of their variability, and the results are presented in Table 2.2. As expected, this variability was much more pronounced in the regulatory control than in the coding region. Thus, considering all mtDNA polymorphisms, the respective percentages of total variability were 50% versus 36%, while when the comparison was made for only the most prevalent SNPs (above 1 per 1000), the values were 26% versus 13%, respectively. Also as expected, the prevalence of transversions was higher in the control region (1:2.9 versus 1:7.5; 1:6.8 versus 1:21.2). However, there are wide differences in variability among the genes, all responsible for the formation of enzymatic complex subunits related to oxidative phosphorylation, generator of energy. Thus, when all polymorphisms were considered, the numbers varied from 29% (*ND3*, *ND4L*) to 57% (*ATP8*), and for those above 1 per 1000 from 10% (*COI*, *ND3*, *ND4L*) to 21% (*ATP8*).

Table 2.2 Diversity present in 5140 human mitochondrial genomes distributed all over the continents.

Mitochondrial DNA regions and characteristics considered	All polymorphisms	Only the most frequent (≥0.1%)
Control region		
% of polymorphic sites	50	26
Transversion/transition ratio	1:2.9	1:6.8
Coding region		
% of polymorphic sites		
ND1	37	15
ND2	33	14
COI	30	10
COII	34	13
ATP8	57	21
ATP6	54	18
COIII	35	11
ND3	29	10
ND4L	29	10
ND4	30	12
ND5	34	12
ND6	38	13
CYTB	43	16
Total	36	13
Transversion/transition ratio	1:7.5	1:21.2

Reproduced from Pereira et al. 2009, [58] with permission of Elsevier.

Other results, not given in the table [58], are as follows: (a) there is less variation in the second codon position (all: first, 24%; second, 13%; third, 63%; most frequent: first, 23%; second, 9%; third, 68%); and (b) the majority of the polymorphic codons codified for apolar, neutral amino acids (66.3%); the other categories were neutral polar amino acids (25.6%), basic polar amino acids (5%), and acidic polar amino acids (3%).

How many SNPs are needed to yield a maximum interpopulation discrimination power? Salas and Amigo [59] screened more than 2000 complete genomes from samples representing the three main continental human groups (Africa, Europe, and Asia), as well as two mixed sets of people ("African Americans" and "Hispanics"), and used this information as a starting point to develop a new simulation-based method. Haplotype diversity was measured for each SNP combination and compared with other panels available from the literature. They found that only a reduced number of SNPs (6–22) are needed to account for 95% of the maximum haplotype diversity of a given sample, but there is not a perfect set of "universal" SNPs suitable for any population comparison.

Nuclear variability

The 1000 Genomes Project Consortium [60] reported the genomes of 1092 individuals living in 14 populations spread over Europe, East Asia, sub-Saharan Africa, and the Americas, who were analyzed through a combination of low-coverage (2–6×) whole-genome sequences, targeted deep (50–100×) exome sequences, and dense SNP genotype data. Table 2.3 shows some key aspects of the results given, separately for autosomes and chromosome X. Not less than 38 million SNPs, 1.4 million biallelic insertions and deletions (indels), and 14,000

Table 2.3 Autosome/chromosome X human genome variability.

Characteristic	Autosomes	Chromosome X
Number of samples	1092	1092
Total number of raw bases (Gb)	19,049	804
SNPs, number of sites (Mb)	36.7	1.3
Average number of SNPs per sample	3.6 Mb	105 kb
Indels, number of sites	1.38 Mb	59 kb
Average number of indels per sample (kb)	344	13
Large deletions, number of sites	13,800	432
Average number of variants per sample	717	26

Reproduced from 1000 Genomes Project Consortium 2012 [60] with permission of Macmillan Publishers Limited.

large deletions were uncovered. The average number of these variants per sample was large, indicating the need for sampling at the local level, instead of relying on just a few nuclear populations, for the proper understanding of our species variability.

A key aspect of this extension from the pilot phase of this project to the indicated study was the characterization of low-frequency variants, since they may give insights into the process of elimination of deleterious mutations, a question of medical importance, and additionally because they tend to be recent in origin, and therefore show increased levels of interpopulation variability.

The results showed that variants present at 10% or above are generally found in all the populations studied. However, 17% of those at low frequency (0.5–5%) were observed in a single ancestry (continental) group, and 53% of those at the 0.5% frequency were observed in a single population. Moreover, (a) 17–343 SNPs have a difference of at least 0.25 between pair of populations within an ancestry group; (b) populations with substantial African ancestry carry up to three times as many low-frequency (0.5–5%) variants as those of European or East Asian origin; and (c) all populations show rare (<0.5%) frequency variants at higher levels than those expected, reflecting recent increases in population sizes and the effects of geographic differentiation.

Arbiza et al. [61] analyzed the data of the indicated project in relation to the X-linked and autosome diversities, considering 569 females from the 14 populations. To properly examine the ratio between the two diversities, it is important to take into consideration at least four factors: (a) differential mutation rates between the X chromosome and the autosomes; (b) sex-biased demographic events or social practices leading to different effective male and female populations; (c) changes in this parameter over time; and (d) the effect of natural selection, especially the exposure of recessive X-linked variants in hemizygous males.

They [61] took these factors into consideration by normalizing the divergence from the rhesus monkeys and other macaques for the X chromosome and autosomes separately. Normalized diversity was lower on the X chromosome than on the autosomes for all populations, but the degree of the difference varied in the diverse continental groups: 0.76–0.77 for the three African, 0.64–0.65 for the five European, and 0.60–0.62 for the three East Asian populations. The three admixed populations from the Americas showed a higher degree of interpopulation variation; the numbers were 0.71 for Puerto Ricans, 0.66 for Colombians, and 0.64 for Mexicans. These differences should be due to contrasting proportions of continental ancestry between them, and the fact that they were determined by diverse degrees of sex-biased matings (in general predominance of European males and non-European females).

Both autosomal and X-linked diversities increased with distance from genes, consistent with the diversity-reducing effect of selection on linked sites through purifying selection (background selection), positive selection (genetic hitchhiking), or both. However, other factors could also be responsible for the differences. The authors [61] tested the possibilities with an array of population models, varying such factors as population bottlenecks, expansions, and differential migration, but they could only account for about half of the observed reductions from the expected X/A diversity value of 1.

Other studies involving the analysis of SNP variability are the following: (a) a novel collapsing method to identify low-frequency (<0.03) variant differences in different regions of the genome [62]; (b) an analysis using 128 populations worldwide that verified a strong role of geography, especially in Asian populations,

giving rise to human population structure [63]; and (c) an approach for the identification of admixture, which identified 100 events occurring over the past 4000 years [64].

Short tandem repeats (STRs) can also be used to identify interpopulation variation, and they were used by Ramachandran and Rosenberg [65] to examine 678 such loci in 68 indigenous populations from Eurasia and the Americas. Genetic differentiation increases more rapidly along lines of longitude in the Americas than along lines of latitude in Eurasia. The authors suggested that the lower latitudinal rate of gene flow occurring among Native Americans could be due to a slower diffusion of crops and technologies through the Americas, compared with the corresponding diffusion in Eurasia.

Genovese et al. [66], on the other hand, described an approach for localizing the human genome missing pieces using the variability due to population admixture. They mapped the location of 70 scaffolds comprising 4 million base pairs in the euchromatic sequence, identifying also eight new large interchromosomal segmental duplications. They verified that most of these sequences are hidden in the genome's heterochromatin, especially in its pericentromeric regions.

Transposable elements (TEs, which migrate from the genome of one organism to another through appropriate vectors) are especially common in the genome of our species; not less than 1.8 million *Alu* recognizable sequence residues, which have been inserted along tens of millions of years of primate evolution, have been identified, and other TE types are also known. Britten [67] examined 655 *Alu* sequences that give perfect full-length matches with those of other primates and that, due to the rate of mutation, should have been inserted in relatively recent times. He concluded, from a comparison with the chimpanzee DNA, that the rate of change of recent (perfect) *Alu* insertions should be 1.5 million years per mutation per *Alu*, a very rapid process. Since *Homo sapiens* is also known for its general rapid evolutionary change, Britten suggested that *Alu* insertions could be one of the factors responsible for this speed, perhaps due to their effects on rates of recombination, affecting the regulation of nearby genes.

Huff et al. [68], on the other hand, tested the frequencies of single-nucleotide polymorphisms around polymorphic *Alu* insertions from two completely sequenced human genomes. They reasoned that the genealogy of a region that contains a mobile element should be on average older than those of the rest of the genome, and estimated that the expected time to the most recent common ancestor (TMRCA) for regions containing a polymorphic insertion is two times longer than the genomic average. They also calculated that the effective population size of human ancestors living before 1.2 million years ago was 18,500, an unusually small population for a species spread across the entire Old World.

Another class of transposable elements is the L1 family. Han et al. [69] identified 73 specifically human deletion events associated with recombination in these elements after the human–chimpanzee divergence. Despite their low frequency, these events deleted 450 kb of our genome. Two different deletion mechanisms could be identified: (a) nonallelic homologous recombination (55 events) and (b) union of nonhomologous extremities (the remaining 18). The deletion positions were not correlated with the local rates of chromosome recombination.

Three almost simultaneous analyses of the variability of structural changes known as copy number variations (CNVs) appeared in 2010 [70–72]. One of them generated a comprehensive map of 11,700 CNVs greater than 443 base pairs, and generated reference genotypes for 4978 of them for 450 individuals of European, African, and East Asian ancestry. Mutation rates of the majority of them are on the order of 10^{-5} per generation, and 104 gave estimations on the order of 10^{-3} per generation, therefore being potential hotspots. The identified changes involve important functions, such as immune or brain development, and therefore should be of crucial importance for the evolution of our species.

Exomes and proteomes

The first question that may be asked is whether it is possible to identify human protein-coding genes that could have originated *de novo*. This event was considered unlikely in the pre-molecular era by persons as important as Susumu Ohno and François Jacob. However, today the consensus is that this process is not impossible. Actually, Wu et al. [73] described a total of 60 protein-coding genes that they propose have originated *de*

Box 2.7 Procedure adopted by Wu et al. [73] to identify the new origin of protein-coding genes.

1. Search for human protein sequences against those of other primates to find those that do not show protein orthologs.
2. Exclude genes without start or stop codons.
3. Check whether the nonhuman homolog sequences present frameshifts, premature stop codons, or no start codon.
4. Verify whether the human sequence presents specific mutations generating intact open reading frames.
5. Search for evidence of expression at the mRNA and protein levels.
6. Verify whether there is no other highly similar coding sequence in the human genome, therefore eliminating the possibility of generation by gene duplication.

novo in the human lineage since divergence from the chimpanzee. The experimental procedure they adopted to identify such genes is schematically given in Box 2.7. Basically, the gene should have no counterpart in other region of the human genome, and should have specific mutations generating intact open reading frames from sequences occurring in other primates that have frameshifts, premature stop codons, or no start codon. Wu et al. [73] also verified that the indicated genes have their highest expression levels in the cerebral cortex and testes, suggesting that they may contribute to phenotypic traits unique to humans that would lead to improved cognitive ability, and reproductive differences. Guerzoni and McLysaght [74] discussed these results, pointing out their significance, stressing the limitations that exist in defining and identifying *de novo* genes, and indicating the future challenges in this area.

The promises and limitations of population exomics for human evolutionary studies, on the other hand, were considered by Tennessen et al. [75]. Exome sequencing using next-generation technology is becoming increasingly common, and is the most cost-effective and practical way to rigorously characterize the spectrum of rare variation. However, there are limitations, which include the difficulty of dealing with extremely large files and in merging data sets, as well as the occurrence of cryptic paralogs. Even with these caveats, the method is a powerful tool for inferring evolutionary patterns in our species in an unbiased and comprehensive manner.

As a first attempt toward understanding how rare variants contribute to risk for complex disease, Tennessen et al. [76] sequenced 15,585 protein-coding genes at high depth (111×) in 2440 individuals of European and African ancestry. They identified more than 500,000 single-nucleotide variants, 86% with a minor allele frequency below 0.5%. On average, 2.3% of the variants each person carries should influence the function of 313 genes, and 96% of them are rare. As already indicated, this finding results from the effects of our recent explosive population growth, coupled with weak purifying selection.

Draft maps of the human proteome using the mass spectrometry method were simultaneously reported in 2014 [77,78]. The first authors [77] studied 30 histologically normal human samples, while the latter [78] considered 27 human tissues and body fluids. Wilhelm et al. [76] also verified the existence of a core proteome of approximately 10,000–12,000 ubiquitously expressed proteins responsible for the general control and maintenance of cells. However, a number of proteins are exclusively or preferentially detected in a given organ, and a selected list of them is given in Box 2.8. Studies like this one are essential for the proper understanding, at the molecular level, of the process from which a single cell develops in a high structured individual with multiple different tissues as is characteristic of our species.

Selection or drift? History

Are the evolutionary mechanisms effective at the organismal level also effective at the molecular level? Two articles published almost half a century ago by Japanese researchers [79,80] had enormous impact at the time, and raised a controversy that lasts till date. They argued that natural selection, considered the main factor of evolution since the seminal work of Darwin [81], would not be important at the molecular level. A definition of what was called the neutral theory of evolution was given by Kimura [82]: "at the molecular level most evolutionary change and most of the variability within species are not caused by Darwinian selection, but by random genetic drift of mutant alleles that are selectively neutral." The important part of the theory is not that

Box 2.8 Proteins with expression levels 10-fold above average in particular organs or tissues.

Organ or tissue	Proteins
Cerebral cortex	Membrane proteins, ion transport, immunoglobulins, angiogenesis, neuron differentiation.
Esophagus	Epithelial development, chloride channels, proteases.
Heart	Muscle process and regulation, respiratory chain, channels, and transporters.
Lung	Glycoproteins, cytokines, defense response, immunoglobulins, angiogenesis.
Breast	Fatty acid and lipid biosynthesis, steroid hormone signaling.
Stomach	*O*-Glycan biosynthesis, glycoproteins, drug metabolism, H^+/K^+ ATPase.
Spleen	Signal transduction, defense response, leukocyte activation.
Pancreas	Proteases and lipases.
Kidney	Oxidoreductases, amino acid metabolism, transporter.
Testis	Neuro/synaptic proteins, collagen synthesis.
Ovary	Growth factors and receptors.
Placenta	Oxygen transport, response to nutrients, steroid biosynthesis.

Source: Reference 78.

molecular mutants are selectively neutral, but that their fate is largely determined by random drift.

Along the 46 years that followed Kimura's proposition, a huge amount of theoretical and empirical analyses considered it in detail. Though a full review of them is outside the scope of this chapter, two relatively recent books by prominent scholars that challenged the Darwinian view of evolution indicate that the controversy will not be settled easily. Lynch [83] asserted that the central point of his analysis was that "many genomic features could not have emerged without a near-complete disengagement of the power of natural selection" and Nei [84], characterizing his mutation-driven theory of evolution, indicated that "genomic conservation and constraint-breaking mutation is the ultimate source of all biological innovations."

Selection or drift? Methods

How it would be possible to identify selection or selection sweeps at the molecular level? One of the possibilities is the comparison between the intra- and interspecific variability (Box 2.9). Increased levels of low-frequency variants would suggest negative directional selection; increased levels of common variants, positive directional selection; and intermediate levels, balancing selection. Selective sweeps would be mainly associated with an increase in rare variants.

Further details about some of the most employed tests for the detection of selection are given in Box 2.10. Tajima's D considers at the intrapopulation level mutation distributions and their frequencies in sequence sets. If purifying selection is present, D is less than zero, while a D higher than zero would suggest balancing selection. The problem is that equivalent results can also be obtained if a population is in expansion ($D<0$) or if there exists population subdivision ($D>1$). Therefore, other tests or observations are necessary to separate the selective process from simple demographic factors.

The four other tests listed after Tajima's D are also not independent of demographic factors. Two of them have linkage disequilibrium (i.e., how variants in different positions of a DNA sequence are correlated) as their basic tool. Selective sweeps are investigated checking whether the expected linkage blocks, and their sizes, occur as expected under the assumption of given evolutionary patterns, the same approach of the next method listed. As for the two others, F_{ST} and HKA (which are the initials of their proponents, R.R. Hudson, M. Kreitman, and M. Aguade) verify whether the genomic region under consideration shows levels of interpopulation variation higher than the genome average (F_{ST}), or whether the variability of a candidate locus significantly differs from that of another, apparently neutral one (HKA).

Box 2.9 Selection and mutation effects in the intra- and interspecific variability.

Evolutionary factor	Variability		Proportion (2)/(1)	Frequency distribution
	(1) Intraspecific	(2) Interspecific		
Increased mutation rate	Increases	Increases	No effect	No effect
Negative directional selection	Reduces	Reduces	Reduces if selection is not too strong	Increase of low-frequency variants
Positive directional selection	May increase or decrease	Increases	Increases	Increase of high-frequency variants
Balancing selection	Increases	May increase or decrease	Reduces	Increase of intermediate-frequency variants
Selective sweep (linked neutral sites)	Decreases	No effect of rate of substitution, but variance increases	Increases	Mostly increase of low-frequency variants

Source: Reference 85.

Finally, the last two tests listed in Box 2.10 use the ratio of nonsynonymous (d_N) to synonymous (d_S) substitutions. In the first case, the mutation would lead to an amino acid different from that which would be coded by the original sequence, while synonymous mutations do not change the amino acid that will be formed (due to the fact that the genetic code is degenerate or redundant; that is, a given amino acid can be codified for up to six

Box 2.10 Methods for the detection of selection based on DNA sequences and single-nucleotide polymorphisms.

Test	**Necessary data**	**Pattern**	**Requires multiple loci?**	**Robust to demographic factors?**
Tajima's *D* and related	Population genetic	Frequency spectrum	No	No
Selective sweep-spatial pattern	Population genetic	Frequency spectrum/spatial pattern	No	No
Linkage disequilibrium (LD)	Population genetic	LD and/or haplotype structure	No	No
F_{ST} and related	Population genetic	Amount of population subdivision	Yes	No
HKA	Population genetic and comparative	Number of polymorphisms/ substitutions	Yes	No
McDonald–Kreitman	Population genetic and comparative	Number of nonsynonymous and synonymous polymorphisms	No	Yes
d_N/d_S ratios	Comparative or population genetic without recombination	Nonsynonymous and synonymous substitutions	No	Yes

Source: Reference 85.

Box 2.11 Methods for detecting selection on the human genome.

Method	Timescale for detection	Advantages and disadvantages
High proportion of function-altering mutations	Many millions of years	Well-determined focus, but typically detects only ongoing or recurrent selection.
Reduction in genetic diversity	Less than 250,000 years	Selective sweeps can lead to the elimination or reduction of variability in nearby regions. This effect may be difficult to distinguish from that of a simple demographic expansion.
High-frequency derived alleles	Less than 80,000 years	Requires knowledge of the ancestral allele, usually inferred from a closely related species.
Interpopulation differences	Less than 50,000 years	Easier detection, but since the differentiation requires at least partial reproductive isolation, the determining events should have occurred after the out of Africa original migration.
Long haplotypes	Less than 30,000 years	Selective sweeps can lead to an allele that has high frequency and shows long-range associations with other alleles. However, these haplotypes may persist for relatively short periods of time.

Sources: References 86–90.

different codons). The great advantage of these tests is that they are not affected by demographic factors. In the absence of selection, $d_N/d_S = 1$; with negative selection, $d_N/d_S < 1$; and with positive selection, $d_N/d_S > 1$. Since DNA sequences are composed of codons (which codify for proteins, being therefore potentially functional) and introns (in which the variability is not directly related to a protein), methods should be used that take into consideration these differences.

Considering now, specifically, the human genome, the analytic methods that have been applied to it could be classified in five categories (Box 2.11). Each of them has merits and difficulties, and detects events in different scales of time. The most direct and with the highest time depth is the first listed. The four others are related to the genomic structure investigated or to the geographical or ecological distributions of given variants. More details about these processes, and about the investigation of evidence for selection in noncodifying, regulatory sequences, as well as data of gene expression can be found in Reference 91.

Akey and Shriver [92] considered how whole-genome sequence data could shed light on human evolution. They indicated some approaches that could be considered: (a) sexual selection and sex-biased dispersal; (b) mutation rate biases; (c) selection from standing variation; (d) directional and balancing selection acting on the same locus; (e) epistatic selection; and (f) the evolution of evolvability. Of course, these approaches are easier to list than to analyze in a rigorous way. New tools would be welcome.

Selection or drift? Analyses

Box 2.12 lists six papers published between 2010 and 2013 that emphasized the role of demography, drift, and different mutation rates for the explanation of human genome diversity. They maintained that classic selective sweeps were rare, that diversity around synonymous and nonsynonymous substitutions was similar, and that amino acid and possible regulatory sites were not significantly enriched in alleles showing wide interpopulation differences. Other findings suggested that the fixation of human-specific pseudogenes and sequence variability could be explained without invoking selection pressures.

However, evidence for the action of selection on the human genome is overwhelming, and Boxes 2.13 and 2.14 list some more recent selected studies. That negative selection is an important evolutionary factor for humans was never doubted, and is exemplified by the vast array of genetic diseases that afflict us. Anyway, Box 2.13 lists

Box 2.12 Selected examples of studies emphasizing the role of mutation, demography, and drift for the explanation of human genomic variability.

Year	Main findings	References
2010	Severe bottleneck in the Middle Pleistocene. Human-specific pseudogene fixation was too rare to account for adaptation events between humans and chimpanzees.	Kim et al. [93]
2010	Variation in equilibrium genome-wide nucleotide composition is largely defined by variation in mutation biases.	Lynch [94]
2010	Human accelerated regions could be due to biased fixation processes.	Katzman et al. [95]
2011	Classic sweeps were not a dominant mode of human adaptation over the past 250,000 years.	Hernandez et al. [96]
2012	The expected signatures of adaptations to new environments are surprisingly lacking at the genomic level. The spatial dimension of human evolution should be taken into account.	Alves et al. [97]
2013	Both SNP and exome sequence data showed that the neutral mutation model fits the empirical distribution quite well.	Miura et al. [98]

five studies published in the last 5 years considering this subject in a quantitative way. No human person is completely "healthy," having at least 100 loss-of-function variants in heterozygosis, many with large effects on viability or fertility. On the other hand, the little impact that our demographic growth had on the burden of deleterious mutations is evidence against the view that we will face a future of genetic deterioration.

Studies considering balancing selection are not numerous in recent years. Four selected examples are as follows: (a) Fernandez-Vina et al. [104] reviewed HLA alleles, lineages, and haplotypes in worldwide populations and concluded that their distributions are consistent with overdominant (heterozygote advantage) selection. (b) Olson [105] suggested that long-term balancing selection appears largely to reflect the consequences of host–pathogen arms races, with local adaptations being widespread. (c) Leffler et al. [106] looked for long-lived balancing selection by searching combinations of SNPs shared between humans and chimpanzees. They maintained that balancing of selection has shaped human variation and, as Olson [105] suggested, that host–pathogen interactions could be identified as the common targets for this factor. (d) Other evidences for the importance of pathogens as main selective pressures were also obtained [107,108].

Box 2.13 Selected examples on the role of negative selection influencing human genomic variability.

Year	Main findings	References
2010	Clinical cohorts could be more extensively considered for selection analyses.	Stearns et al. [99]
2011	At least 5.5% of the human genome has undergone purifying selection, and 4.2% showed constrained elements.	Lindblad-Toh et al. [100]
2012	A typical "healthy" genome contains at least 100 loss-of-function variants, many with large effects on human fitness.	MacArthur et al. [101]
2013	The genetic risks for 102 diseases in 1043 unrelated persons across 51 populations indicate that they go well beyond of what is expected by genetic drift.	Corona et al. [102]
2014	Recent human demography had little impact on the average burden of deleterious mutations in human populations.	Simons et al. [103]

Box 2.14 Selected examples of studies emphasizing the role of positive selection for the explanation of human genomic diversity.

Year	Main findings	References
2010	Evidence for positive selection based on accelerated evolution regions suggests that neural development and function have adapted mainly through noncoding changes, while adaptation via coding changes involved mainly immunity, olfaction, and male reproduction.	Haygood et al. [109]
2010	Differences in allele frequencies across populations are associated with polar ecoregions, foraging, and a diet rich in tubers and roots.	Hancock et al. [110]
2011	Genic and nonsynonymous SNPs are enriched relative to nongenic SNPs for correlations with climate variables.	Hancock et al. [111]
2011	Phylogenetically orthogonal parallel patterns of local adaptation caused by subtle shifts at many widespread polymorphisms were found.	Tennessen and Akey [112]
2011	*ABCC11* rs17822931 is significantly associated with latitude, suggesting an adaptation to a cold climate.	Ohashi et al. [113]
2011	*ADH1B**48His* has increased frequencies in East Asian populations probably due to positive selection.	Li et al. [114]
2012	Genetic and genomic changes that are specific to *Homo sapiens* showing accelerated evolution display adaptive changes in complex loci that are highly enriched for disease associations.	O'Bleness et al. [115]
2012	List of 146 copy number variants showing evidence of positive selection were investigated based on the following criteria: (a) population differentiation, 104; (b) nonsynonymous substitution, 32; (c) function, 4; (d) high population differentiation + nonsynonymous substitution, 3; (e) linkage, 2; and (f) high population differentiation + linkage, 1.	Iskow et al. [116]
2012	Coevolution of four genes on haplotypes carrying the Alzheimer disease loci, interconnected through multiple interacting proteins, showed the action of recent positive selection.	Raj et al. [117]
2012	*ABCA1**Arg230Cys* polymorphism showed several evidences of positive selection, and especially of coevolution with maize domestication in Central America.	Hünemeier et al. [118]
2013	Regulatory elements (transcription factors) seem to contribute substantially to both adaptive substitutions and deleterious polymorphisms.	Arbiza et al. [119]
2013	The ectodysplasin receptor, EDARV 370A, shows signals of strong positive selection in both mice and humans. It is associated with an increased number of active eccrine glands in the Han Chinese.	Kamberov et al. [120]
2013	Clusters of human-specific substitutions that occurred since the chimpanzee–human divergence showed, among other factors, evidence for positive selection.	Xu et al. [121]
2013	A long-range haplotype method applied to data on 14 populations across the world indicated that positive selection events tended to cluster in populations of the same ancestry. A total of 405 regions were identified as positively selected, and 212 (52%) were shared by at least two populations.	Liu et al. [122]
2014	Significant extended haplotype homozygosity was detected in around 4 of the 14 LINE-1 retrotransposon insertions tested (29%).	Kuhn et al. [123]

Fourteen studies published between 2010 and 2014, which found evidence for positive selection in a series of genomic regions, are listed in Box 2.14. The types of approaches employed to detect the influence of this factor varied. Half of them relied on population differentiation, but other methods were employed as well. Special mention should be made of the paper by Iskow et al. [116], who searched for positive selection in 146 copy number variants. As with the studies listed in Box 2.14, the evidence for 104 (71%) of them was based on population differentiation, with 4 others (an additional 3%) being based on population differentiation in association with another method. Although, as indicated in Box 2.10, this approach is not robust enough to eliminate the influence of demographic factors, it is clear that the influence of positive selection

in a considerable fraction of our genome cannot be dismissed.

Nervous system and culture

What makes us human? Alfred Russel Wallace was one of the first to question, in the 19th century, the role of natural selection in the development of culture, which placed us in sharp contrast with our cousins, the chimpanzees. Three groups of researchers [124–126] argued that the road to humanity should have been paved by an increasing dependence on learned behaviors and culture, which in turn conditioned an extraordinary population growth, independently of natural selection.

Irrespective of the primary factors, however, our life history and behavior are strongly influenced by biological factors, and two examples of the differences between the brain function of humans and chimpanzees, as ascertained through molecular studies, can be cited.

First, Prabhakar et al. [127] investigated not less than 110,549 conserved noncoding sequences in the human lineage and verified that the strongest signal of accelerated evolution occurred near genes specifically involved in neuronal cell adhesion. The process seems to have been independent in the human and chimpanzee lineages, leading to the suggestion that *cis*-regulatory and other noncoding changes may have contributed to the modifications in brain development and function that gave rise to uniquely human cognitive traits.

That human ontogenesis proceeds at a slower rate than in other primates, a process called neoteny, has been well known for many years now. A molecular approach to this question, however, was lacking until Somel et al. [128] examined the messenger RNA expression in the prefrontal cortex of humans, chimpanzees, and rhesus macaques. They observed a significant excess of genes showing neotenic expression in humans, with dramatic heterochronous remodeling during postnatal development. More specifically, these changes were preferentially expressed in the gray matter, and this delayed gray matter maturation in the prefrontal cortex may have extended the period of neuronal plasticity associated with active learning, therefore providing humans with additional time to acquire knowledge and skills. Further discussion about these matters is presented in Chapter 7.

Conclusions

Science is not free from extraneous ideological and monetary influences, as was well exemplified by the episodes related to the work that resulted in the first draft of our genome, at the end of the 20th and beginning of the present century. On the other hand, progress in the knowledge of our genetic material is increasing at a phenomenal rate, involving huge gains in our understanding of its structure, function, and their interrelationships. Methodological developments have made possible, also, the extension of these studies beyond the extant genomes, giving adequate information about our ancient DNA. The peculiarities of human variability are being considered at different organismal and population levels, and alternative hypotheses about the factors influencing it developed. At present, the information we have about the human genome is unparalleled by the knowledge acquired in any other organism. The complexities of our evolutionary process, due to the decisive influence of sociocultural factors, however, make definitive conclusions difficult. It is hoped that the increased sophistication of the molecular and bioinformatic analyses will lead to important developments in the understanding of these process in a not very distant future.

Review questions and exercises

1. Envisage what would have been the studies related to the first draft of the human genome, if Craig Venter had not decided to investigate it using private, non-public, funds.

2. Describe the main characteristics of our genome. The whole picture suggests homogeneity or heterogeneity?

3. Search in the literature and databases and identify five normal and five abnormal conditions not listed in Box 2.2, obtaining their precise position in the genome.

4. If, for your investigation, you had to choose, which of the three approaches to function would you prefer, genetic, evolutionary, or biochemical?

5 How, in genetic and evolutionary terms, would you define sex?

6 If a Neanderthal, dressed in modern clothes, crossed by you in the street, would you immediately recognize its taxonomic state or not?

7 Which is better, an accelerated evolutionary rate or evolutionary conservation?

8 What are the advantages and disadvantages of an evolutionary approach considering nucleotides or amino acids?

9 Is the world, and your specific life, governed primarily by random or by deterministic factors?

10 In which ways sociocultural factors could influence our genome?

References

1 Dennis C et al. 2001. Everyone's genome. *Nature* **409**:813.
2 Sulston J, Ferry G. 2001. *The Common Thread: A History of Science, Politics, Ethics, and the Human Genome*. Joseph Henry Press, Washington, DC.
3 Waterston RH et al. 2002. On the sequencing of the human genome. *Proc Natl Acad Sci USA* **99**:3712–3716.
4 Waterston RH et al. 2003. More on the sequencing of the human genome. *Proc Natl Acad Sci USA* **100**:3022–3024.
5 Adams MD et al. 2003. The independence of our genome assemblies. *Proc Natl Acad Sci USA* **100**:3025–3026.
6 Rogers J. 2003. The finished genome sequence of *Homo sapiens*. *Cold Spring Harb Symp Quant Biol* **68**:1–11.
7 Dhand R. 2006. The "finished" landscape. *Nature* (Human Genome Collection Suppl): 1–305.
8 Weigmann K. 2004. The code, the text, and the language of God. *EMBO Rep* **5**:116–118.
9 McKusick VA. 1988. *Mendelian Inheritance in Man: Catalog of Autosomal Dominant, Autosomal Recessive, and X-Linked Phenotypes*, 8th edition. Johns Hopkins University Press, Baltimore, MD.
10 Gregersen PK. 2014. A genomic road map for complex human disease. *Science* **343**:1087–1088.
11 Kellis M et al. 2014. Defining functional DNA elements in the human genome. *Proc Natl Acad Sci USA* **111**:6131–6138.
12 Doolittle WF. 2013. Is junk DNA bunk? A critique of ENCODE. *Proc Natl Acad Sci USA* **110**:5294–5300.
13 Ziller MJ et al. 2013. Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**:477–481.
14 Li J, Zhang Z. 2013. miRNA regulatory variation in human evolution. *Trends Genet* **29**:116–124.
15 Sholtis SJ, Noonan JP. 2010. Gene regulation and the origins of human biological uniqueness. *Trends Genet* **26**:110–118.
16 Hünemeier T et al. 2010. Population data support the adaptive nature of *HACNS1* human–chimpanzee differences in a limb expression domain. *Am J Phys Anthropol* **143**:478–481.
17 Rocha PP et al. 2013. Finding the right partner in a 3D genome. *Science* **342**:1333–1334.
18 Sayres MAW, Makova KD. 2012. Gene survival and death on the human Y chromosome. *Mol Biol Evol* **30**:781–787.
19 Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet* **14**:113–124.
20 Clark AG. 2014. The vital Y chromosome. *Nature* **508**:463–465.
21 Cortez D et al. 2014. Origins and functional evolution of Y chromosomes across mammals. *Nature* **508**:488–493.
22 Bellott DW et al. 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508**:494–499.
23 Skaletsky H et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**:810–813.
24 Lambert CA et al. 2010. Highly punctuated patterns of population structure on the X chromosome and implications for African evolutionary history. *Am J Hum Genet* **86**:34–44.
25 Mueller JL et al. 2013. Independent specialization of the human and mouse X chromosome for the male germ line. *Nat Genet* **45**:1083–1087.
26 Pääbo S. 2014. *Neanderthal Man: In Search of Lost Genomes*. Basic Books, New York.
27 Green RE et al. 2006. Analysis of one million base pairs of Neanderthal DNA. *Nature* **444**:330–336.
28 Briggs AW et al. 2009. Targeted retrieval and analysis of five Neanderthal mtDNA genomes. *Science* **325**:318–321.
29 Carpenter ML et al. 2013. Pulling out the 1%: whole genome capture for the targeted enrichment of ancient DNA sequencing libraries. *Am J Hum Genet* **93**:852–864.
30 Skoglund P et al. 2014. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc Natl Acad Sci USA* **111**:2229–2234.
31 Gokhman D et al. 2014. Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. *Science* **344**:523–527.
32 Handt O et al. 1994. Molecular genetic analyses of the Tyrolean Ice Man. *Science* **264**:1775–1778.
33 Krings MA et al. 1997. Neandertal DNA sequences and the origin of modern humans. *Cell* **90**:19–30.
34 Ovchinnikov IV et al. 2000. Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* **404**:490–493.
35 Krings M et al. 2000. A view of Neandertal genetic diversity. *Nat Genet* **26**:144–146.

36 Serre D et al. 2004. No evidence of Neandertal mtDNA contribution to early modern humans. *PLoS Biol* **2**:313–317.
37 Noonan JP et al. 2006. Sequencing and analysis of Neandertal genomic DNA. *Science* **314**:1113–1118.
38 Green RE et al. 2006. Analyses of one million base pairs of Neanderthal DNA. *Nature* **444**:330–336.
39 Green RE et al. 2008. A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**:416–426.
40 Gilbert MTP et al. 2008. Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* **320**:1787–1789.
41 Rasmussen M et al. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**:757–762.
42 Green RE et al. 2010. A draft sequence of the Neandertal genome. *Science* **328**:710–722.
43 Krause J et al. 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464**:894–897.
44 Reich D et al. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**:1053–1060.
45 Rasmussen M et al. 2011. Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**:94–98.
46 Meyer M et al. 2013. A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**:403–406.
47 Fu Q et al. 2013. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc Natl Acad Sci USA* **110**:2223–2227.
48 Rasmussen M et al. 2014. The genome of a Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**:225–229.
49 Castellano S et al. 2014. Patterns of coding variation in the complete exomes of three Neandertals. *Proc Natl Acad Sci USA* **111**:6666–6671.
50 Prüfer K et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**:43–49.
51 Shapiro B, Hofreiter M. 2014. A paleogenomic perspective on evolution and gene function: new insights from ancient DNA. *Science* **343**:1236573-1–1236573-7.
52 Wall JD, Slatkin M. 2012. Paleopopulation genetics. *Annu Rev Genet* **46**:635–649.
53 Vernot B, Akey JM. 2014. Resurrecting surviving Neandertal lineages from modern human genomes. *Science* **343**:1017–1021.
54 Herrera KJ et al. 2009. To what extent did Neanderthals and modern humans interact? *Biol Rev* **84**:245–257.
55 Lowery RK et al. 2013. Neanderthal and Denisova genetic affinities with contemporary humans: introgression versus common ancestral polymorphisms. *Gene* **530**:83–94.
56 Sankararaman S et al. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**:354–357.
57 Yotova V et al. 2011. An X-linked haplotype of Neandertal origin is present among all non-African populations. *Mol Biol Evol* **28**:1957–1962.
58 Pereira L et al. 2009. The diversity present in 5140 human mitochondrial genomes. *Am J Hum Genet* **84**:628–640.
59 Salas A, Amigo J. 2010. A reduced number of mtSNPs saturates mitochondrial DNA haplotype diversity of worldwide population groups. *PLoS One* **5**:e10218.
60 1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**:56–65.
61 Arbiza L et al. 2014. Contrasting X-linked and autosomal diversity across 14 human populations. *Am J Hum Genet* **94**:827–844.
62 Moore CB et al. 2013. Low frequency variants, collapsed based on biological knowledge, uncover complexity of population stratification in 1000 Genomes Project data. *PLoS Genet* **9**:e1003959.
63 Wang C et al. 2012. A quantitative comparison of the similarity between genes and geography in worldwide human populations. *PLoS Genet* **8**:e1002886.
64 Hellenthal G et al. 2014. A genetic atlas of human admixture history. *Science* **343**:747–751.
65 Ramachandran S, Rosenberg N. 2011. A test of the influence of continental axes of orientation on patterns of human gene flow. *Am J Phys Anthropol* **146**:515–529.
66 Genovese G et al. 2013. Using population admixture to help complete maps of the human genome. *Nat Genet* **45**:406–414.
67 Britten RJ. 2010. Transposable element insertions have strongly affected human evolution. *Proc Natl Acad Sci USA* **107**:19945–19948.
68 Huff CD et al. 2010. Mobile elements reveal small population size in the ancient ancestors of *Homo sapiens*. *Proc Natl Acad Sci USA* **107**:2147–2152.
69 Han K et al. 2008. L1 recombination-associated deletions generate human genome variation. *Proc Natl Acad Sci USA* **105**:19366–19371.
70 Conrad DF et al. 2010. Origins and functional impact of copy number variation in the human genome. *Nature* **464**:704–712.
71 Fu W et al. 2010. Identification of copy number variation hotspots in human populations. *Am J Hum Genet* **87**:494–504.
72 Sudmant PH et al. 2010. Diversity of human copy number variation in multicopy genes. *Science* **330**:641–646.
73 Wu D-D et al. 2011. *De novo* origin of human protein-coding genes. *PLoS Genet* **7**:e1002379.
74 Guerzoni D, McLysaght A. 2011. *De novo* origins of human genes. *PLoS Genet* **7**:e1002381.
75 Tennessen JA et al. 2011. The promise and limitations of population exomics for human evolution studies. *Genome Biol* **12**:e127.
76 Tennessen JA et al. 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**:64–69.

77 Kim M-S et al. 2014. A draft map of the human proteome. *Nature* **509**:575–581.
78 Wilhelm M et al. 2014. Mass-spectrometry-based draft of the human proteome. *Nature* **509**:582–587.
79 Kimura M. 1968. Evolutionary rate at the molecular level. *Nature* **217**:624–626.
80 Kimura M, Ohta T. 1971. *Theoretical Aspects of Population Genetics*. Princeton University Press, Princeton, NJ.
81 Darwin C. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London.
82 Kimura M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.
83 Lynch M. 2007. *The Origins of Genome Architecture*. Sinauer, Sunderland, MA.
84 Nei M. 2013. *Mutation-Driven Evolution*. Oxford University Press, Oxford.
85 Nielsen R. 2005. Molecular signatures of natural selection. *Annu Rev Genet* **39**:197–218.
86 Sabeti PC et al. 2006. Positive natural selection in the human lineage. *Science* **312**:1614–1620.
87 Harris EE. 2008. Searching the genome for our adaptations. *Evol Anthropol* **17**:146–157.
88 Akey JM. 2009. Constructing genomic maps of positive selection in humans: where do we go from here? *Genome Res* **19**:711–722.
89 Zhong M et al. 2010. A powerful score test to detect positive selection in genome-wide scans. *Eur J Hum Genet* **18**:1148–1159.
90 Peter BM et al. 2012. Distinguishing between selective sweeps from standing variation and from a *de novo* mutation. *PLoS Genet* **8**:e1003011.
91 Hellmann I, Nielsen R. 2008. Human evolutionary genomics. In: Pagel M, Pomiankowski A (eds), *Evolutionary Genomics and Proteomics*. Sinauer, Sunderland, MA, pp 271–297.
92 Akey JM, Shriver MD. 2011. A grand challenge in evolutionary and population genetics: new paradigms for exploring the past and charting the future in the post-genomic era. *Front Genet* **2**:e47.
93 Kim HL et al. 2010. Divergence, demography and gene loss along the human genome. *Philos Trans R Soc B* **365**:2451–2457.
94 Lynch M. 2010. Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci USA* **107**:961–968.
95 Katzman S et al. 2010. GC-biased evolution near human accelerated regions. *PLoS Genet* **6**:e1000960.
96 Hernandez RD et al. 2011. Classic selective sweeps were rare in recent human evolution. *Science* **331**:920–924.
97 Alves I et al. 2012. Genomic data reveal a complex making of humans. *PLoS Genet* **8**:e1002837.
98 Miura S et al. 2013. Random fluctuation of selection coefficients and the extent of nucleotide variation in human populations. *Proc Natl Acad Sci USA* **110**:10676–10681.
99 Stearns SC et al. 2010. Measuring selection in contemporary human populations. *Nat Rev Genet* **11**:611–622.
100 Lindblad-Toh K et al. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**:476–482.
101 MacArthur DG et al. 2012. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**:823–828.
102 Corona E et al. 2013. Analysis of the genetic basis of disease in the context of worldwide human relationships and migration. *PLoS Genet* **9**:e1003447.
103 Simons YB et al. 2014. The deleterious mutation load is insensitive to recent population history. *Nat Genet* **46**:220–224.
104 Fernandez-Vina MA et al. 2012. Tracking human migrations by the analysis of the distribution of HLA alleles, lineages and haplotypes in closed and open populations. *Philos Trans R Soc B* **367**:820–829.
105 Olson MV. 2012. Human genetic individuality. *Annu Rev Genomics Hum Genet* **13**:1–27.
106 Leffler EM et al. 2013. Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science* **339**:1578–1582.
107 Fumagalli M et al. 2011. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genet* **7**:e1002355.
108 Novembre J, Han E. 2012. Human population structure and the adaptive response to pathogen-induced selection pressures. *Philos Trans R Soc B* **367**:878–886.
109 Haygood R et al. 2010. Contrasts between adaptive coding and noncoding changes during human evolution. *Proc Natl Acad Sci USA* **107**:7853–7857.
110 Hancock AM et al. 2010. Human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. *Proc Natl Acad Sci USA* **107** (Suppl 2): 8924–8930.
111 Hancock AM et al. 2011. Adaptations to climate-mediated selective pressure in humans. *PLoS Genet* **7**:e1001375.
112 Tennessen JA, Akey JM. 2011. Parallel adaptive divergence among geographically diverse human populations. *PLoS Genet* **7**:e1002127.
113 Ohashi J et al. 2011. The impact of natural selection on an *ABCC11* SNP determining earwax type. *Mol Biol Evol* **28**:849–857.
114 Li H et al. 2011. Diversification of the *ADH1B* gene during expansion of modern humans. *Ann Hum Genet* **75**:497–507.
115 O'Bleness M et al. 2012. Evolution of genetic and genomic features unique to the human lineage. *Nat Rev Genet* **13**:853–866.

116 Iskow RC et al. 2012. Exploring the role of copy number variants in human adaptation. *Trends Genet* **28**:245–257.
117 Raj T et al. 2012. Alzheimer disease susceptibility loci: evidence for a protein network under natural selection. *Am J Hum Genet* **90**:720–726.
118 Hünemeier T et al. 2012. Evolutionary responses to a constructed niche: ancient Mesoamericans as a model of gene-culture coevolution. *PLoS One* **7**:e38862.
119 Arbiza L et al. 2013. Genome-wide inference of natural selection on human transcription factor binding sites. *Nat Genet* **45**:723–729.
120 Kamberov YG et al. 2013. Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell* **152**:691–702.
121 Xu K et al. 2013. The evolution of lineage-specific clusters of single nucleotide substitutions in the human genome. *Mol Phylogenet Evol* **69**:276–285.
122 Liu X et al. 2013. Detecting and characterizing signatures of positive selection in global populations. *Am J Hum Genet* **92**:866–881.
123 Kuhn A et al. 2014. Linkage disequilibrium and signatures of positive selection around LINE-1 retrotransposons in the human genome. *Proc Natl Acad Sci USA* **111**:8131–8136.
124 Hawks J et al. 2007. Recent acceleration of human adaptive evolution. *Proc Natl Acad Sci USA* **104**:20753–20758.
125 Varki A et al. 2008. Explaining human uniqueness: genome interactions with environment, behaviour and culture. *Nat Rev Genet* **9**:749–763.
126 Chiaroni J et al. 2009. Y chromosome diversity, human expansion, drift, and cultural evolution. *Proc Natl Acad Sci USA* **106**:20174–20179.
127 Prabhakar S et al. 2006. Accelerated evolution of conserved noncoding sequences in humans. *Science* **314**:786.
128 Somel M et al. 2009. Transcriptional neoteny in the human brain. *Proc Natl Acad Sci USA* **106**:5743–5748.

CHAPTER 3
Population structure

We need to reach that happy stage of our development when differences and diversity are not seen as sources of division and distrust, but of strength and inspiration.

—*Josefa Iloilo* [1]

SUMMARY

A number of markers capable of assessing population structure are currently employed, including morphological traits, behavioral characteristics, and molecular variants. Although all of these marker systems are still in use, advances in biochemical genetics and subsequent discoveries in the field of molecular biology have provided students of human evolution with molecular tools with excellent phylogenetic resolution. The early investigations utilizing molecular markers were based on protein and isozyme/allozyme polymorphisms detected subsequent to electrophoretic separation or antigen–antibody reactions. Since the 1970s, a series of techniques have been developed that allow us to examine genetic variability directly at the DNA level. Procedures such as the polymerase chain reaction (PCR) and large-scale DNA sequencing have ushered a new era of assessment of diversity by mass producing specific sites within the DNA and detecting the exact locations of mutations, respectively. At present, scientists are capable of whole-genome sequencing at the population level. To analyze the huge amount of data generated at the bench work, population geneticists rely on sophisticated bioinformatic tools that annotate, store, and compare the DNA sequences. Students of human evolution benefit from these technological advances by feeding massive amounts of DNA data into computer programs based on population genetic principles. Some of these genetic paradigms have origins traceable to Mendel's laws. For example, the Hardy–Weinberg equilibrium formulation dates back to the beginning of the 20th century. A multitude of programs with algorithms designed to ascertain phylogenetic relationships among human populations are now available. In addition to determining population relationships among extant groups, the different parameters of genetic variability have been applied to ancient DNA and archaic human groups such as Neanderthals and Denisovans. Furthermore, these same tools are routinely used to investigate the internal organization of populations such as subpopulation structure that may signal admixture of different groups and/or nonrandom mating. Knowledge of genetic differences among populations as well as internal differentiation within populations is paramount to a number of related disciplines such as forensic DNA fingerprinting and epidemiology since the calculation of inclusion probabilities in criminal cases, for example, and cause-and-effect relationships involving diseases, respectively, assume genetic homogeneity within groups.

DNA-based marker systems

At present, the majority of genetic analyses on populations are performed by scoring variability directly on the DNA. Although diversity at the protein level and morphometric parameters are still in use as markers, currently, a number of molecular techniques that prove the DNA directly allow investigators to examine a large number of individuals in a short period of time. These methodologies are employed to assess diversity in different parts of the genome. For example, mitochondrial DNA (mtDNA) is routinely examined to assess maternally derived genetic diversity [2]. The mitochondrion is an organelle found in multiple copies inside eukaryotic cells. Mitochondria are thought to derive from a symbiotic relationship that has evolved for billions of years. Today, it is accepted that mitochondria started as

Genomes, Evolution, and Culture: Past, Present, and Future of Humankind, First Edition.
Rene J. Herrera, Ralph Garcia-Bertrand, and Francisco M. Salzano.

© 2016 John Wiley & Sons, Ltd. Published 2016 by John Wiley & Sons, Ltd.

prokaryotic parasites that developed gradually into a mutualistic existence with eukaryotic cells. After a long evolutionary process of interdependence, most of the genes essential for independent life were transferred to the nuclear genome or lost, making mitochondrial metabolic self-sufficiency impossible. Of course, eukaryotic cells cannot live either without this organelle essential for energy production. Thus their existence has become completely interdependent. A number of characteristics such as its double membrane, circular haploid genome, and no histone proteins bound to DNA, point to the symbiotic origin of mitochondria from an ancient parasitic or mutualistic existence. In some metabolically active tissues, such as the heart and liver, their numbers are in the thousands. In humans, the size of the mitochondrial genome is about 16,500 base pairs. Since the mitochondria are always inherited from mothers and they are not involved in DNA recombination, their DNA is used to follow maternal lineage as a large complex or haplotype. A haplotype is a portion of DNA that tends to stay together generation after generation due to lack of recombination. The mitochondria of sperm normally never reach the inside of the ova and are left out of the embryo. That is why our fathers are not contributors to our mtDNA genome. There are several characteristics that make mtDNA an excellent marker system. For example, their large numbers inside each cell and small circular DNA allow scientists to study mtDNA even in ancient samples where the nuclear DNA is too degraded or chemically modified to be useful.

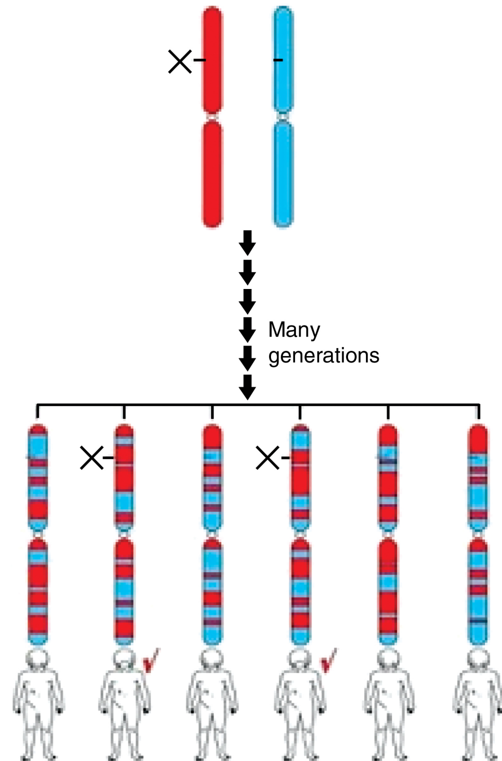


Figure 3.1 Consequence of DNA recombination. Notice how parental DNA recombines generating chromosomes with different amounts and locations of maternal and paternal segments. (See the Color Plates section.)

The nuclear genome, on the other hand, is diploid, composed of a maternal and a paternal set of linear chromosomes that exchange DNA or recombine during meiosis. Except for the nonrecombining portion of the Y chromosome (NRY), the maternal and paternal nuclear DNAs are capable of exchanging corresponding sequences. This process of DNA recombination has fundamental consequences in the processes that govern the evolution of eukaryotes as well as in the methodologies employed for the analysis of nuclear inheritance. Since the DNA of most of the nuclear genome engages in this process of crossing over, the genes involved assort independently as cells undergo meiosis in the process of gametogenesis (egg and sperm formation). The consequences of this reshuffling are chimeric linear chromosomes in subsequent generations made up of intermittent maternal and paternal DNA segments (Figure 3.1). In other words, instead of DNA that tends to be inherited as a whole (mtDNA and NRY), as units or haplotypes, most of the nuclear genes move from generation to generation independently from each other. If genes are located on different chromosome pairs or distant from each other on the same chromosome (about 50 million bases or more), they tend to assort independently from each other. On the other hand, if genes are closer to each other (less than 50 million bases) within the same linear DNA, they tend to segregate independently from each other as cells undergo meiosis due to limited recombination [3]. Most of the time the frequency of separation or recombination involving nearby genes is directly proportional to the distance between them; however, recombination is not entirely random (see Chapter 5 on linkage disequilibrium and recombination hotspots). In other words, the greater the distance separating the genes, the greater the frequency of recombination events between the two. In fact, early gene mapping experiments dating back to the turn of the 20th century were performed by ascertaining the prevalence of the exchanges between homologous

chromosomes resulting from recombination. It was assumed that the greater the percentage of crossing over, the greater the distance between the genes in question and this proved true for most genes.

As a result of the differences between uniparental (mtDNA and Y chromosome) and nuclear inheritance, the analytical methodologies used to study genetic diversity differ. Algorithms dependent on the principles of independent genetic assortment of genes would not be appropriate for the analyses of Y-specific and mitochondrial inheritance since these gene complexes tend to move together as blocks from generation to generation. As a result, some analytical programs (e.g., Rst distances for short tandem repeats or STRs on the Y chromosome; see Box 3.1) take into consideration the fact that some sequences are inherited as part of a complex that rarely recombines. Such programs account, for example, for the number of mutations within haplotypes (e.g., two mutations are not equivalent to seven mutations) and the number of repeats among variants (e.g., a locus possessing three repeats is not equivalent to a locus exhibiting nine repeats). In other words, all available software programs have limitations and the variables and parameters in each need to be adjusted to cater to the marker system employed.

Another variable to consider when studying genetic diversity is that the rate of mutation (μ) is not uniform for all types of DNA. For example, the mutation rate of human mitochondrial DNA has been estimated to be about 3×10^{-5} per base per 20-year generation. This relatively high rate of mutation is partially due to the infidelity of the mitochondrial DNA polymerase in duplicating DNA. The mutation rate that affects the number of repeats in short tandem repeat DNA sequences (nuclear genome) ranges between

Box 3.1 Some major analytical methods employed in population genetic studies.

Methodology	Purpose/function	Type of data	References
Neighbor joining (NJ)	Phylogenetic comparisons based on a bottom-up (agglomerative) clustering method	Phylogenetic tree	Saitou and Nei [4]
Principal component analysis (PCA)	Phylogenetic comparisons based on orthogonal transformation	Plot	Pearson [5]
Multidimensional scaling (MDS)	Phylogenetic comparisons based on dissimilarity matrix and Euclidean distances	Plot	Cox and Cox [6]
Hardy–Weinberg equilibrium (HWE)	Impact of evolutionary forces in allele and genotype frequencies	Allele and genotype frequency changes	Hardy [7]; Weinberg [8]; Edwards [9]
Genetic diversity (GD)	Genetic heterogeneity/inbreeding	Proportion of polymorphic or heterozygous loci or alleles per locus	Nei [10]
Heterozygosity	Genetic heterogeneity/inbreeding	Proportion of heterozygotes at a locus	Allendorf [11]
Analysis of molecular variance (AMOVA)	Correlations	Estimation of population differentiation directly from molecular data	Excoffier et al. [12]
Estimation of coalescent times	Age of mutations	Traces alleles to a single ancestral copy	Kingman [13]
Percent unique haplotypes	Haplotype diversity	Frequency of unique haplotypes	
Network analysis	Phylogenetic analysis of haplotypes	Haplogroups' relationships	Cooper et al. [14]
Contour analysis	Diversity/frequency distribution in space	Geographical representation of alleles	Chiaroni et al. [15]
Gst	Genetic distance	For multiple alleles	Nei [10]
Fst	Genetic distance	For biallelic markers	Wright [16]
Bootstrap	Random sampling with replacement	Statistical significance	Efron [17]

Box 3.1 (*Continued*)

Methodology	Purpose/function	Type of data	References
Gene counting	Allelic frequency	Allelic abundance	Ceppellini et al. [18]
Genotypic frequency	Predicting genotypes	Frequency of combinations of alleles	Shields et al. [19]
Haplotype diversity	Uniqueness of particular haplotype	Genetic heterogeneity/inbreeding	Nei and Tajima [20]
Haplotype discrimination capacity	Frequency of haplotypes	Genetic uniqueness	Chang et al. [21]
R^2 test	Linkage disequilibrium	Nonrandom association of alleles at two or more loci	Slatkin and Excoffier [22]; Reich [23]
μ value	Probability of a mutation/generation	Mutation rate	Excoffier and Slatkin [24]
Maximum likelihood	Estimation of parameters of a statistical model	Phylogenetic comparisons	Excoffier and Slatkin [24]
Demographic expansion	Finite-sites model with heterogeneity of mutation rates	Population expansion	Kingman [25]
Ewens–Watterson test	Based on the comparison of homozygosity	Assess selection pressure	Ewens [26]; Slatkin and Excoffier [22]
Shared haplotypes	Percentage of haplotypes in common	Phylogenetic comparisons	Excoffier et al. [27]
F-statistics	Statistically expected level of heterozygosity	Inbreeding test	Wright [28]
Exact test	Statistical significance in which all assumptions are met	Assess whether data fit the results expected from theory	Raymond and Rousset [29]
Assignment test	Likelihood-based ancestry	Assigning genotypes to populations	Paetkau et al. [30]
Mantel test	Correlation between two matrices	Genetic distance correlations	Smouse [31]
Chi-square test	Statistical significance	Assess whether data fit the results expected from theory	Roff and Bentzen [32]
Bonferroni correction	Family-wise error rate	Correct errors resulting from multiple comparisons	Dunn [33]
Watterson estimator	Based on coalescent theory	Population mutation rate	Watterson [34]

10^{-3} and 10^{-4}. Furthermore, some hypermutable, medically important trinucleotide repeats are capable of expanding their number of reiterated units within a few generations. STR mutations affecting the number of repeat units generally lead to increments in the number of reiterated sequences, at a rate depending on the actual locus in question. These frequencies are much higher than the point mutations occurring in the nuclear genome that generate substitution, deletion, and addition of one or a few nucleotides per site per generation on the order of 10^{-6} to 10^{-8}. Mutations in the form of insertion of repetitive elements, such as the members of the *Alu* family, insert at the rate of one *de novo* insertion approximately every 20 births [35].

SNPs, STRs, and Indels as DNA markers

SNPs

A number of DNA-based marker systems are routinely employed. DNA marker systems, in general, can be grouped into single-nucleotide polymorphisms (SNPs), short tandem repeats (STRs), and insertion/deletion polymorphisms (indels). Some SNPs are rather common in the genome and are found at polymorphic frequencies (>1%). Some are rare making them difficult to identify and confirm, although it is thought that they are abundant in any given individual. In general, polymorphic

markers such as SNPs are particularly useful in population genetic studies since their frequencies and distribution allow for phylogenetic assessment. These polymorphisms often take the form of frequency and diversity clines within geographical regions indicating gradual changes. Yet, at times SNPs exist at low frequencies present in one or a few populations as private (in one or a few populations) variants.

It is interesting that most SNPs are located in unique sequences [36]. SNPs are typically visualized with a number of techniques including direct DNA or RNA sequencing, restriction enzyme digestions (restriction fragment length polymorphisms, or RFLPs), and high-resolution HPLC. SNPs can be characterized based on a number of criteria. For example, some SNPs are located in unique DNA sequences of known function at times affecting gene regulation and expression. These SNPs may be located within regulatory elements such as promoters, enhancers, and silencers and may impact transcription initiation, intron splicing, post-transcription modification, and mRNA stability. SNPs within protein-coding sequences could lead to alteration in protein structure/function and/or provide for evolutionary change under selection pressure. Yet, it is thought that the majority of SNPs are functionally silent and therefore tend to be excellent as markers for population genetic studies (see Chapter 5 for more discussion on SNPs and how haplotypes are created from SNPs).

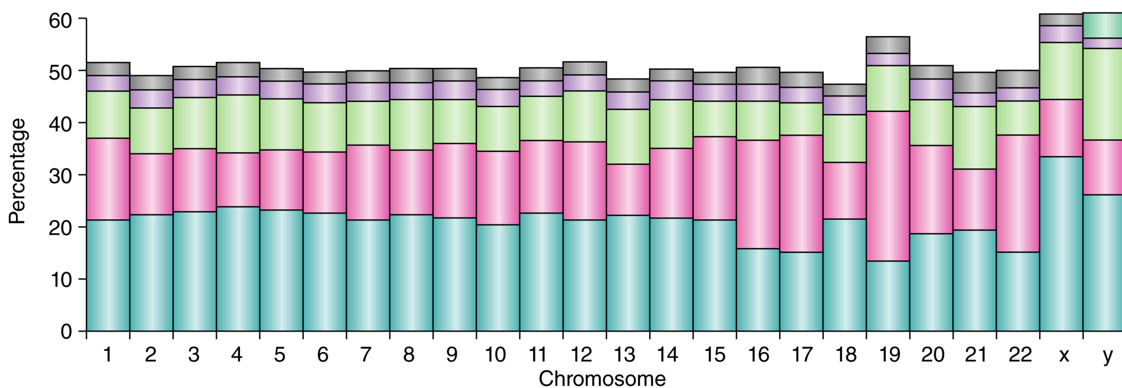
STRs

STR sequences, the other major marker system, represent one type of repetitive DNA in our genome. The human genome is made up of about 50% reiterated DNA. Figure 3.2 lists the various classes of iterated DNA

(a)

Repeat class	Repeat type	Number (hg19)	Cvg	Length (bp)
Minisatellite, microsatellite, or satellite	Tandem	426,918	3%	2–100
SINE	Interspersed	1,797,575	15%	100–300
DNA transposon	Interspersed	463,776	3%	200–2000
LTR retrotransposon	Interspersed	718,125	9%	200–5000
LINE	Interspersed	1,506,845	21%	500–8000
rDNA (16S,18S,5.8S, and 28S)	Tandem	698	0.01%	2000–43,000
Segmental duplications and other classes	Tandem or Interspersed	2.270	0.20%	1000–100,000

(b)



Nature Reviews | Genetics

Figure 3.2 Families of repetitive elements. Cvg indicates percentage of the genome occupied by repeat class. (Reproduced from Treangen & Salzberg 2012 [37] with permission of Macmillan Publishers Limited.) (See the Color Plates section.)

sequences present in humans. Figure 3.2 also provides information on the size of the repeated unit, the copy number for each family, and the percentage of the genome occupied by the redundant DNA types. The bar graph in Figure 3.2b illustrates the color-coded proportions of each repetitive class found in each human chromosome. STR polymorphisms are routinely employed in population genetic studies as well as forensic and medical cases. These simple sequences are made up of two to six nucleotide units reiterated a few to hundred of times, each copy in the same orientation in tandem. The nonrecombining region of the Y chromosome in humans is particularly rich in STRs. A similar class of tandem repeats is referred to as variable number of tandem repeats or VNTR. These reiterated sequences are a bit larger than STRs, each repeat being 15 to a few dozen nucleotides. They share the same head to tail direct repeat organization and their repetitiveness with the STRs as well as the mechanism of mutation that generates the various numbers of tandem repeats, DNA replication slippage (see Chapter 4 under mutation types). Due to the simple nature of these reiterated sequences, it is likely that they do not perform functions except possibly as spacer DNA keeping, for example, proper distance among regulatory elements and structural genes. Therefore, most STR and VNTR loci are thought to be selectively neutral in terms of nucleotide substitutions. These tandem repeated sequences in fact provide two marker systems in one since in addition to the genetic diversity in the form of various numbers of repeat units, internal point mutations within the repeats can be scored. This dual and superimposed genetic variability provides for greater resolution in population genetic studies and identity assessments.

In forensic scenarios, these hypervariable tandem repeat loci are ideal as an investigative tool since highly sensitive diagnostic, PCR-based kits are available to detect nanogram quantities of DNA left as "touch DNA," genetic material left behind on a surface, for example, from the touch of a hand. Currently, this technology is so sensitive that picogram quantities of DNA, at a crime scene, are enough to generate a complete or partial STR profile of individuals who have visited the crime scene. These marker systems are so precise that we are rapidly approaching the time in which perpetrators will wonder not whether they are going to be identified but when. Also, some of these STR loci are so variable that they exhibit alleles (variants in numbers of reiterated units) in excess of 80. This diversity allows for easier discriminating among individuals since with a few loci individuals could be rendered genetically unique. At present, STR markers are capable of detecting and typing DNA left on a surface such as the trigger of a gun and the system is being optimized to co-amplify dozens of loci in a single PCR (multiplex PCR). Thus, even if only partial profiles are obtained, sufficient information may be generated to identify putative perpetrators. In addition, the application of these marker systems in forensic studies is facilitated by the availability of local, national, and international databases of previous offenders allowing for the apprehension of suspects during the identification phase of the investigation. Cold cases (dormant crimes that have remained unsolved for years) can be activated and successfully resolved. Using this technology, individuals can be exonerated or convicted.

Medically, STR markers have been particularly useful, for example, in detecting any residual malignant bone marrow tissue subsequent to chemotherapy, and before the introduction of normal cells from a donor. In practice, chemotherapy is provided to destroy all endogenous cancerous cells and typing with STR markers before and after treatment allows clinicians to verify that all of the malignant tissue has been destroyed prior to introducing normal cells. In addition, a growing number of medically important diseases have their basis in the rapid amplification of short tandem repeats that occur in a few human generations. These dramatic increments in the number of repeats, usually trinucleotide sequences, seem to be triggered by specific point mutations within the reiterated sequences. These trinucleotide STR disorders generally exhibit a phenomenon known as genetic anticipation, where the severity of the conditions can increase with each successive generation that inherits them. Some well-known examples of this type of human maladies include fragile X syndrome, Huntington's disease, and dentatorubral-pallidoluysian atrophy. The progression of these illnesses across generations is routinely monitored by STR typing technology.

STR loci also have found their niche in basic population genetic research. Simple tandem repeating DNA markers are hypervariable with the different alleles varying in the number of repeating elements. This phenomenon is likely the result of the lack of fidelity of the DNA polymerase resulting in the occurrence of DNA replication slippage as these highly reiterated sequences are being copied. At mutation rates on the order of 10^{-3} to 10^{-4}, these highly polymorphic loci provide the resolution to examine microevolutionary and recent events in human

evolution [38]. For example, these selectively neutral repetitive markers have shown to have the sensitivity to provide phylogenetic and migrational information on recent human diasporas including events within historical times. Specifically, the combination of SNP and STR markers is currently shedding light on issues such as the Bantu dispersal, the Austronesian expansion, the settlement of America by Paleolithic humans from Asia, and the different independent agricultural/domestication revolutions worldwide, among others. Some of these migrations will be discussed in more detail in Chapter 6.

Indels

Indels represent a broad class of biallelic polymorphisms characterized by the presence (one allele) or absence (the alternative allele) of a sequence of DNA. In its most basic form, segments of DNA from random locations in the genome are missing in some individuals of a population, at times creating polymorphisms. It is likely that these deleted sequences are not vital for the survival of the individual or are marginally under selection pressure. Although any deleted DNA sequence could be used as a genetic marker, individual members of specific reiterated DNA classes (see Figure 3.2 for listing) have been useful in phylogenetic studies. For example, recently inserted repetitive members of the SINEs (short interspersed nuclear elements) [39], LINEs (long interspersed nuclear elements) [40], and retroviral [41] families exist in polymorphic state in human populations. It is likely that many of these individual members represent recent insertional events and thus they are not fixed in the population.

As markers, indels such as *Alu*s provide for unequivocal identification of the ancestral and derived states since the probability of an independent insertion into exactly the same site on the DNA sequence is highly unlikely and there are no known mechanisms for the precise removal of an insertion without leaving a residual sequence behind. In other words, SINE/LINE indel polymorphisms are essentially homoplasy-free markers that can be used to study human populations. This characteristic of indels allows for polarity of the lack of insertion as the ancestral state. This knowledge is highly valuable in phylogenetic studies since it provides for the directionality of the mutation.

Independent of their value as markers, *Alu* insertion elements represent an enigma in terms of evolutionary biology. It turns out that the *Alu* family started as a duplication of the gene that codes for the small 7SL RNA moiety of the signal recognition particle of the rough endoplasmic reticulum [42]. Since their origin, *Alus* have been amplifying and dispersing throughout the genome of primates. At novel insertion sites, *Alus* may coevolve with nearby flanking sequences. This possibility is suggested by the eclectic function of some *Alus*. In other words, since specific *Alu* sequences are known to perform a number of functions such as controlling elements and coding sequences, it is possible that *Alu*s are capable of coevolving with preexisting genes generating novel or modifying functions [43].

Ancient retroviral insertions, also known as endogenous retroviruses (ERVs), are reiterated sequences of particular interest to students of virology and human genome evolution. About 100,000 of these sequences or 8% of the human genome is made up of this type of repeated DNA. As with SINEs and LINEs, these ERV insertions facilitate deletions and duplications of host DNA providing for diversity and raw material for evolution. The diversity provided by ERVs has been employed in phylogenetic studies [41].

Since most of the ERVs represent ancient insertion events, they are fixed in our genome. Yet, one subfamily, HERV-K (HML2), which constitutes less than 1% of the human ERVs, is thought to have integrated recently within the past few hundred thousand years and some members are still polymorphic. In other words, particular ERVs exist as polymorphic retroviral insertions (PRVIs) in three possible genotypic states in human populations: individuals homozygous for the insertion, homozygous for the lack of insertion, or heterozygous. Since individuals and populations differ in the presence and absence, and frequency, respectively, of these polymorphic ERV insertions, PRVI loci have been used for identity and population genetic studies [41]. Specifically, two members of the human-specific subfamily HERV-K (HML2), HERV-K106 and HERV-K116, were active integrating in the last 800,000 years and HERV-K106 may have infected modern humans as recent as 150,000 years ago, just prior to the out of Africa migration. It is not clear whether these recent integrations were endogenous retroposition events (secondary insertions within individuals) or external viruses infecting individuals.

In light of this discussion, we cannot help wonder whether retroviral epidemics in the past have initiated pandemic infections leading to striking and sudden reduction in population sizes. It is interesting to contemplate the possibility that some ERVs originated as pandemic retroviral infections that may have driven

humanity to the brink of extinction. Could retroviral infections have been responsible for some of the bottleneck episodes that have taken place in recent human evolution? And, further, could HIV represent the present-day version of retroviral epidemics that may evolve to a new ERV?

Population genetic tools for analyzing population structure

To analyze the voluminous amount of data generated at the bench, population geneticists rely on bioinformatic tools that annotate, store, and examine the DNA sequences [44]. Students of human evolution benefit from these technological advances by feeding massive amounts of DNA variability data into computer programs based on population genetic principles. Some of these genetic paradigms have origins traceable to Mendel's laws. For example, the Hardy–Weinberg equilibrium formulation dates back to the beginning of the 20th century. A number of programs with algorithms designed to ascertain phylogenetic relationships among human populations (interpopulation methods) are now available, while other software programs are designed to mine the internal genetic structure of populations (intrapopulation methods). For example, bioinformatic tools are routinely used to investigate the internal organization of populations such as subpopulation structure that could signal admixture of different groups and/or nonrandom mating. In addition to determining population relationships among extant groups, the different parameters of genetic variability have been applied to ancient and archaic human groups such as the Neanderthal and Denisova [45].

Bioinformatic tools are designed to assess a number of parameters. Independent of whether the trait in question is morphological, physiological, or molecular, variants are counted and scored usually utilizing the gene counting method [46]. Simply put, in this procedure the number of variants of a trait are counted and the frequency of each is ascertained. Variants in genetic studies take the form of alleles, the alternative forms of genes. Allelic measurement is the basic raw data that feed into a number of phylogenetic algorithms employed in population genetics. Allelic frequencies can be utilized individually when genes assort independently during meiosis. In other words, when genes are located on different chromosome pairs (e.g., chromosomes 1 and 7), they segregate independently from each other, generation after generation, and their allelic frequencies should be utilized individually in the various computational analyses. Conversely, allelic frequencies of genes located on nonrecombinant DNA, including most of the Y chromosome (NRY), mtDNA, and chloroplast DNA (ctDNA or cpDNA) in plants or genes in close proximity to each other on the same chromosome (approximately within 50 centimorgan units), cannot be analyzed as independently segregating units. In such situations, genetic comparisons may be performed at the level of haplotypes, a haplotype being a group of genes that tend to remain together from generation to generation. In other words, since genes within haplotypes do not assort independently during meiosis and are linked in a complex, their frequencies may be analyzed at the haplotype level instead of individual allelic proportions. Some phylogenetic software programs such as multidimensional scaling (MDS) are capable of taking into consideration the nonrecombining nature of certain complex of loci utilizing algorithms in which the genetic data are in the form of haplotypes. In humans, DNA on the NRY and mtDNA are routinely utilized as uniparental markers of inheritance.

Box 3.1 illustrates some of the routinely employed bioinformatic programs in population genetics. Some of the software programs are basic and designed to ascertain mutation rates (e.g., μ value), frequency, and diversity of variants or haplotypes (e g., network analysis). Also, a number of these programs aim to assess phylogenetic relationships (e.g., neighbor joining, principal component analysis, and multidimensional scaling) as well as genetic similarities, differences, and distances (e.g., Mantel test) among populations. Other more specialized tests provide for correlations between genetic diversity and a number of parameters such as geographical space (e.g., AMOVA and contour analysis). Age of mutations and coalescent times (time of separation of two or more lineages) are also routinely computed. In addition, programs for evaluating Hardy–Weinberg equilibrium expectations (e.g., exact test), linkage disequilibrium (e.g., R^2 test), inbreeding (e.g., F_{is}), and selection pressure (e.g., Ewens–Watterson test) are available. Statistical significance is usually determined using algorithms such as bootstrap, χ^2, and Bonferroni analyses.

To illustrate the type of data output routinely generated by some of the above-mentioned analyses, a number of specific applications are illustrated below (also see Appendix A). For example, phylogenetic routines are usually performed to assess relationships among populations. The results of tests are at times visualized in the form of phylogenetic trees. Figure 3.3 represents an

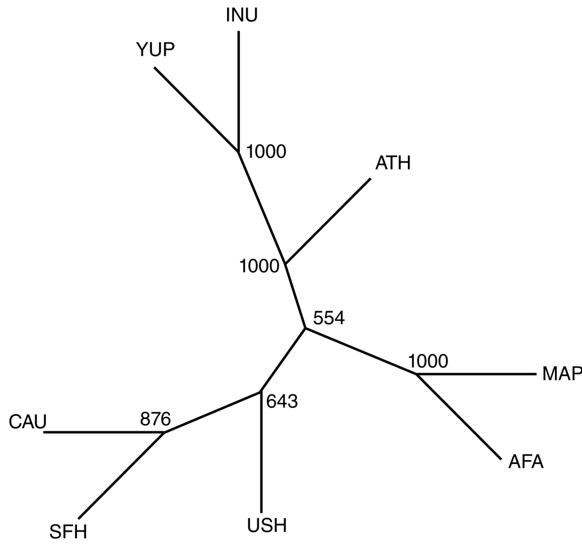


Figure 3.3 Phylogenetic tree with bootstrap values. The threeletter acronyms indicate different populations. ATH = Athabascans; INU = Inupiat; YUP = Yupik; AFA = African Americans; CAU = Caucasians; MAP = Maputo; SFH = Southeast Florida Hispanics; and USH = USA-wide Hispanics. Numbers at bifurcations represent bootstrap values.

example of how populations are arranged according to their genetic similarities in a phylogenetic tree. In this type of diagram, the lengths of the lines connecting populations are directly proportional to genetic distances. Populations that bifurcate from a single stem are considered sister groups and reflect close affinities. A number of genetic distance formulations such as Nei's (see Box 3.2) are commonly employed to generate these dendrograms. Assessment of statistical significance is usually determined using bootstrap algorithms (see Box 3.1) that compute the percentage of time the same bifurcations are produced as independent trees are generated. Routinely about 1000 permutations or reiterations (trees) are analyzed in bootstrapping and nodes with percentages of 90% and above are considered statistically significant, although lower percentages are often seen. Notice the bootstrap values at the bifurcation of the phylogenetic tree in Figure 3.3. The main criticism with this type of analysis is that populations are forced to connect with lines in making the branches of the tree. This shortcoming has motivated a number of investigators to report their data in the form of two- or three-dimensional plots that do not assume explicit connections among populations. For example, principal component analysis and multidimensional scaling (Figure 3.4) are software programs employed to represent phylogenetic relationships within a plot. The closer the distance separating populations on the graph, the greater their genetic affinities. Principal component analysis (PCA) outputs provide the percentage of the total genetic diversity used to partition populations along the axes (i.e., *x*, *y*, and *z*) on the projection. Combined

Box 3.2 Some commonly employed genetic distances in population genetic studies.

Distance	Characteristics	Name of program	References
Cavalli-Sforza's chord measure	Genetic drift only	PHYLIP/GENDIST	Cavalli-Sforza and Edwards [48]
Nei's standard genetic distance	Driven by mutation and genetic drift	PHYLIP/GENDIST	Nei [49]
Edwards' angular distance	Angular distance	POPPR	Edwards [50]
Reynolds, Weir, and Cockerham's genetic distance	Coancestrality coefficient/genetic drift only	PHYLIP/GENDIST	Reynolds et al. [51]
Rogers' distance	Classical Euclidean/fixation index/no mutation	TFPGA	Rogers [52]
Prevosti's distance	Absolute genetic distance	POPPR	Prevosti et al. [53]
Nei's minimum genetic distance	Driven by mutation and genetic drift	TFPGA	Nei and Roychoudhury [54]
Nei's D_A distance	Driven by mutation and genetic drift/ microsatellite DNA data	DISPAN	Nei et al. [55]
Goldstein's distance	Stepwise mutation for microsatellites		Cooper et al. [56]
Bruvo's distance	Stepwise mutation for microsatellites	POPPR	Bruvo et al. [57]

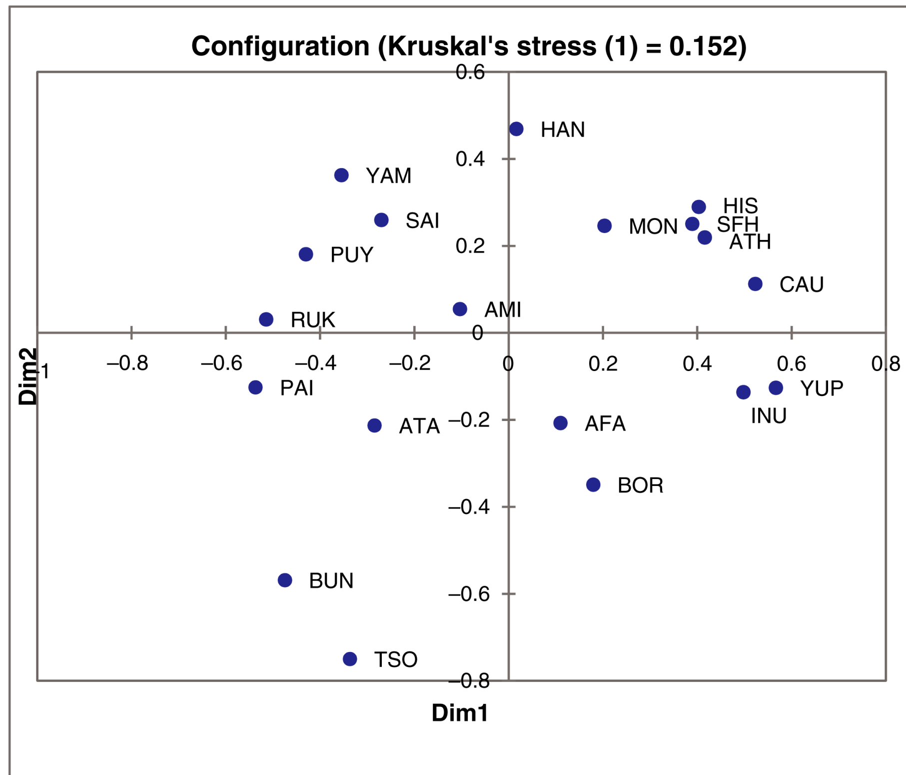


Figure 3.4 MDS plot of Taiwanese aborigines and reference populations: Ami (AMI), Atayal (ATA), Bunun (BUN), Paiwan (PAI), Puyuma (PUY), Rukai (RUK), Saisiyat (SAI), Tsou (TSO), Yami (YAM), Mongolian (MON), Hans in China (HAN), Polynesian Society Islands (BOR), Europe Caucasians (CAU), North American Athabascans (ATH), Inupiat (INU), Yupik (YUP), U.S. Hispanics (HIS), Southeast Florida Hispanics (SFH), and African American (AFA). (Reproduced from Zeng et al. 2014, [47] with permission of Elsevier.)

percentages of 50 or greater from all the dimensions are indicative of reliable data. In the case of MDS analyses, dependable results are assessed by the stress value level of the run. Specifically, the lower the stress value, the greater the congruency of the data since the genetic distances (e.g., pairwise Gst distances) do not need to be too strained to fit the distribution of populations in the plot. Projections with stress values of 0.1 or less are generally acceptable.

AMOVA is a frequently used program to assess the statistical significance of geographical and linguistic distributions as a function of genetic diversity. For example, let us consider a scenario in which there is an interest in ascertaining whether the geographical locations of certain groups of populations are congruent with the genetic frequency or diversity data. The AMOVA test is designed to determine whether genetic differences among populations correlate with the geographical distances separating them and whether the parallelism is statistically significant. The same test can be employed, for example, with correlations between linguistic and genetic characteristics and for any other type of data in which the strength of correlations needs to be assessed. In performing AMOVA, populations are partitioned according to, for example, geographic groups. The expectation is that populations that belong to the same geographic category would be more similar compared with populations that do not. Contour maps, on the other hand, are generated with programs that provide for the visualization of genetic distribution in the context of geographical maps. Figure 3.5a and b represents typical contour outputs in which the frequencies and abundance, respectively, of a genetic marker are plotted on geographical maps. These maps are particularly useful in representing gradual or clinal differences as a function of geographic distances.

In network analyses, on the other hand, the genetic relationships of individuals from a number of populations

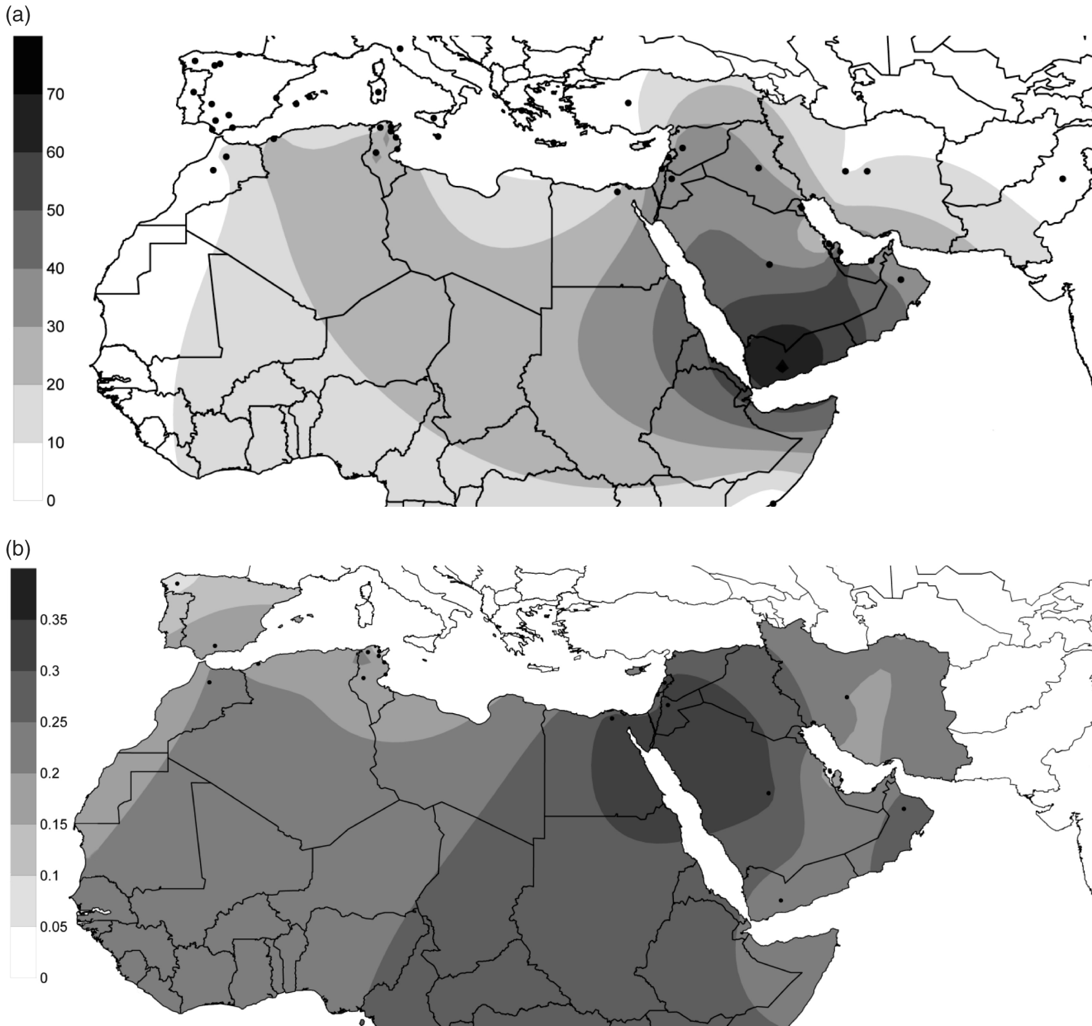


Figure 3.5 Contour plots illustrating the (a) frequency and (b) diversity distribution of individuals under Y-specific haplogroup M267 in northern Africa and neighboring regions. (Reproduced from Regueiro et al. [58] with permission of Elsevier.)

are examined (Figure 3.6). The test is employed to assess genetic relationships among single people and how they partition in relation to other samples from the same or different populations. In practice, networks are well suited for uniparental marker systems such as Y chromosomal and mtDNA haplotypes. The samples are represented as circles and color coded according to the population they belong to (each population takes a different color or shade) and the size of the circles reflects the number of persons exhibiting a given haplotype. Since the samples are color coded according to populations, the distribution of samples within the network also provides information on the phylogenetic relationships among the groups. The smallest circles indicate singletons or unique haplotypes. Circles with two or more colors indicate sharing of haplogroups among individuals from different populations. The program generates lines connecting all the people creating a sort of network, the characteristic from which the name derives. These lines connect individuals that are most closely related and the lengths of the connecting lines are directly proportional to the number of mutational steps. Therefore, the degree of genetic differences among people can be ascertained by counting the number of mutational steps separating the samples.

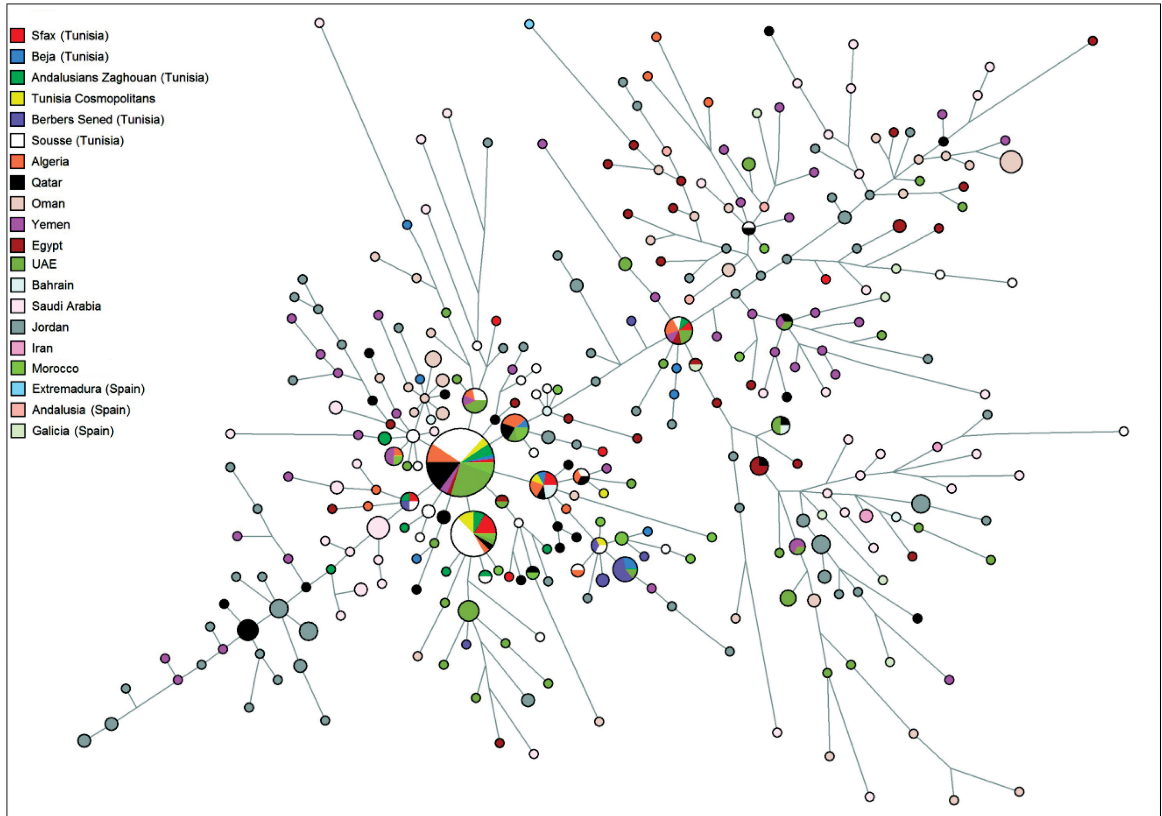


Figure 3.6 Network analysis of individuals under Y-specific haplogroup M267. Size of circles is proportional to number of individuals. (Reproduced from Regueiro et al. 2015 [58] with permission of Elsevier.) (See the Color Plates section.)

Networks are particularly informative when they exhibit a "spider-type" topology as seen in the branches emanating from the large multicolor circle (node) in Figure 3.6. This morphology may provide information about the location and/or populations where a mutation originated or a regional founder effect was followed by diversification. The branches represent lineages of sequentially related individuals with the samples at the terminals of each offshoot being the most genetically differentiated individuals. Networks that show partitioning of individuals from specific populations into particular branches suggest minimal gene flow among the populations or groups, while a lack of compartmentalization is indicative of gene flow among the populations in the network.

Structure analysis is an increasingly used method to study population diversity. It is based on Bayesian statistics, algorithms that regard parameters of populations as random variables having known probability distributions. Structure analysis is designed to examine the genetic similarities and differences among populations and individuals (Figure 3.7). Structure analysis is particularly useful as a first-step test to examine population structure prior to further more sophisticated genetic analyses. Structure analysis, for example, provides information on the genetic contribution to populations from source or ancestral groups, and addresses the common ancestry of individuals and populations. It is often utilized to infer the origins of people and populations of unknown admixed genetic backgrounds. Specifically, this test is commonly employed in genome-wide studies to ascertain the uniqueness of populations as well as their degree of admixture. This is done utilizing reference populations for comparison. The samples are represented in a bar graph. Each person appears as a bar with colors representing its various genetic components and their relative proportions. In this type of analysis, the greater the number of genetic markers investigated, the more comprehensive and thorough the results are expected to

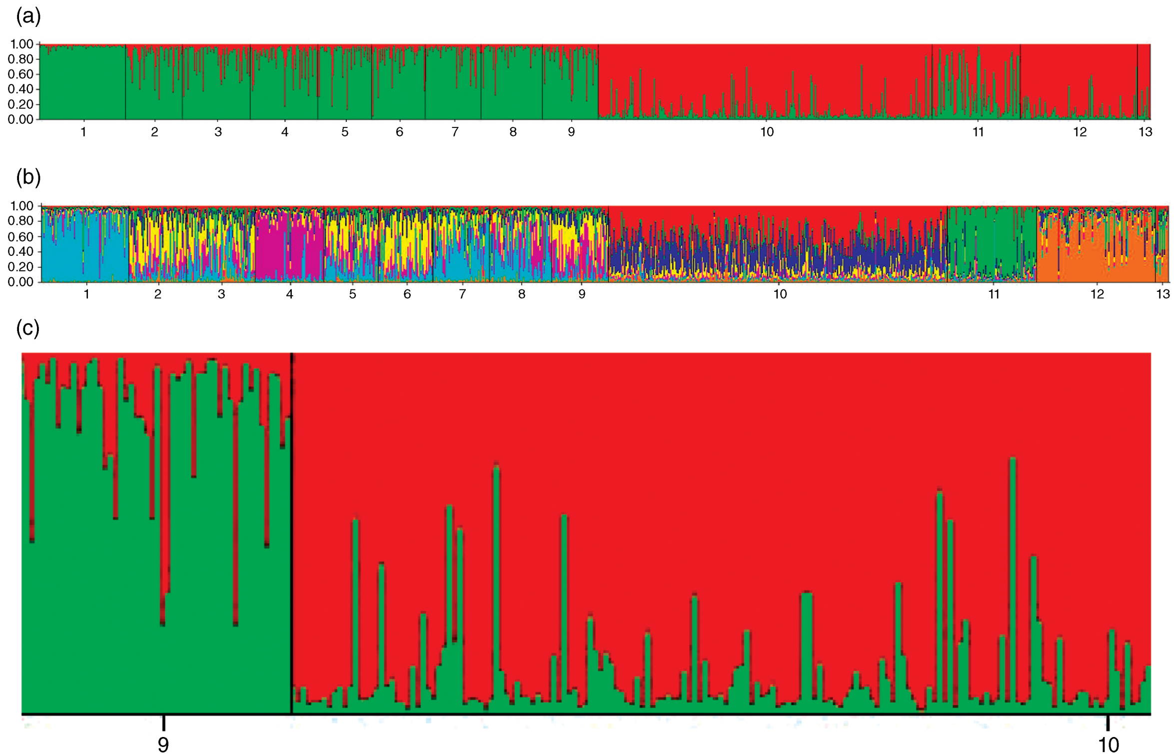


Figure 3.7 Structure analysis of individuals from 13 populations assuming two and seven genomic components. Figure 3.7a represnts a Structure analysis performed assuming two genomic components ($K=2$). Figure 3.7b was generated at seven genomic components ($K=7$). Figure 3.7c represents a magnification of the Puyuma-Sweden portion of Fugure 3.7a. Key to populations:(1) Bunun, (2) Taiwan, (3) Saisiyat, (4) Yami, (5) Ami, (6) Rukai, (7) Atayal, (8) Tsou, (9) Puyuma, (10) Sweden, (11) Society Islands, (12) Yucatan, and (13) Buczotcz. (See the Color Plates section.)

be, since a greater proportion of the genome contributes to the estimations.

In the graph illustrated in Figure 3.7a, each vertical bar indicates a single individual. In this example, the individuals are grouped into 13 predetermined populations designated with the numbers 1–13 on the *X*-axis. The colors indicate the proportion of each individual's loci that are drawn from each of $K=2$ clusters. *K* values represent the number of clusters or groups that the imputed data are assigned by the investigator and partition in any given test. In practice, different numbers of clusters are tried in independent structure tests to ascertain the maximum genetic definition among populations afforded by each *K* value.

The graph in Figure 3.7a exhibits four types of general profiles based on the amounts of red and green (the two clusters) displayed by the individuals in each population. One type of profile is represented by population 1 (the Bunun tribe from Taiwan). A second type of profile is illustrated by populations 2–9 (see key for population names in Figure 3.7), while a third and a fourth type are noted in populations 10, 12–13 and 11, respectively. In addition, although these results indicate that at $K=2$ structure analysis can discriminate among the Taiwanese tribes as a group, Society Islands from Polynesia and Sweden, it fails to differentiate between the Bunun (population 1) and the rest of the Taiwanese tribes as well as between Swedes (population 10) and the Mayans from Yucatan and Buczotcz, populations 12 and 13, respectively. Results like this would justify employing higher *K* values to discern among the populations. Higher resolution and discrimination among populations is in fact seen when the structure analysis was performed at $K=7$ (Figure 3.7b).

Figure 3.7c zooms into an area of Figure 3.7a ($K=2$) that includes population 9 (the Puyuma tribe from Taiwan) and the first portion of population 10 (Sweden) to visualize individual samples. When this area is

magnified, it is possible to discern, for example, that the leftmost individual in population 9 exhibits around 15% of the red constituent that is the predominant genetic component in the Swedish population and 85% of the green cluster, the major constituent of all Taiwanese tribes (populations 1–9). This type of examination allows for the assessment of the genetic composition of any given individual within a population.

Forces affecting population dynamics, structure, and evolution

Selection

A number of forces are known to act and affect the genetic constitution of populations. Selection for or against specific variants represents one mode by which the diversity of a population usually changes. Specific alleles can change in frequency, be deleted, or reach fixation depending on their benefit to the individuals in a population in a given environment.

It has been a paradigm of biology since the beginning of the 20th century that the majority of *de novo* mutations are deleterious or even lethal since the genomes of contemporary species are the result of 3 billion years of trail-and-error mutations followed by selection pressure. Under such conditions, in geological time, the best-suited variants for a given environment have been selected for. Therefore, it is likely that any new random change in the DNA would not be expected to lead to an improvement in the survival potential or fitness of individuals that experience the mutation.

Balancing the negative selection for some of the newly created mutations are the benefits that stem from genetic diversity in the gene pool. In a changing environment, a premium exists for a certain level of diversity in the population. In other words, since the environment is not constant and the nature of the selection pressure changes with time, a healthy gene pool requires a certain level of genetic variability and flexibility that would allow survival. Populations have become extinct as a consequence of genetic homogeneity. This phenomenon has been repetitively observed in many species when the number of organisms in certain populations diminishes creating genetically unhealthy gene pools possibly due to the high frequencies of homozygous deleterious or lethal conditions. In the short term, limited genetic diversity is known to promote homozygosity of deleterious alleles and expression of rare medical conditions. In addition, in the long term, minimal variation could compromise the survival of a population in a changing environment since it limits alternative evolutionary paths. In fact, one of the premises of conservation biology is that species are susceptible to extinction when the gene pool is limited in diversity. Species become genetically unhealthy and vulnerable when the number of individuals reaches a threshold minimum.

A related question is how much neutral or marginally deleterious alleles can be tolerated in the gene pool for the sake of retaining variability; in other words, when does diversity turn into generic load. The *Alu* SINE family represents a good example of insertional mutations that may provide for DNA raw material for the creation of new genes and for genetic flexibility. Clearly, *Alu*s represent a highly successful family of middle repetitive elements that expanded from a single copy, at the beginning of the explosive speciation of primates, to 1 million copies in the human genome today. Since most of the *Alu* subfamilies are not active in retrotransposing into other parts of the genome, it is thought that only a limited number of master *Alu* sequences are currently fertile transcribing and inserting into new sites in the genome. It is postulated that the *Alu* family has reached a point of dynamic equilibrium in which no further net increment of elements is taking place. It is possible that no further net increase in the number of *Alu*s is the result of negative selection from the impact of insertions on functional sequences and as facilitators of illegitimate recombination. Conversely, positive selection may be provided by the benefits these sequences have to the population as source of novel DNA (see Chapter 5 for specific examples of transposons involved in the creation of new genes and their role in evolution).

Selection is also a dynamic evolutionary force in the sense that specific variants can change their fitness value depending on the demands imposed by the environment. In other words, an allele under negative selection pressure may become beneficial if the environmental changes favor people possessing it or vice versa. For example, let us take the classical example of sickle-cell anemia. In certain areas of Africa, India, and the Mediterranean basin, malaria is endemic and a serious health issue. Currently, the sickle-cell allele provides to the heterozygous carriers of the mutation some protection against the malaria parasite. As a result, the sickle-cell variant is under positive selection pressure due to this

partial protection. Red blood cells contain the protein hemoglobin that is composed of two units of the polypeptide alpha-globin and two units of the polypeptide beta-globin, forming the heterotetramer hemoglobin. It turns out that the malaria parasite avoids or does not do well when feeding on the mutant type beta-globin chain, which makes up sickle-cell hemoglobin (SHb). SHb possesses a single-nucleotide point mutation that alters the sixth amino acid of the beta-globin peptide changing it from glutamic acid to valine. This change in the open reading frame of the gene promotes the premature crystallization of the functional heterotetramer hemoglobin under physiological conditions. This abnormal behavior of the mutant hemoglobin leads to the sickle shape of red blood cells as the hemoglobin crystals grow from the inside out pressing on the inside of the cell membrane outward. As the partial oxygen pressure decreases in affected individuals under physical stress (e.g., running), the growing crystals push so hard on the inside of the membrane that it ruptures releasing the hemoglobin into the plasma. Free hemoglobin (not membrane bound) is highly toxic, especially to neurons, as it is metabolized into bilirubin. This condition has negative medical outcomes and as such provides a negative selection force against homozygous individuals for the sickle-cell allele. Hypothetically, if malaria were to be eradicated, the positive selection pressure for the sickle-cell allele would no longer exist and only the negative selection due to hemolysis would remain. Under those conditions, the frequency of the mutant allele should decrease. In this scenario, a change in the environment (no malaria) alters the nature of selection forces.

Genetic drift

Genetic drift is an evolutionary force that relies on random chance. In other words, similarly to tossing a coin 100 times and not getting exactly 50 heads and 50 tails, genetic drift generates fluctuations in allelic or haplotype frequency fortuitously. The concept of drift implies that, over time, the frequencies of characters or markers gradually oscillate toward fixation or deletion from the gene pool.

In real life, a number of circumstances can promote random changes in frequencies of DNA alleles. First, change in allelic levels can occur in the absence of any selection pressure. Frequencies may drift as certain genetic variants gradually fail to be represented in the span of several generations, just as heads of a coin may consistently turn up in a series of tosses. In other instances, for example, a dramatic event could trigger the annihilation of individuals in a population, and if the people who perish are not a random representation of the genetic composition of the population at large, allelic frequencies may be altered in the gene pool. Events such as epidemics and geological catastrophes may randomly eliminate certain genetic types from the gene pool. These phenomena have the potential to promote rapid changes in the frequencies of specific alleles.

Migration and founder effect

As a result of migration, humans have colonized regions previously uninhabited. An often-mentioned example is the incursion and settlement of Paleolithic humans into the Americas. It is well established that humans migrated across the Bering Strait into the New World about 10,000–20,000 years ago. The original source of the migrants is thought to be the Altaic region of south-central Siberia [60, 61]. Radiocarbon dating indicates that humans reached the extreme end of South America in what is now southern Chile about 14,800 years ago. At that time, the Bering Land Bridge between Alaska and Siberia would have been impassable. These data may suggest that Paleolithic humans may have taken a coastal route south. Nevertheless, the proposed dates of the crossing to what is currently Alaska and archeological sites such as Monte Verde in Chile necessitate a rather rapid north to south movement of humans.

Although investigators are not certain on the number of penetrations and migrants, it is thought that the number of individuals colonizing the Americas was limited, and possibly in a few waves. This dispersal in recent human evolution most likely involved repeated instances in which a small group of people moved from one location to another. Each of these stepwise, periodic relocations, likely, did not involve long distances. We can envision a number of scenarios that could have motivated humans to move to a new location including feuds with other clan members, limited food resources, following game, or simply satisfying our insatiable desire to find out what is beyond the horizon, the unknown. Under these conditions, it is possible that the small number of individuals moving to a new location were not representative, genetically, of the original group prior to the split. Therefore, the genetic composition of the departing party and the group left behind would differ. It is likely that these episodes of colonization and new

settlements occurred repeatedly in tandem as humans spread south and eastward. We can envision that by the time humans reached Tierra del Fuego at the extreme end of South America, the groups of migrants could be genetically quite different from the original Asians who ventured into the New World.

Inherent in this process of consecutive displacement of a finite number of people is a reduction of genetic variability. Every time a subgroup separated from the main group and independently settled in a new location, the genetic diversity represented in the migrant party was likely less than the variability of the original group. It is improbable that the genetic composition of a limited number of people would represent the entire genetic pool of the original population prior to the separation. This process is referred to as founder effect. In addition, since the advancing population is usually very small, it is sensitive to genetic drift and inbreeding (see below). Although this is a random phenomenon, reduction of genetic heterogeneity has occurred in a number of human dispersals including the settlement of the Americas and Oceania. This topic will be further discussed in Chapter 6.

Isolation and inbreeding

The settlement of previously uninhabited territories oftentimes brings about the isolation of individuals. At times, distances are sufficiently large that movement of people among settlements becomes rare. Under these conditions, the isolates breed primarily within the group. This type of inbreeding is experienced by migrating groups that settle on islands and in inaccessible regions such as the Amazon. The peopling of Oceania provides a good example of this phenomenon. As humans dispersed from west to east in the Pacific, navigating from island to island, small parties of individuals established themselves in small islands separated by hundreds or thousands of miles of ocean. Except for the expected occasional crossing among islands, sometimes by accident, these communities would live in total isolation and all procreation would take place within the group. This scenario invariably leads to inbreeding.

The consequence of inbreeding is an increasingly genetically limited gene pool. Often the migration to new locations is undertaken by small parties of people. This sets the stage for a founding gene pool with low levels of diversity. Compounding this situation, these enclaves become isolated and inbred. Subsequently, forces such as genetic drift may promote the random deletion of variants from the gene pool, gradually generating additional genetic homogeneity in the secluded population.

Inbreeding decreases the number of alleles of each locus or gene. This, in turn, increases both dominant and recessive allele homozygosity. An increment in the homozygous state augments the possibility of recessive deleterious and lethal alleles becoming expressed. This compromises the well-being of individuals as well as the population at large. In addition, these human groups, due to their low level of genetic diversity, possess a susceptible gene pool with limited capacity to adapt in a changing environment. For these reasons, some isolates may become extinct with time. Medically, these genetically restricted populations are characterized by large incidents of unique recessive illnesses. An interesting example is the large occurrence of cardiovascular diseases in populations from Polynesia that live in the vast expanse of the Pacific Ocean [62]. As previously mentioned, these groups have been subjected to isolation, inbreeding, and genetic drift as a result of migration into small, widely separated islands.

Nonrandom mating

Our species is particularly prone to situations that lead to nonrandom mating. Many of these circumstances involve our cultural environment. In general, for example, there is a tendency to marry individuals that belong to the same ethnic, socioeconomic, educational, or religious group. This selection takes the form of personal preference and/or family pressure to select a mate within their kind, whatever that category may be. At times, nonrandom reproduction is simply driven by personal preferences or prejudices. In other instances, individuals often rationalize and argue that potential mates from the same group are bound to be, overall, more compatible, and therefore increasing the probabilities of a good fit resulting in a successful marriage, and offspring.

Specific cases of nonrandom mating include the preference by individuals, in a number of human populations, to marry within their own immediate or extended family (consanguineous marriage). This has been a practice among royal families from different countries and may derive from an effort to keep power and richness within the clan. A number of religious groups such as Muslims also promote intrafamily marriage, especially between first and second cousins. In many religions, marriage involving a different faith is not desirable or simply forbidden. In other instances, inbreeding is not optional but it is driven by geographical

isolation. Island isolates are a good example as has been observed in the aboriginal tribes of Taiwan and Oceania [63]. Conversely, a number of primitive human populations specifically encourage marriage among members of different villages or clans. These mating arrangements seem to be ancient practice. It is possible that a realization of the negative effects of inbreeding by these groups may have driven this tradition. The consequences of inbreeding include a general decrease in genetic diversity allowing for an increment in the levels of homozygosity in the population. As previously discussed, limited genetic diversity may lead to an unhealthy gene pool and susceptibility to extinction due to inflexibility to evolve in a changing environment. Also, a more immediate outcome is the promotion of the homozygous state involving deleterious alleles. Cases of abundant genetic diseases among inbred human populations are many. The incidence of these genetic conditions is so frequent in certain populations that genetic screening programs have been instituted to identify heterozygous carriers. Some of these racial-based screening programs include Canavan disease, cystic fibrosis, familial dysautonomia, hemoglobinopathies (inherited blood diseases), hemoglobin E, sickle-cell diseases, thalassemia, and Tay–Sachs disease [64].

Bottlenecks

Fundamental to the bottleneck phenomenon is a relatively drastic reduction in population size (Figure 3.8). Bottlenecks may be caused by random events such as geological catastrophes that lead to reduction of population size. A likely consequence of a decrease in population size is a diversity-limited gene pool that is unrepresentative of the variability of the original group. A sharp reduction in the number of individuals and genetic variability can lead to a drop below the minimum viable population size compromising the survival of the group. In other words, just like the previously discussed mechanisms that minimized genetic diversity, bottlenecks may reduce the robustness of the gene pool to survive environmental changes. Yet, bottlenecks differ from general genetic drift in that bottlenecks are usually associated with sudden events such as limited resources, epidemics, earthquakes, floods, fires, droughts, migration of a subgroup of a population, climatic changes, and geological catastrophes, as well as genocide. Since bottlenecks result in a smaller population size, they may lead to inbreeding and fixation or deletion of alleles from the gene pool.

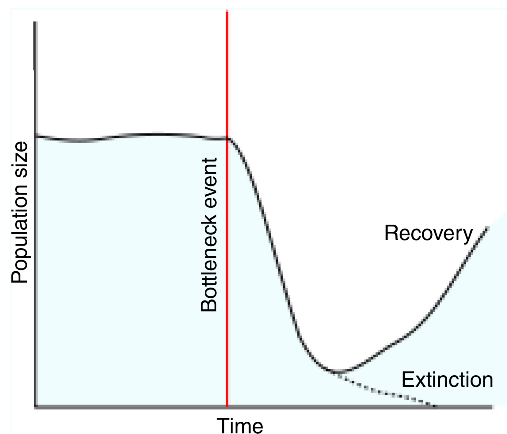


Figure 3.8 Population bottleneck. (*Source*: TedE. https://commons.wikimedia.org/wiki/File%3APopulation_bottleneck.svg. Used under CC BY-SA-3.0, http://creativecommons.org/licenses/by-sa/3.0/.)

In instances in which bottlenecks do not reduce the size of the population randomly, like in selective elimination of the less fit, shrinking of the population may improve the genetic pool. Conditions capable of selecting for the best-fit individuals could bring about a reduction or elimination of unfavorable alleles in the new environment.

It has been postulated that our species has experienced a number of bottleneck episodes. One notorious recent event that may have led to a bottleneck incident in our species was the Toba supervolcanic eruption that occurred about 70,000 years ago in Lake Toba in the island of Sumatra, Indonesia. It is theorized that the dust generated by this eruption provoked a major dramatic environmental collapse similar to a nuclear winter, triggering a global volcanic winter that lasted approximately 6–10 years and a cooling process of at least a millennium [65]. It has been speculated that this catastrophe reduced the worldwide human population to about 10,000 individuals [66]. The Toba supereruption occurred about the same time that the last glacial period commenced. Thus, the geological data corroborate the low level of genetic variation observed in humans today. Also, it is interesting that the Toba event coincided with the original migration of humans out of Africa, as suggested by mtDNA data. It is possible that our species was within the confine of Africa prior to the eruption and the event may have provided the motivation for migration in search for food and resources. Further, prior to the out of Africa dispersal, it is thought that humans experienced a

major bottleneck event that caused a drop of the population to as low as 2,000 for as long as 100,000 years [67].

Admixture

Unlike all the previously discussed evolutionary forces that tend to promote genetic homogeneity, admixture may bring an increase in diversity. The fundamental premise of admixture is the unification of people from two or more populations, creating a novel group of individuals with constituents from the contributing populations. If the participating populations are genetically different, this unification process would promote greater diversity.

Admixture may provide genetically limited populations with an infusion of variability that could rescue the group from extinction. This principle is often employed in conservation biology to salvage endangered species that suffer from genetically compromised gene pools. The idea is that bringing together and mating individuals from different geographical populations would lead to greater genetic diversity and improve the survival of the species.

In humans, admixture episodes are a common occurrence and can lead to linkage disequilibrium (see Chapter 5). Starting with early modern humans, their interactions with Neanderthals may have allowed for introgression between the two groups. It is not clear whether modern humans and Neanderthals represent a single or separate species; yet, it is known that modern humans initially migrated out of Africa by way of the Levant about 120,000 years ago encountering Neanderthals in what is today the Near East [68]. Although that early dispersal out of Africa likely was not successful for modern humans as a species, it clearly brought about contact and the possibility of admixture between the two hominins. Subsequently, another window of opportunity for admixture presented itself when contemporary humans finally ventured into Europe 45,000 years ago. Neanderthals are thought to have become extinct approximately 23,000 years ago. Considering the time of human penetration into Neanderthal territory, at least roughly a 20,000-year period of contact may have allowed admixture to occur. Although various levels of common DNA between humans and Neanderthals have been reported, it is not clear whether this sharing of DNA is the result of contamination, introgression, and/or ancient ancestral polymorphisms (see Chapter 6).

In prehistoric times, well-documented examples of admixture are abundant. In the upper Paleolithic, about 10,000–12,000 years ago, for example, one dramatic event that led to admixture was the agricultural revolution and subsequent expansion of humans into Europe and Asia from Central Anatolia (present-day Turkey). The agricultural revolution ushered our species from a hunting–gathering existence to an agricultural one. This transformation provided for surplus food accumulated during the summer and fall for storage and winter consumption as well as domesticated animals for meat and dairy products. These new practices freed humans from a daily dependence on procuring food. In turn, this allowed for the establishment of homesteads, towns, and communities, since the need for relocation and migration in search of food was not as crucial.

It is still debated to what extent the practice of agriculture and domestication resulted from acculturation or migration of people carrying with them not only the new technologies of farming but their genes as well. Although percentages differ from study to study as to how much cultural diffusion as opposed to gene flow contributed to the spread of this new way of life, today, the orthodoxy acknowledges that some degree of migration, originating in the Near East, reached Europe introducing agriculture and the establishment of homestead in the form of city-states. As a result, most genetic studies involving autosomal and uniparental genetic markers (Y-specific and mtDNA) indicate various frequencies of Paleolithic (before the Agricultural Revolution) and Neolithic (after the Agricultural Revolution) markers in different parts of Europe today. This dispersal and penetration of people certainly resulted in admixture.

In more recent times, the Bantu expansion represents another major migratory event that brought about admixture in sub-Saharan Africa. The term Bantu initially indicated a group of related languages belonging to the Niger–Congo family with wide distribution throughout sub-Saharan Africa. With time, the name Bantu evolved to represent also a culture. It is believed that the Bantu people originated in West Africa in what in now North Cameroon about 5000 years ago. Approximately 4000–3000 years ago, the Bantu people initiated a major human diaspora and associated cultural transformation that rapidly propagated agriculture and ironwork along with the Bantu language to most subequatorial Africa [69]. It is likely that the Bantu were motivated to migrate as a result of limited resources. It is theorized that the Bantu dispersed in two waves, one path from the Bantu homeland along a southwestern course and a second path, also from North Cameroon, in a

southeastern direction. Yet, more recent data argue for an initial single southwestern migration from the Bantu homeland and a subsequent longitudinal dispersion eastward to populate East Africa. During this massive geographical and cultural diffusion, the Bantu migrants encountered and interacted with the indigenous sub-Saharan tribes practicing animal husbandry or hunting–gathering [70].

The Bantu expansion was so impacting that today their language and culture dominate the sub-Saharan landscape. In addition, according to Y-specific haplogroups and haplotype frequency distributions as well as mtDNA markers, the infusion of Bantu DNA into the gene pools of the original non-Bantu populations was profound. For the most part, Bantu DNA predominates over autochthonous genes in most of sub-Saharan Africa. Therefore, like with the agricultural dispersal in Europe, the Bantu expansion was not just simply an acculturation phenomenon but it was driven by migration as well. Overall, the admixture and assimilation processes of Bantu genetic elements into native populations have been both highly complex and region-specific, with frequencies of Bantu and non-Bantu markers fluctuating depending on geographical location [71].

In general, humanity has experienced a dramatic increment in admixture. This has been primarily due to the increased ease and speed of worldwide transportation and multiple invasions. In addition, changes in social norms and a decrease in racial stigmas among contemporary humans are facilitating marriage among most groups of people. Particularly in major cities, admixture involving individuals from different parts of our planet is now a regular occurrence and not an oddity. This type of admixture differs from the examples provided above since they do not involve finite groups of people but individuals from many different backgrounds reproducing in an almost endless number of combinations and permutations of ethnicity and origins. This trend is likely going to continue and augment, and, with time, the concept of discrete populations may be academic.

Applications of population genetics

Introduction

Knowledge of population structure is important to a number of disciplines. For example, internal differentiation within populations or subpopulation structure is paramount to a number of fields including forensic DNA fingerprinting [72], medicine [73], and epidemiology [74] since the calculation of inclusion probabilities and cause-and-effect relationships involving diseases, respectively, should be made in genetically homogeneous groups. Basic research on the origin of humans also requires an understanding of population genetics. Although genetic diversity is limited in our species and admixture is constantly blurring any uniqueness among human groups, the existence of population-specific differences still impacts various aspects of our lives and continues to be important.

The reasons for the differences among ethnic groups stem from a number of factors. One likely contributing element is the impact of genetic drift as humans dispersed from Africa to the rest of the world. The colonization of new land by a limited number of people provided for bottleneck, founder, and genetic drift events that partitioned the genetic diversity of the original out of Africa migrants into geographically separated groups with non-identical gene pools. Although mutations have contributed to the genetic diversity among populations, due to the recent departure of modern humans from their homeland in Africa, the number of *de novo* mutations that occurred subsequent to the diaspora does not account entirely for the group-specific differences that we see today. In addition, our species possesses a cultural universe that often differs among populations creating barriers that inhibit gene flow, preventing genetic homogenization and facilitating genetic drift as well as uniqueness.

The vast majority of the genetic differences that exist among human populations are quantitative in nature. In other words, they are not due to private alleles that are unique to one or a few human groups. This type of diversity in humans likely originated as people dispersed and colonized new territories out of Africa. In this scenario, the partitioning of individuals into migrating groups, in tandem, one location after another, unequally segregated their gene pools. Oftentimes, this lack of qualitative disparity is cited as evidence for the uniformity of our species and the nonexistence of races.

Medicine

In medicine, it is well known that certain diseases not only have a genetic basis but also partition nonuniformly among human populations. Therefore, a better understanding of the genetic differences and similarities at the population level should benefit our understanding of the maladies and the human condition. Diabetes, for

example, is a condition impacted by interethnic differences. Studies reveal that the incidence of the disease is three times higher in African Americans compared with non-Hispanic whites. African Americans, Native Americans, and Hispanics possess the highest risk of developing type 2 diabetes, which makes up 90–95% of all diabetes cases (National Institute of Diabetes and Digestive and Kidney Diseases, 2007). And yet, at the beginning of the 20th century, diabetes was associated with Jews [75]. It is likely that the medical opinion in the early 1900s was influenced by the concept of race and stereotypes that existed at the time.

Proponents of race-based research argue that ethnicity could be a strong predictor of health outcomes for some illnesses and that a better understanding of population-specific susceptibilities will increase the chances of reducing health disparities and improve health in general. Although it is not clear to what extent these ethnic differences result from environmental conditions due to specific diet, limited health care, and/or other cultural factors [76], it is likely that there are some genetic bases to these observations. In the case of diabetes, it is thought that multifactorial inheritance, involving a number of major and minor genes, governs the trait as opposed to Mendelian single-gene genetics. As discussed in Chapter 4, multifactorial genetics coupled with a high environmental component makes it difficult to identify and study the individual genes responsible for multifactorial diseases.

In addition to multifactorial characteristics, a number of Mendelian traits tend to segregate differentially among human populations. A well-known case is cystic fibrosis. Cystic fibrosis exhibits a marked geographical and ethnic distribution. It affects mainly people of European descent and it represents the most frequent genetic condition among this group. Although over 1,000 mutations have been linked to the disease, a three-nucleotide deletion, F508del, is responsible for about two-thirds of the cases and only four other mutations are found at frequencies greater than 1%. Its mode of inheritance is autosomal recessive and impacts mostly the lungs but also other tissues including the pancreas, liver, and intestine. The most obvious symptom is difficulty in breathing, which often leads to lung infections. Mechanistically, the illness is the result of abnormal transport of chloride ions across the membranes of organs. This, in turn, generates a thick viscous secretion that needs to be suctioned out or loosened from the lungs either mechanically by pounding the back and chest or by treatment with DNase pharmaceuticals, or the patients can literally drown in their mucous.

An interesting health-related mutation that also partitions ethnically and geographically is a deletion mutation of 32 base pairs (Δ32) in a gene known as CCR5. The CCR5 gene codes for a membrane glycoprotein that typically acts as a chemokine receptor on macrophages. All strains of HIV require the CD4 receptor for infection. Most HIV viruses use the CCR5 coreceptor and the CD4 receptor to incorporate HIV into macrophages and monocytes. Mutant variants of HIV can arise in macrophages and monocytes that can use the CD4 receptor and CXCR4 as a coreceptor to infect helper T cells. Heterozygous carriers for the CCR5Δ32 deletion exhibit reduced susceptibility to M-tropic strains of HIV and delayed onset to the disease, while homozygotes enjoy resistance to specific strains of HIV. In the deletion homozygotes, the altered glycoprotein cannot fulfill its normal function as an HIV coreceptor, resulting in immunity. It is not clear whether homozygous individuals exposed to the virus are still vectors and capable of infecting others.

From linkage disequilibrium data (described in Chapter 5), it has been assessed that the CCR5Δ32 allele is 700–3,500 years old and it has been under intense positive selection pressure. Although HIV is a recent new arrival to the world of infections, it is theorized that older epidemics such as the smallpox provided the initial selection pressure for the rapid increase in the frequency of the mutation. In addition, from historical records, a fast-moving disease such as smallpox was also likely responsible for the robust positive selection pressure on the CCR5 deletion allele [77]. This CCR5Δ32 mutation exhibits a well-defined geographical distribution (Figure 3.9). The

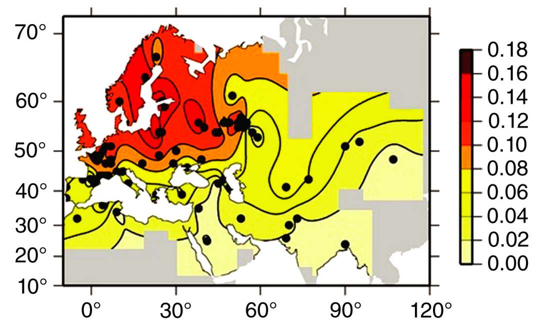


Figure 3.9 Contour map illustrating the frequency distribution of the Δ32 mutation of the CCR5 gene. (*Source*: Novembre et al. 2005, [77]. Used under CC BY-4.0, http://creativecommons.org/licenses/by/4.0/.) (See the Color Plates section.)

highest frequencies of CCR5Δ32 are seen in Northern Europe, specifically in the Baltic region, at 16%, with a north to south clinal distributions into Southern Europe (4% in Greece), Eastern Europe, and West Asia. In addition, Ashkenazi Jews possess high frequencies of CCR5Δ32, likely due to founder effects and isolation unique to their history and not to the process of the mutation dispersal.

Pharmacogenetics

Pharmacogenetics explores the genetic differences in metabolic pathway genes that may impact people's responses to drugs, including the therapeutic effects as well as adverse side effects. For instance, persons of non-African populations respond to several hypertension medications, including angiotensin-converting enzyme inhibitors (ACEIs), angiotensin receptor blockers (ARBs), calcium channel blockers (CCBs), or thiazide-type diuretics as first-line treatments, by lowering blood pressure and reducing cardiovascular and cerebrovascular episodes. In contrast, the African-derived persons have better outcomes using either thiazide-type diuretics or CCBs alone. When ACEI was used as a first-line therapy to treat African American hypertensive patients instead of CCB, the patients had a 51% increase in stroke that coincided with sustained higher blood pressure [78]. However, the problem that physicians face is that it is now difficult to determine, for example, what constitutes being African American because of admixture over the last few centuries. In other words, how much African background do you need to decide one drug will outperform another, and it is no longer sufficient to assess African ancestry by skin color.

DNA fingerprinting

In the field of DNA fingerprinting, an understanding of population structure is paramount. DNA fingerprints are employed in forensic science to identify individuals. In the legal system, this technology is used to calculate the probabilities that a given person provided the evidence DNA. In criminal cases, for example, DNA is usually collected from the scene of the crime and the genetic profiles generated are compared with those from known standards such as victim's and/or defendant's. In civil cases, on the other hand, the scenario may involve paternity disputes and immigration situations where familial relationships are tested prior to admission into a country. Independent of the legal or civil issues at hand, the question that is ultimately asked is what is the random probability that an individual, other than the accused, for instance, is the source of the evidence DNA. To answer this question, it is necessary to ascertain how frequent is the DNA evidence profile in an appropriate population database. This number is referred to as the probability of inclusion and, in the current state of the science, these values are highly incriminating, ranging on the order of one person in several trillions. In other words, on average, trillions of people need to be screened to find another with the same DNA profile as the one found at the evidence site.

Since the probabilities of inclusion are based and dependent on the frequency of specific alleles in the database of specific populations, it is important to assess the appropriateness of the populations used to generate these data sets. For example, a nonauthentic database from a foreign location (e.g., Mongolia) could underestimate the probability of two individuals sharing the same profile in Venezuela. This is due to the fact that the DNA frequencies in Venezuela are rarely found in a Mongolian database and the probability of detecting an individual with the same profile is very small or zero. The issue also becomes relevant in the case of specific polymorphic variant alleles that show higher probability of occurrence in certain populations, where several orders of magnitude differences are observed.

Subpopulation structure could also affect DNA fingerprinting assessments. Also referred to as population stratification, it is the result of differences in allelic frequencies between subgroups within a population. These internal genetic differences within populations may represent violations of the concept of populations as a unit of freely interbreeding individuals part of a single gene pool. Differences in ancestry, geographical proximity, migrational patterns, and cultural and social practices among factions within a population may promote subpopulation structure requiring further study into its effects on the probability of inclusion values.

From populations to races and species

Organisms have been divided and organized into a hierarchical continuum of designations known as taxonomic classification. Primarily for convenience and logistical reasons, this system of nomenclature arranges living

things into groups such as kingdom, phylum, class, order, family, genus, species, races (subspecies), and populations according to degrees of similarity. Our current system of biological classification has its genesis in the work of the Swedish botanist and zoologist Carl Linnaeus (1707–1778) who grouped species according to shared physical characteristics (see Chapter 1). These taxa have since been modified incorporating the Darwinian principle of common descent. Although our current taxonomic system aims at objectivity, in practice, it is often subjective due to the relative weight (importance) given to different diagnostic characters (traits) in assigning organisms to classifications (e.g., same or different genera). In addition, the system is somewhat arbitrary since there are no set rules or parameters to decide, for example, how different groups of organisms need to be in order to be considered as different races or subspecies or any other taxa. Since the age of molecular biology, additional objectivity in the form of genetic markers as parameters is employed to ascertain taxa. Yet, the system still remains somewhat loose.

Keeping the above issues in mind, in a portion of this hierarchical sequence of taxa in our classification system, we have populations, races, and species. Most biologists agree that groups of organisms that are not able to reproduce among themselves are members of different species. A number of mechanisms are known to be responsible for providing for speciation including geographical and reproductive isolation that with time leads to genetic incompatibility. Organisms potentially capable of successfully interbreeding may be members of the same population or they may belong to populations with different gene pools. Individuals from certain populations sometimes are capable of interbreeding, but do not interbreed due to geographical and/or social separation. Races or subspecies are thought, by most biologists, to represent populations with differences greater than that observed among populations but still capable of interbreeding.

Historically, the term race was employed to describe nationalities and individuals with a common language [79]. Later in the 17th century, the term race was used to describe people with different physical attributes, and more recently, in the 1800s, it began to be used in a taxonomic sense to categorize populations with different sets of unique characteristics [79]. Then, in the second part of the 20th century, especially in the aftermath of World War II, the concept of human races was subjected to intense criticisms. Some of the arguments advanced were the subjectivity of the race classification, while others seemed to stem from reactions to the sociopolitical environment of the time. At present, a substantial number of scientists believe that the concept of race has no biological significance as it applies to humans and point out that *Homo sapiens* is a single uniform species. The term has become so politically charged that most people, including scientists, avoid using it and use terms such as ethnic groups, populations, or communities. It is interesting that even an organization such as UNESCO, in the mid-20th century, used *ethnic groups* instead of races to communicate the concept—a change that many considered an issue of semantics.

Today, many biologists and physical anthropologists think that race is a valid biological concept and feel that, other than being stigmatic in nature, there is no real scientific reason for abstaining from subdividing our species into groups, populations, or races. According to the classical concept of race, humanity can be partitioned into three to five subspecies (Figure 3.10). However, opponents to the use of the term race argue that the classification is arbitrary and, therefore, meaningless.

Proponents of the concept of race contend that independent of the name given, physically and genetically different groups of humans exist. They indicate that the

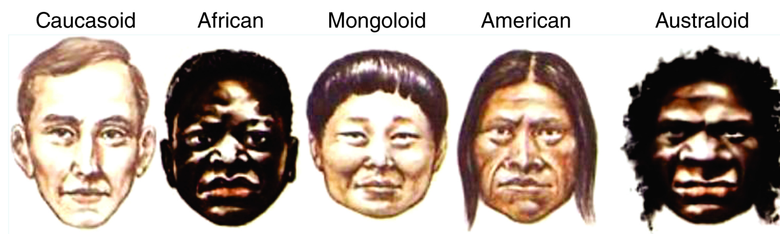


Figure 3.10 Commonly recognized human races. (*Source*: https://en.wikipedia.org/wiki/File:Human_race.jpg. Used under CC BY-SA 3.0, http://creativecommons.org/licenses/by-sa/3.0/.) (See the Color Plates section.)

actual use of the name race should not be banned because of negative connotations and associations. What is important, they state, is that although most of the population differences are quantitative and clinal in nature and not private (population-specific), they are sufficiently evident to partition humanity into groups. These scientists claim that opponents of the concept are in a state of racial denial for nonscientific reasons.

Conversely, highly notable scientists argue that there are a number of irrefutable facts that undermine the biological concept of race as applied to humans. Richard Lewontin was the first to note that most genetic variation resides among individuals within populations and not among populations or races. It turns out that about 85% of our diversity is found internally within populations and only the remaining 15% of the variation falls along the classically defined racial groups. In other words, most human genetic variation does not exhibit race clustering and, in addition, most of the variation of our species is continuous without sharp boundaries. It was also argued that since the existing genetic differences among populations are in decline due to interracial marriages, and are destined to disappear soon, most people are going to be mixed and any original differences will disappear.

Most of the arguments in this debate stem from the fact that there are no clear criteria for assessing how much difference among populations of genetically related individuals is necessary to substantiate the existence of races. In this regard, we must keep in mind that taxonomic classifications are to some degree arbitrary and subjective. In addition, it is likely that many people, including some scientists, consider the concept of race socially dangerous—dangerous to the point of promoting racism as previously seen in human history. Thus, many people have a tendency to undermine the concept of race not because it is scientifically invalid but because they fear potential negative outcomes. The term race is politically incorrect to many people, independent of the evidence.

Review questions and exercises

1. Describe the main types of DNA-based marker systems used in population genetic studies.
2. Explain what SNPs, STRs, and indels are.
3. How much of DNA in our genome is repetitive?
4. What are *Alu* sequences? How they originated? How abundant are they? What function they perform?
5. Define introgression.
6. What are ancient retroviral insertions (ARIs)? What percentage of our genome is represented by ARIs?
7. Could HIV evolve into stable retroviral insertions in our genome?
8. What advantages, if any, have DNA markers over morphological characteristics in assessing phylogenetic relationships?
9. Define SINEs and LINEs, elaborate on their putative function, and explain how are they kept in our gene pool.
10. What are uniparental markers?
11. Defend or argue against the statement “The nonrecombining portion of the Y chromosome represents a large haplotype.”
12. How different types of DNA markers with various mutation rates should be used to study ancient and recent human evolution? *Hint*: Should we employ hypermutable STR loci to explore speciation events dating back to the origin of primates?
13. What advantages provide insertional elements such as *Alu*s in the study of population genetics?
14. How could STR sequences function?
15. Define genetic anticipation.
16. In indel mutations how can you tell whether the insertion state or the deletion state is the ancestral condition?
17. How interspersed repetitive sequences promote deletions and additions?
18. Is natural selection a constant force in evolution or is dynamic, subject to the demands of the environment? Elaborate providing examples.

19 Provide possible explanations why the human species is relatively homogeneous.

20 Explain why genetic diversity provides for a healthy gene pool.

21 How the Toba supervolcanic eruption may have impacted recent human evolution?

22 Describe how the agricultural revolution helped shape the gene pools of European populations.

23 Defend and argue against race-based medical research.

24 How the Δ32 mutation in the CCR5 glycoprotein provides immunity to HIV infection?

25 Defend and argue against the biological concept of human races.

26 Agree or disagree with the statement "Many people have a tendency to undermine the concept of race not because it is scientifically invalid but because they fear potential negative outcomes. The term race is politically incorrect to many people, independent of the evidence."

27 Enumerate three programs routinely used to assess phylogenetic relationships among populations. Explain how the data generated are interpreted.

28 Describe the evolutionary forces affecting genetic diversity within populations.

29 Explain how population genetics impact other disciplines such as medicine, forensic science, and human evolution.

References

1 Lane C. 2013. The beauty of diversity. *Motivated* **4**: 11.
2 Behar DM et al. 2008. The dawn of human matrilineal diversity. *Am J Hum Genet* **82**:1130–1140.
3 Li W, Freudenberg J. 2009. Two-parameter characterization of chromosome-scale recombination rate. *Genome Res* **19**:2300–2307.
4 Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**:406–425.
5 Pearson K. 1901. On lines and planes of closest fit to systems of points in space. *Philos Mag* **2**:559–572.
6 Cox TF, Cox MAA. 2001. *Multidimensional Scaling*. Chapman & Hall, London, UK.
7 Hardy GH. 1908. Mendelian proportions in a mixed population. *Science* **28**:49–50.
8 Weinberg W. 1908. Über den Nachweis der Vererbung beim Menschen. *Jahresh Vereins für vaterländ Naturkunde in Württemberg* **64**:368–382.
9 Edwards AWF. 1977. *Foundations of Mathematical Genetics*. Cambridge University Press, Cambridge, MA.
10 Nei M. 1973. Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* **12**:3321–3323.
11 Allendorf FW. 1986. Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biol* **5**:181–190.
12 Excoffier L et al. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA data. *Genetics* **131**:479–491.
13 Kingman JFC. 2000. Origins of the coalescent 1974–1982. *Genetics* **156**:1461–1463.
14 Cooper G et al. 1996. Network analysis of human Y microsatellite haplotypes. *Hum Mol Genet* **5**:1759–1766.
15 Chiaroni J et al. 2010. The emergence of Y-chromosome haplogroup J1e among Arabic-speaking populations. *Eur J Hum Genet* **18**:348–353.
16 Wright S. 1969. *Evolution and the Genetics of Populations*, Vol. **2**. University of Chicago Press, Chicago, IL.
17 Efron B. 1979. Bootstrap methods: another look at the jackknife. *Ann Stat* **7**:1–26.
18 Ceppellini R et al. 1955. The estimation of gene frequencies in a random-mating population. *Ann Hum Genet* **20**: 97–115.
19 Shields DC et al. 1999. The "thermolabile" variant of methylenetetrahydrofolate reductase and neural tube defects: an evaluation of genetic risk and the relative importance of the genotypes of the embryo and the mother. *Am J Hum Genet* **64**:1045–1055.
20 Nei M, Tajima F. 1981. DNA polymorphism detectable by restriction endonucleases. *Genetics* **97**:145–163.
21 Chang YM et al. 2007. Haplotype diversity of 16 Y-chromosomal STRs in three main ethnic populations (Malays, Chinese and Indians) in Malaysia. *Forensic Sci Int* **167**:170–176.
22 Slatkin M, Excoffier L. 1996. Testing for linkage disequilibrium in genotypic data using the EM algorithm. *Heredity* **76**:377–383.
23 Reich DE. 2001. Linkage disequilibrium in the human genome. *Nature* **411**:199–204.
24 Excoffier L, Slatkin M. 1995. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* **12**:921–927.

25 Kingman JFC. 1982. The coalescent. *Stoch Process Appl* **13**:235–248.
26 Ewens WJ. 1972. The sampling theory of selectively neutral alleles. *Theor Popul Biol* **3**:87–112.
27 Excoffier L et al. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform* **1**:47–50.
28 Wright S. 1950. Genetical structure of populations. *Nature* **166**:247–249.
29 Raymond M, Rousset F. 1995. GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *J Hered* **86**:248–249.
30 Paetkau D et al. 1997. An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. *Genetics* **147**:1943–1957.
31 Smouse PE. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Syst Zool* **35**:627–632.
32 Roff DA, Bentzen P. 1989. The statistical analysis of mitochondrial DNA polymorphisms: X^* and the problem of small samples. *Mol Biol Evol* **6**:539–545.
33 Dunn OJ. 1961. Multiple comparisons among means. *J Am Stat Assoc* **56**:52–64.
34 Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* **7**:256–276.
35 Cordaux R et al. 2006. Estimating the retrotransposition rate of human *Alu* elements. *Gene* **373**:134–137.
36 Kellis M et al. 2014. Defining functional DNA elements in the human genome. *Proc Natl Acad Sci USA* **111**: 6131–6138.
37 Treangen TJ, Salzberg SL. 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* **13**:36–46.
38 Rowold DJ, Herrera RJ. 2000. *Alu* elements and the human genome. *Genetica* **108**:57–72.
39 Batzer MA, Deininger PL. 2002. *Alu* repeats and human genomic diversity. *Nat Rev Genet* **3**:370–379.
40 Pereira LMH et al. 2005. Phylogenetic information in polymorphic L1 and *Alu* insertions from East Asians and Native American populations. *Am J Phys Anthropol* **128**:171–184.
41 Herrera RJ et al. 2006. Ancient retroviral insertions among human populations. *J Hum Genet* **51**:353–362.
42 Mighell AJ et al. 1997. Alu sequences. *FEBS Lett* **417**: 1–5.
43 von Sternberg RM et al. 1992. Genome canalization: the coevolution of transposable and interspersed repetitive elements with single copy DNA. *Genetica* **86**:215–246.
44 Levitt M. 2001. The birth of computational structural biology. *Nat Struct Biol* **8**:392–393.
45 Green RE et al. 2010. A draft sequence of the Neandertal genome. *Science* **328**:710–722.
46 Ceppellini R et al. 1955. The estimation of gene frequencies in a random-mating population. *Ann Hum Genet* **20**:97–115.
47 Zeng Z et al. 2014. Extreme genetic heterogeneity among the nine major tribal Taiwanese island populations detected with a new generation Y23 STR system. *Forensic Sci Int Genet* **12**:100–106.
48 Cavalli-Sforza LL, Edwards AWF. 1967. Phylogenetic analysis: models and estimation procedures. *Am J Hum Genet* **19**:233–257.
49 Nei M. 1972. Genetic distances between populations. *Am Nat* **106**:283–292.
50 Edwards AWF. 1971. Distance between populations on the basis of gene frequencies. *Biometrics* **27**:873–881.
51 Reynolds JB et al. 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* **105**:767–779.
52 Rogers JS. 1972. Measures of genetic similarity and genetic distances. *Stud Genet Univ Texas Publ* **7213**:145–153.
53 Prevosti A. 1975. Distances between populations of *Drosophila subobscura,* based on chromosome arrangements frequencies. *Theor Appl Genet* **45**:231–241.
54 Nei M, Roychoudhury AK. 1974. Genic variation within and between the three major races of man, Caucasoids, Negroids, and Mongoloids. *Am J Hum Genet* **26**:421–443.
55 Nei M et al. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J Mol Evol* **19**:153–170.
56 Cooper G et al. 1999. An empirical exploration of the genetic distance for 213 human microsatellite markers. *Am J Hum Genet* **65**:1125–1133.
57 Bruvo R et al. 2004. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Mol Ecol* **13**:2101–2106.
58 Regueiro M et al. 2015. From Arabia to Iberia: a Y chromosome perspective. *Gene* **564** (2): 141–152.
59 Dulik MC et al. 2012. Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and indigenous Altaians. *Am J Hum Genet* **90**:229–246.
60 Regueiro M et al. 2013. On the origins, rapid expansion and genetic diversity of Native Americans from hunting-gatherers to agriculturalists. *Am J Phys Anthropol* **150**: 333–348.
61 Wahi S et al. 1997. Risk factors for coronary heart disease in a population with a high prevalence of obesity and diabetes: a case-control study of the Polynesian population of Western Samoa. *J Cardiovasc Risk* **4**:173–178.
62 Zeng Z et al. 2014. Taiwanese aborigines: genetic heterogeneity and paternal contribution to Oceania. *Gene* 240–247.
63 Bach G et al. 2001. Tay–Sachs screening in the Jewish Ashkenazi population: DNA testing is the preferred procedure. *Am J Med Genet* **99**:70–75.
64 Savino J, Jones MD. 2007. *Supervolcano: The Catastrophic Event That Changed the Course of Human History: Could Yellowstone Be Next*. Career Press, Pompton Plains, NJ.
65 Dawkins R. 2004. *The Grasshopper's Tale: The Ancestor's Tale, A Pilgrimage to the Dawn of Life*. Houghton Mifflin Company, Boston, MA.

66 Behar DM et al. 2008. The dawn of human matrilineal diversity. *Am J Hum Genet* **82**:1130–1140.
67 Herrera KJ et al. 2009. To what extent did Neanderthals and modern humans interact? *Biol Rev* **84**:245–257.
68 Diamond J. 1997. *Guns, Germs, and Steel: The Fates of Human Societies*. W.W. Norton & Company, New York, USA.
69 Cavalli-Sforza LL. 1986. *African Pygmies*. Academic Press, Orlando, FL.
70 Rowold D et al. 2014. At the southeast fringe of the Bantu expansion: genetic diversity and phylogenetic relationships to other sub-Saharan tribes. *Meta Gene* **2**:670–685.
71 Jeffreys AJ et al. 1984. Hypervariable 'minisatellite' regions in human DNA. *Nature* **314**:67–73.
72 Light KC et al. 1987. Effects of race and marginally elevated blood pressure on cardiovascular responses to stress. *Hypertension* **10**:555–563.
73 Facinelli B et al. 1990. Use of DNA fingerprinting in an epidemiologic study of outbreak-specific and non-specific strains of group C *Neisseria meningitides*. *Eur J Epidemiol* **6**:80–83.
74 Thomas WH. 1904. Medical treatment of diabetes. *Albany Med Ann* **25**:358–363.
75 James SA. 1984. Socioeconomic influences on CHD in black populations. *Am Heart J (Suppl)* **108**:669–672.
76 Novembre J et al. 2005. The geographic spread of the CCR5 Δ32 HIV-resistance allele. *PLoS Biol* **3** (11): e339.
77 James PA et al. 2014. Evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). *J Am Med Assoc* **311**:507–520.
78 Keita SOY et al. 2004. Conceptualizing human variation. *Nature* **36**:S17–S20.

CHAPTER 4
Genetic variability

At bottom every man knows well enough that he is a unique being, only once on this earth; and by no extraordinary chance will such a marvelously picturesque piece of diversity in unity as he is, ever be put together a second time.

—*Friedrich Nietzsche* [1]

SUMMARY

A number of forces act at the population level to shape gene pools. Mutations, the raw material of evolution, are the sources of variability that are then acted on by mechanisms that alter the abundance of genetic variants. These forces include genetic drift, selection, bottleneck events, founder effects, isolation, nonrandom mating, and inbreeding. Mutations, for the most part, occur at random throughout the genome and happen spontaneously or can be induced by chemicals or radiation. DNA areas rich in repetitive sequences, short tandem repeats (STRs), and GpC dinucleotides experience elevated mutation rates. Some mutations are subject to natural selection, yet others are selectively neutral or almost neutral. Mutations that are under selection pressure usually are retained in genomes as a result of balanced polymorphisms in which positive and negative selection pressures reach a state of dynamic equilibrium. In the state of balanced polymorphisms, heterozygotes have a higher fitness than both homozygotes, and directional selection occurs when one homozygote is favored (positive selection) compared with the other (negative selection). Sickle-cell anemia is a classical example. Neutral mutations are particularly useful in studies of genetic variability since they represent markers that reflect true ancestry more faithfully. The phenotype not only is dictated by our DNA but always has an environmental component. Mutations occur in somatic or germline tissues with different consequences, the former being a contributor to the genesis of cancer and the latter as a source of evolutionary change. The environment is never constant and since selection pressure affects allelic frequencies of many genes, the frequencies of specific genetic variants fluctuate. These selection-driven alterations in allelic frequencies may obscure the true phylogenetic relationships among groups of organisms.

On the nature of variability

The diversity observed in organisms results from the interaction of two factors: information stored in the genetic material of inheritance and the impact of the environment. We see variability at a number of levels, including molecular, cellular, organismal, and population. Yet, clearly, diversity is more apparent in the way individuals look to the naked eye. Revealing differences at the molecular and cellular levels requires instrumentation and/or technical expertise. In our species, the genetic material or inheritance is stored as deoxyribonucleic acid (DNA). All of the DNA information in the nucleus is collectively known as the genome. DNA and the information contained in it is transferred to ribonucleic acid (RNA) and, then, to different types of proteins that are the direct contributors to the phenotype in combination with the environment. In some instances, the RNA impacts the phenotype directly since some genes are not used to make proteins and just make RNA that is not converted to proteins. Examples of these RNA genes include transfer RNA (tRNA), ribosomal RNA (rRNA), and small nuclear RNA (snRNA).

The variability observed among human groups ultimately derives from the diversity of individual members

Genomes, Evolution, and Culture: Past, Present, and Future of Humankind, First Edition.
Rene J. Herrera, Ralph Garcia-Bertrand, and Francisco M. Salzano.

© 2016 John Wiley & Sons, Ltd. Published 2016 by John Wiley & Sons, Ltd.

of each population. In other words, a given population is characterized by the total genetic composition of its individual members or its gene pool. Therefore, a knowledge of the basis of the individuals' genetic differences is imperative for an understanding of diversity at the population level. Genetic diversity stems from changes in the chemistry of the genome or DNA. These changes are often referred to as mutations and, for the most part, they occur randomly within the genome [2]. They can be characterized as spontaneous or induced (by chemicals or radiation). Mutations can also be defined according to the nature of the change. For example, the so-called point mutations involve a single or a few nucleotide alterations. A nucleotide is the basic unit and building block of DNA (as explained in Chapter 2). Larger numbers of nucleotides are affected when pieces of DNA are deleted, duplicated, or their orientation is reversed. Traditionally, when additions and deletions of DNA are big enough to be observed with the light microscope (several meganucleotides in size), the mutations are considered gross chromosomal aberrations (see Chapter 5).

At the population level, mutations that are not under negative selection pressure or are mildly selected against may accumulate in the gene pool [3]. Such mutations are considered variants and they make up the diversity within populations. When these variants reach levels of 1% or higher in a population, they are considered polymorphic. Certain mutations can further become fixed when they replace the original wild-type form of the gene. Mutations that are selectively neutral or even under slightly negative selection pressure may increase in frequency within the gene pool just by chance [4]. In some instances, mutations exist in a state of dynamic equilibrium in the population due to negative and positive selection acting on them and balancing each other [5]. For example, a mutant allele could be selected against due to its detrimental effect on the phenotype and, at the same time, be positively selected for the impact on other aspects of the phenotype. A classical case of balanced polymorphism is sickle-cell anemia in which the selection against affected individuals, due to poor oxygen- and CO_2-carrying capacity of a mutant form of the beta-hemoglobin in the blood, is balanced by the relative immunity enjoyed by carriers against malaria (Figure 4.1) [6]. Although it is suspected that a good portion of our genome is made up of selectively neutral mutations or those under balancing selection, it is not clear what are the exact proportions. Intuitively, DNA sequences that have a negative effect on the survival of

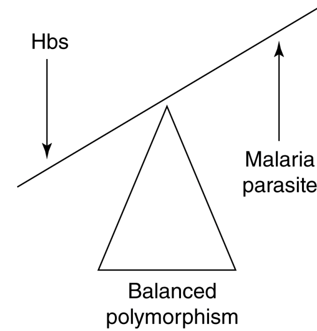


Figure 4.1 Balanced polymorphism. The arrow label "Hbs" pushing down on the scale represents negative selection resulting from the poor O_2 and CO_2 carrying capacity of sickle-cell hemoglobin while the arrow label "Malaria parasite" pushing up on the scale indicates positive selection due to the partial protection that heterozygous sickle-cell carriers enjoy against malaria.

an individual would be negatively selected in the population preventing them from reaching polymorphic levels. Yet, it is likely that some mutations only impact function mildly affecting the phenotype and survival so minimally that they could drift to polymorphic levels just by chance.

Mechanisms responsible for generating genetic variability

There are two general types of mutations (somatic and germinal) based on whether the DNA change is going to be passed to the next generation or not (Figure 4.2). In order for changes in the DNA to be transmitted to the offspring, they need to occur in germline tissue (germinal), cells that are destined to generate gametes (sperms and ova). Somatic tissue is composed of cells that make up most of our bodies except for the cells destined to produce gametes. The outcomes of the two types of mutations are biologically very significant but their consequences are very different. Somatic mutations may occur in tandem in the same cell leading to a number of medical conditions including cancer. Usually cancer originates and becomes more deadly resulting from an accumulation of changes in the DNA of a cell. In this scenario, the normal wild-type alleles of a cell are mutated to forms that allow cells to become unregulated, fast-growing, and immortal, capable of growing into localized tumors (benign), or invading metastasizing tumors (malignant) if not detected and destroyed by

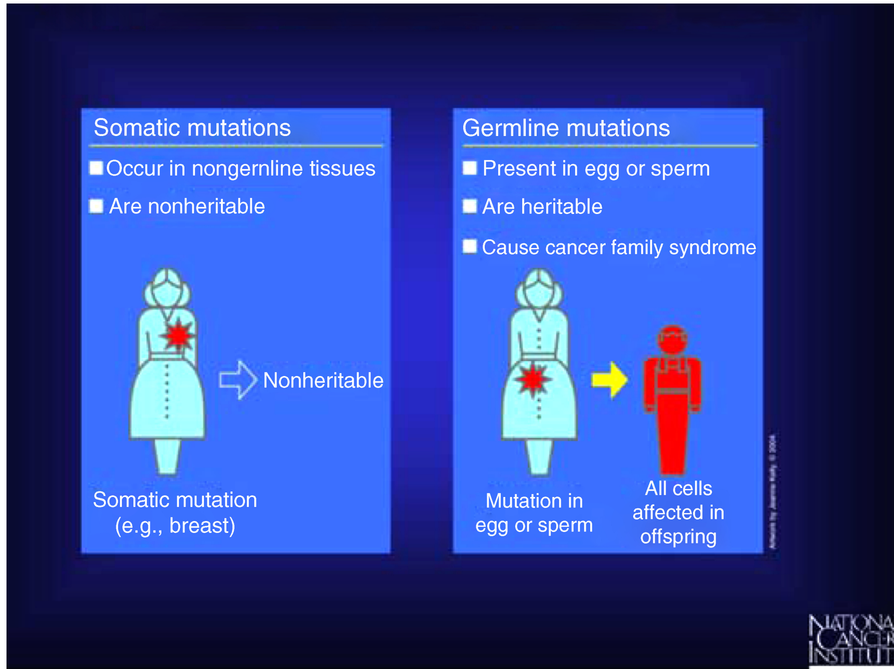


Figure 4.2 Somatic and germline mutations. (*Source*: National Cancer Institute, U.S. Department of Health.)

surveillance mechanisms such as the immune system. Evolutionarily speaking, these types of mutations are silent and of no consequence except when they limit the life span of potential parents and indirectly prevent them from reproducing and passing their DNA to the next generation. Germinal mutations, on the other hand, have the potential to pass altered DNA into offspring and if the mutations are detrimental they can impact the health, life expectancy, and quality of life of the offspring. These mutations can be incorporated into the gene pool and in doing so affect the course of evolution.

Various mechanisms are responsible for causing changes in the DNA. The same mechanisms are responsible for changes in the genes of all cells, whether they will pass the mutations to the next generation or not. One category involves changes caused by environmental agents such as ultraviolet light, nuclear radiation, cosmic rays, and certain chemicals. Ultraviolet light, a component of sunlight, for example, usually induces thymidine bridges that covalently link the nitrogenous bases of adjacent thymidine nucleotides to each other. At the time of DNA replication, DNA polymerase cannot copy these chemically modified nucleotides and stops. Subsequently, the polymerase either halts replication and truncates the replication process or bypasses the paired thymidines, producing a DNA strand with a deletion of two or more nucleotides.

Another type of mutation involves mistakes made by the DNA replication machinery, basically errors by the DNA polymerase enzyme, which reads the template strand of the DNA, to make a new complementary strand, completing the double-stranded molecule.

Overexposure or accidents may increase mutations driven by environmental agents. For example, extended exposure to sunlight could induce mutations resulting from ultraviolet light [7]. The mechanism by which mutations take place differs depending on the inducing agent. For example, ionizing radiation can promote single- and double-stranded breaks in the double helix severing the phosphodiester bonds between oxygen (O) and phosphate (P) of the DNA backbone. This type of damage, if incorporated into a gene, would code for a truncated gene incapable of producing a complete protein that may be partially or completely nonfunctional. Another possible outcome from this type of damage is that the cell will attempt to fix the broken ends by joining these free ends to other pieces of DNA (e.g., telomeres) within the cell. In instances of double-stranded breaks in

two or more chromosomes, translocations may occur in which pieces of DNA are exchanged between chromosomes. This creates a type of mutation called reciprocal translocation. Also, double breaks within the same chromosome could lead to inversions of sequences by 180° [8] (see Chapter 5 for the section on chromosome rearrangement). If a translocation breakpoint occurs within or near a gene, that gene's function may be affected since abnormal chimeric proteins or aberrant transcription, respectively, may occur.

Furthermore, particular chemicals such as methylating and deaminating agents are capable of introducing methyl groups or removing amino groups, respectively, from the nitrogenous bases of nucleotides. These alterations in the nitrogenous bases make the nucleotides look like different nucleotides that are recognized incorrectly by the DNA polymerase at the time of DNA replication. Then, as a result, the DNA polymerase incorporates an incorrect nucleotide in the newly synthesized DNA strand. For example, a modified G could pair with a T, instead of its normal complement, a C.

Mutations may also occur without the intervention of chemicals or radiation. DNA polymerase makes mistakes introducing incorrect nucleotides [9]. This happens about once every 100,000,000 bases. These errors by the replicating machinery lead to unpaired single-stranded nucleotides that look like "bubbles" in the vastness of the double-stranded DNA. This region of unpaired single-stranded DNA may be detected by DNA repair mechanisms and then proceed to cut out the nucleotides of one of the two unpaired DNA strands, at random. Then, the remaining single-stranded DNA is copied completing the double helix (Figure 4.3). If the nucleotide that is removed is the nucleotide that was incorrectly incorporated by the DNA polymerase, the resulting repaired DNA would be identical to the original sequence. In this scenario, no mutation is generated. If, on the other hand, the nucleotide that is removed is the original nucleotide and not the one introduced in error, the DNA sequence will change in relation to the original double helix and in doing so a mutation would be introduced (Figure 4.3). Gene conversion is another way of creating unpaired regions of nucleotides that the cell may attempt to repair by a similar process (see Chapter 5 for more on gene conversion) [10]. During gene conversion, a sequence replaces the homolog sequence in the complementary chromosome such that both DNA molecules become identical at the end of the process. The cell is also capable of creating unpaired nucleotides by gene conversion involving homologous but not identical DNA double helices pairing during gamete formation.

Although we are capable of limiting our exposure to mutagens, it is impossible to eliminate them entirely from the environment, or shield ourselves from them. Every day we are exposed to ultraviolet light, solar radiation, and chemical agents, some of which are known mutagens and/or carcinogens. Similarly, errors due to the limited fidelity of the DNA polymerase during DNA replication are inevitable. We should also keep in mind that since the genetic code is redundant and more than one nucleotide triplet may code for the same amino acid, some mutations would not alter the amino acid that is specified. Therefore, these mutations are considered evolutionarily silent since the protein that is produced is unchanged, and thus, these DNA changes are of no consequence to the fitness of future generations.

Randomness of mutations

Mutations may be random with respect to their effect on the fitness of the organisms carrying them [11]. In other words, potentially beneficial mutations do not occur at higher frequency compared with mutations that confer a disadvantage, or vice versa. Yet, the distribution of genetic diversity is not random within the genome of humans and other organisms. At first glace, the data that illustrate the degree of diversity along the DNA may appear as if certain areas of our genome experience different rates of mutations. Yet, if we examine the degree of variability in different types of coding and noncoding DNA sequences, it is seen that sequences that have functional value exhibit less diversity than regions that do not (Figure 4.4). This correlation between function and variability is thought to derive mostly not from different mutation rates but from the various intensities of selection pressure on specific sequences or regions of DNA. Further examination of the plot in Figure 4.4 reveals that there are a number of levels of diversity, with protein-coding sequences displaying the least and intergenic DNA the most. Other areas with various known functions, such as transcription regulatory DNA elements and RNA processing signal sequences, exhibit different degrees of conservation commensurate with how critical their functions are.

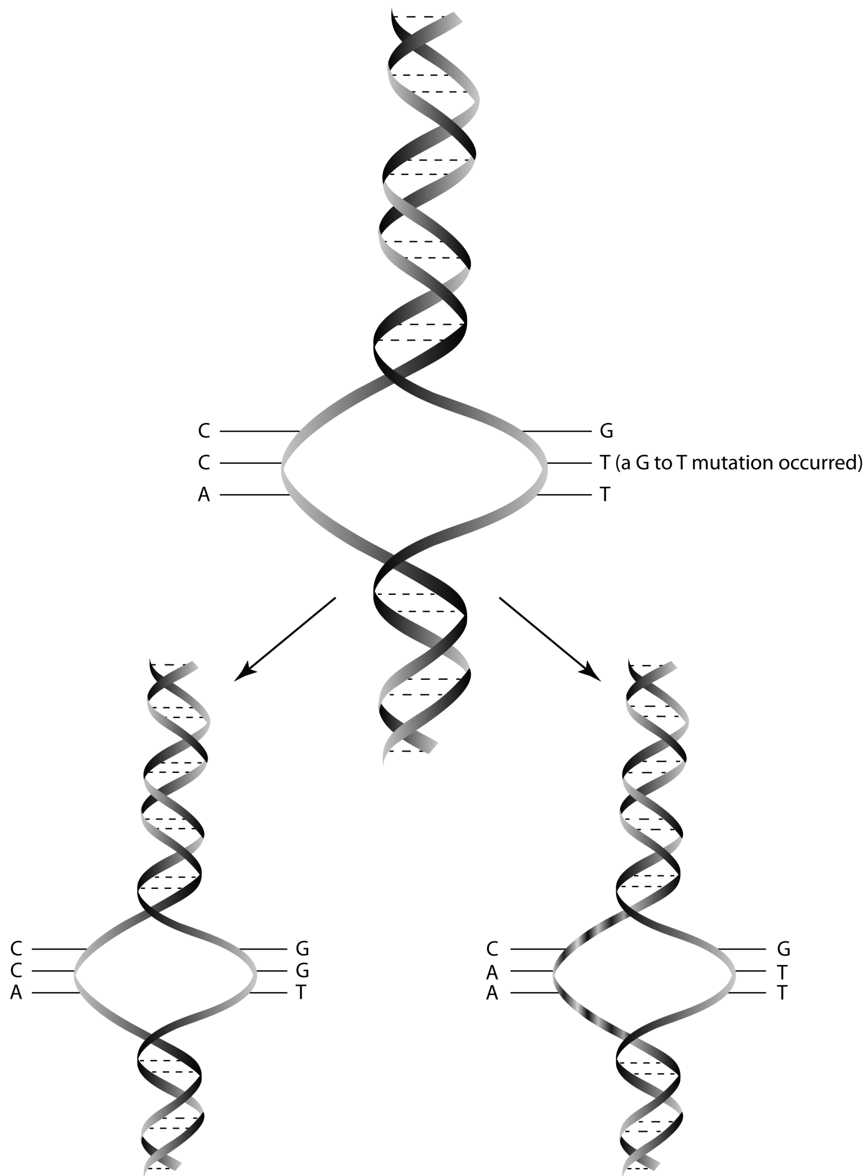


Figure 4.3 Excision repair of DNA. A nucleotide change occurs substituting a "T" for a "G" (top sequence, right strand, middle nucleotide). As a result of the unpaired nucleotide, DNA strands separate. Deletion of one of the two unpaired strands is performed by repair mechanisms (bottom sequences). Deletion of the nonmutated DNA strand leads to a change in the resulting DNA double helix (bottom right sequence).

This interpretation of the data is intuitive if we consider the expected consequences of changes in the DNA responsible for a hierarchy of functional importance. In other words, most mutations affecting critical areas are strongly selected against due to the severity of the consequences of randomly altering the sequences that have been under selection pressure for millions or billions of years. Humans, and all living things, are the product of evolution and selection of the fittest since the beginning of life and the likelihood of a random change in the DNA improving the function of such an optimized sequence is small.

Yet, the statement that mutations are random is not entirely correct. Spontaneous point mutations occur at a

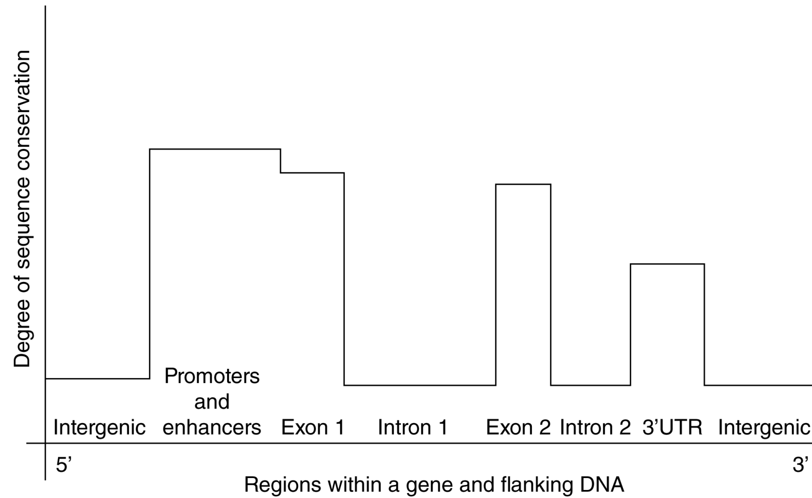


Figure 4.4 Levels of diversity along DNA. Sequence conservation along a stretch of DNA varies directly proportional to degree of functionality.

frequency of about 1.1×10^{-8} per site per generation [12]. Mutation rates change with time as a result of fluctuations in exposure to environmental insults and the rate of mutation varies across the genome. This variability has profound impact in human evolution. Some of these differences in mutation rates have explanations, for others the reasons remain obscure [13]. In terms of regional differences in mutation rate within the genome, it is known that certain DNA types are more prone to undergo change. For instance, repetitive DNA sequences exhibit a tendency to engage in pairing with other identical or similar sequences located in different parts of the genome during the process of recombination or DNA exchange. DNA pairing and recombination take place during both somatic cell division or mitosis and gamete formation or meiosis. The resulting products of these types of illegitimate pairing and recombination are deletions and duplications in chromosomes (Figure 4.5). These kinds of mutations are frequent and sequences rich in repetitive sequences with significant homology among them experience higher frequencies of DNA alterations. Since this type of aberration could lead to deletions and duplications of coding or regulatory DNA sequences, the resulting genetic alterations could result in genetic diseases in subsequent generations if the germline is involved. In somatic tissue, it could be the genesis of localized abnormal cell growth and malignancy.

In somatic tissue, DNA recombination between repetitive sequences, even if the pairing is exact, involving the same repetitive element in homologous chromosomes could generate states of homozygosity of detrimental alleles. It is thought that many of these localized conditions of homozygosity are responsible for the genesis of tumors, some of them malignant capable of metastasis. In addition, regions within the genome that possess high concentration of certain repetitive elements, such as *Alu*s, experience elevated levels of DNA changes due to the mechanism of gene conversion (Figure 4.6). Further, *Alu* repetitive elements with their elevated content of G–C (63%) are more prone to experience high gene conversion rates. Genomic areas with a dense distribution of *Alu* elements would be more susceptible to this type of DNA change (see Chapter 5).

DNA regions made up of short tandem repeats also experience high rates of mutations. The mechanism

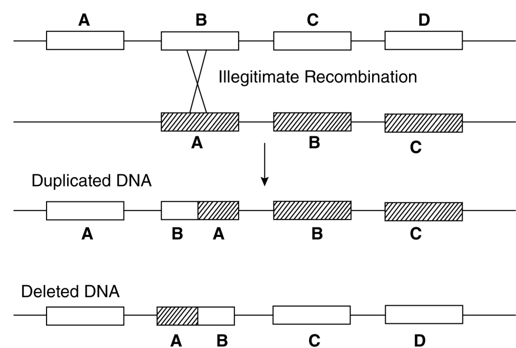


Figure 4.5 Illegitimate recombination leading to deletion and duplication.

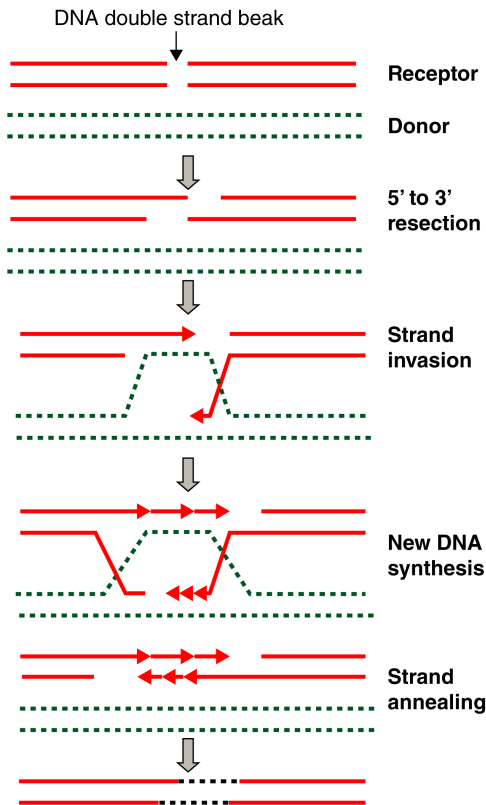


Figure 4.6 Mechanism of gene conversion. (*Source*: Azugje et al. 2007 [14]. Used under CC BY-2.0, http://creativecommons.org/licenses/by/2.0.) (See the Color Plates section.)

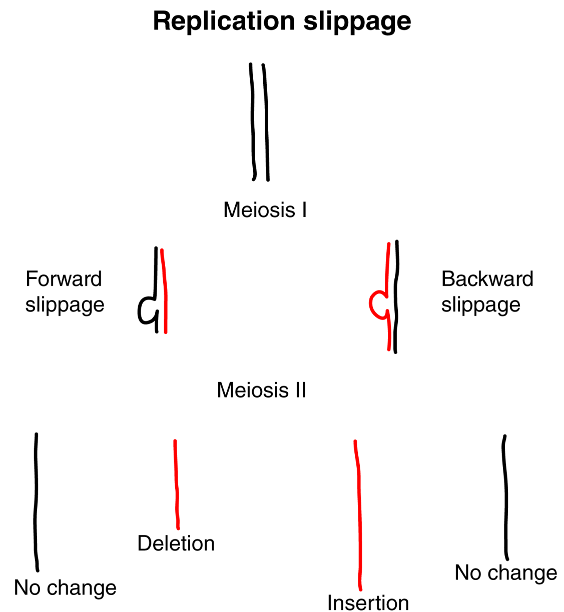


Figure 4.7 Replication slippage. (See the Color Plates section.)

responsible for these changes involves a process known as replication slippage in which the DNA polymerase replicates additional repeat units within the STR sequence (Figure 4.7) [15]. Some triplet STR repeat types experience such high mutation frequencies that the number of repeat units changes during the course of a few generations and are responsible for a growing number of human maladies. Most STR loci experience mutation rates involving alterations in number of repeats in the range of 10^{-3} to 10^{-4} and because of their high mutation rate they are hypervariable and exist in highly polymorphic states in our genome. STRs not only experience high mutation rates derived from replication slippage but also undergo internal point mutations like any other DNA sequence. As a result, these regions of DNA are hotspots of mutation and diversity. Due to this high mutation rate and the resulting hypervariability, these DNA regions are targeted by researchers, forensic scientists, and clinicians in population genetic studies, criminal cases, and as markers of genetic diseases, respectively.

Other sequence types including areas rich in thymidine and GpC islands experience high mutation rates for different reasons [16]. GpC dinucleotides, for example, are hypermutable because of their tendency to convert to TpGs (and CpAs). The rate of transition mutations (interchanges of a purine for another purine or a pyrimidine for another pyrimidine) is elevated at GpCs in humans and mammals because the cytosines in these dinucleotides are subject to methylation and methylated cytosine is unstable as it undergoes deamination to thymine, which, if left uncorrected, yields a $C \rightarrow T$ mutation. The rate of transversions (interchanges of purines for pyrimidines or vice versa) is also elevated by a few fold at GpC dinucleotides. The overall effect of GpC dinucleotides is a mutation rate 10-fold higher than other sites in the genome. Areas abundant in thymidine are also prone to mutate at higher frequencies by virtue of their propensity to develop thymidine bridges that can lead to single-strand breaks as well as deletion of nucleotides.

Other known hypermutable areas include telomeric and late replicating sequences. Telomeric regions are thought to promote alterations in chromosomes by virtue of the free DNA ends at the telomeres. It is postulated that DNA free ends facilitate the fusion of chromosomes

that could lead to aberrations, including translocations, inversions, deletions, and additions, at the time of chromosomal separation during cell division. It is likely that late replicating sequences experience elevated mutation rates due to the limited amount of time that they have to achieve any DNA repair resulting from damage generated during DNA replication.

Inheritance and environment

Inheritance and environment, the two inevitable contributors of variability, always combine to generate our biochemical and physical characteristics, known as the phenotype. The phenotype is visualized in the form of traits, such as eye color, skin color, insulin deficiency, and depression, to name a few. This interplay of inheritance and environment is oftentimes highly intricate making the assessment of the contribution of each very difficult. Although the relative inputs of these two components to variation differ from trait to trait, all characteristics are impacted by both. Environment can consist of various factors including our internal body milieu (physiology, biochemistry, etc.) and our external surroundings and exposure (the womb, geography, diet, climate, etc).

Traits can be primarily governed by a single major gene, or by multiple genes. The level of impact by the environment on the phenotype can differ from significant to minimal. Single-gene traits are traditionally known as Mendelian or unifactorial characteristics since we can follow their inheritance as the result of discrete single factors from generation to generation, according to the laws of Gregor Mendel. We will learn more about Mendel's laws in subsequent sections. Characteristics that fall into this comparatively simple mode of inheritance include sickle-cell anemia and cystic fibrosis, to name a few. In these single-gene trait examples, the environment does not always play a prominent role. At the other end of the spectrum, we find traits that are under the strong control of the environment. These are traits such as skin color, weight, and IQ (intelligence quotient) [17]. Most traits are dictated by different proportions of genetic and environmental factors in complex interactions leading to the final phenotype [18].

In addition to the difficulties in discriminating between genetic and environmental components of any given trait, the contribution of inheritance is obscured by the cumulative interaction of a number of genes, sometimes dozens or more. In genetics, this is known as epistasis. In epistasis, the expression of a gene is impacted by the activity of other modifying genes. Some of these individual genes affect traits minimally and there are so many of them that currently it is difficult to identify them individually and to discriminate between their expression from environmental factors. Most characteristics are polygenic or controlled by many genes, and although Mendel's law of segregation of alleles is in force for all, the impact of individual genes is obscured by the influence of their large numbers. In these multifactorial traits (traits controlled by multiple genes), some of the genes are of major influence and some of minor influence. In fact, the vast majority of human characteristics result from the cumulative effects of the environment and several genes. As advances in the molecular biology of gene expression and bioinformatics (computational genetics) progress, it is becoming clear that strict single-gene traits are rare.

The implications of the various degrees of influences by the environment and the complex contribution of a multitude of genes and alleles to traits are that different characteristics are better markers than others to assess diversity, both at the individual and at the population level. Characteristics can be morphological, physiological, behavioral, and biochemical. Yet, unless we are examining directly the molecule that stores and maintains inheritance, environment will contribute and alter the trait or phenotype. Take, for example, height. Height is a multifactorial trait affected by a number of genes and it is highly influenced by the environment. The polygenic nature of height is easily visualized in the bell-shaped distribution of different sized people in the population as well as in the diversity in the size of offspring of short and tall parents. Clearly, human height is not under the control of a single gene as Mendel observed in his pea plants. The environmental component of height is usually important since deficiencies in diet, vitamins, cofactors, injuries, and disease could limit the full height potential encoded in the DNA of any given individual. Thus, if we were to use height as a marker to study diversity, it will be impossible to ascertain precisely how much of the trait has been dictated by inheritance alone, and when comparing populations, how much of the differences only derive from the gene pool. Therefore, since environment varies constantly, the expression of a trait will change in any direction independent of the genetic makeup of the person and the observed variability will not relate necessarily to the genotype. This renders the trait less ancestry dependable,

and therefore less favorable, in studies of heredity and population diversity.

The consequences of this spectrum of environmental impact on traits are that, in general, characteristics under the control of inheritance alone are desirable in studies designed to assess genetic relationships and ancestry. Furthermore, since characteristics are expressed at the molecular (e.g., enzyme activity), tissue (e.g., amount of actin and myosin in muscle), and organismal (e.g., skin color) levels, with an intrinsic environmental input, it is best to examine and score markers directly at the DNA level. Unfortunately, easily accessible characteristics are morphological or behavioral at the organismal level and are usually multifactorial. Traits at the organismal level are readily observed in the field and have been employed since the 19th century in anthropomorphic studies. Even today morphometric parameters such cranial size and shape are being examined and employed in phylogenetic studies. There are instances, such as in fossil remains, in which these traits are all we have accessible for examination and they provide crucial data. For example, many ancient human remains do not yield useful DNA for assessing variability since nucleic acid suffers from gradual degradation and base modification with time. In those instances, tissues such as bone, teeth, and hair are the only material available and morphometric data are routinely collected and analyzed (see Chapter 2).

Selection works on the phenotype

In the process of natural selection, the environment is blind to the actual DNA constitution or sequence. The environment selects for or against specific variants of a gene, the alleles, by allowing the fittest organisms to survive based on their phenotype. The phenotype may represent behavioral, anatomical, or physiological traits. For example, when a point mutation occurs in a gene, it may change the amino acid coded in a given position of the peptide. This, in turn, could alter the structure and/or function of the peptide. If the resulting protein happens to be an enzyme, this change may render it partially or entirely inactive. This compromised or lack of enzyme activity could be reflected as an accumulation of a product in a given metabolic pathway, affecting the survival of the individual. It is at this point that natural selection acts.

As mentioned earlier, most human traits are under the control of many genes and environment, multifactorial inheritance. As our knowledge of our genome increases, the number of traits clearly under the control of single genes become less numerous. As the structure–function relationships in our genomes are being uncovered, intricate and subtle interactions among genes are found. It turns out that when traits are investigated at the molecular and cellular levels, complex interactions involving products from different genes are observed. Clearly, single-gene phenotypes are not the rule but the exception. Even traits traditionally thought to be under the control of single genes have subsequently been shown to involve multiple genes. Take, for example, eye color. The old paradigm of a dominant allele capable of producing the pigment melanin that is then deposited on the iris of our eyes and a recessive variant sequence incapable of making the protein is inaccurate [19]. Even a casual inspection of eye color in a number of people illustrates that the eye color phenotype exhibits a gradient of colors and shades (Figure 4.8). Colors range from deep brown, light brown, tones of green to blue and gray. It seems as if no two eye colors are the same. Unifactorial genetics cannot explain these observations. New advances in molecular biology have detected a number of different major genes that are responsible for eye color. In addition, a number of independent minor or modifying genes as well as the environment (e.g., age and angle of incoming light) contribute to the trait creating the observed range of diversity in color and shade.

In multifactorial inheritance, natural selection is acting on all the genes that contribute to the characteristic in question and not on individual genes. In other words, natural selection would favor individuals exhibiting the optimal phenotype for the given environment, independent of the number of genes responsible for the

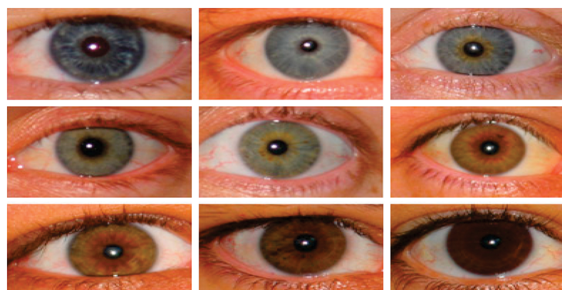


Figure 4.8 Gradient variation of eye colors. (*Source*: Reproduced from Sturm & Frudakis 2004 [20] with permission of Elsevier.) (See the Color Plates section.)

characteristic. The implications of this are that if any of the genes responsible for a function is mutated, rendering its protein product partially or totally inactive, the phenotype will be affected and natural selection will act on it. Furthermore, in a hypothetical scenario, environmental changes could trigger selection against a trait. In such case, all the genes responsible for the phenotype will be selected against, reducing with time, the number of individuals expressing the nonbeneficial trait and all the genes responsible for it, independent of the alleles they possess. Another example involves cascades of enzymes responsible for producing certain gene products in metabolism. In a metabolic pathway such as glycolysis, a number of genes work in tandem to generate a number of products. Mutations affecting any one of the enzymes that work in sequence would impact the levels of the end products. Therefore, the inactivation of any of the genes coding for intermediary enzymes would halt the products down the pathway and the phenotypes derived from those products would be affected. This is an example of epistasis where the inactivation of one gene in a pathway affects the expression of genes downstream in the pathway and results in the accumulation of intermediary products. Medically, this situation makes it difficult to assess which gene within a cascade is actually responsible for the abnormal phenotype in a given individual. In other words, any given function can be stopped or reduced by mutations in any one of a number of interacting but independent genes. Another ramification of multifactorial inheritance is that deleterious mutations in any of the genes responsible for a trait could affect the fitness of the individual even if the other genes produce functional gene products. Therefore, if a population possesses an allele or variant of a given gene of a multifactorial trait under negative selection pressure, all the individuals with the mutated sequence would be selected against, independent of whether the alleles of the other interacting genes are under positive, negative, or no selection pressure.

The impact of selection

A number of genetic sequences are clearly under selection pressure. For example, DNA that dictates the synthesis of enzymes required for metabolism and structural proteins that are building blocks of organelles, cytoskeleton, nuclear matrix, and ribosomal, spliceosomal, or transfer RNA, are subject to selection. Considering that organisms have been evolving for over 3 billion years, it is not likely that random mutations in functional genes are going to improve fitness. The reason is that fine-tuning of basic biological processes for billions of years is bound to generate efficient mechanisms to guide the business of life. For this reason, most random mutational change would be neutral, if not deleterious or even lethal.

Certain mutations that affect genes that control major developmental events are capable of dramatically altering the phenotype. These changes may affect the behavior, anatomy, or physiology of individuals leading to reproductive isolation. Also, mutations in master genes responsible for key developmental steps or DNA coding for essential molecules that regulate transcription, RNA processing, or translation are expected to impact the phenotype of individuals often generating conditions incompatible with life.

Some of these mutations are so detrimental that they are never seen in an organism since the resulting embryos die shortly after conception and are reabsorbed into the uterine wall undetected. These highly deleterious mutations that lead to the early death of the embryo have little impact on the energy expenditure of parents or care takers. These lethal mutations that result in early termination of pregnancy may be *de novo* and the impact on the genetic health of the population would be minimal. A greater contribution to the genetic load of populations is incurred when deleterious alleles are carried by heterozygous parents and masked by the wild-type alleles and only in homozygous children the trait is fully expressed. Deleterious mutations that allow the prenatal or postnatal survival of the individual would result in a higher genetic load due to the greater energy drain on the population.

Mutations and frequencies of alleles are expected to change not only as a result of an increment in environmental insults such as mutagens but also as a function of alterations in selection pressure. A number of genes subject to selection exist in a state of balanced polymorphisms in which opposite selection forces have established a state of dynamic equilibrium. A classical example of this condition is sickle-cell anemia [6]. It is known that the sickle-cell anemia trait is under negative selection pressure resulting from the physiological disadvantage that homozygous individuals have since the mutant defective beta-hemoglobin, coded by the variant DNA, is a poor transporter of oxygen and CO_2. Yet, this same variant allele provides heterozygous carrier individuals

with a certain protection from malaria because the malaria parasite cannot reproduce well in people who possess the mutant hemoglobin. These two selection forces, acting in opposite directions, usually reach a state of dynamic equilibrium in which the frequencies of the beta-hemoglobin wild-type allele and the sickle-cell variant remain constant. If, on the other hand, the environmental pressure changes, decreasing, for example, the population of parasites, it is likely that the selection for the sickle-cell allele would decrease. This, of course, would affect the previously established state of equilibrium as the negative selection pressure increases as the benefits for the variant trait decrease since the danger of being infected with malaria decreases. The environment is never constant. The biological and physical universe that surrounds us changes and in doing so it affects the conditions necessary for survival of organisms including humans. Selection pressure on specific variants and phenotypes can be relaxed or intensified depending on environmental demands. Thus, as a function of time, the frequencies of specific alleles are bound to fluctuate. Even in the absence of genetic equilibrium, the frequencies of certain alleles are expected to increase when positive selection pressure for them increases and decrease when negative selection increases.

In addition to the expected fluctuations in the environment, changes in the genetic constitution of individuals can bring about changes in levels of positive or negative selection pressure. For example, in instances of gene duplication in a diploid organism, one of the duplicated sequences is expected to experience reduced selection pressure since only one copy of the gene is required for the normal physiological function and survival. Therefore, one of the duplicated copies is free to mutate without being subject to negative selection. Under these conditions, these sequences can accumulate mutations rapidly, increasing diversity. It is this mechanism, in fact, that is responsible for the creation and evolution of families of genes that expand the physiological possibilities and genetic complexities of populations. Thus, gene duplication is thought to be a significant force in providing for novel sequences and functions as well as allowing for new evolutionary venues and speciation. Examples of this process are many. The hemoglobin gene family is a classic example in which a number of related proteins have been generated over time by the process of sequence duplication, mutation, and relaxed selection pressure [21]. Oftentimes, although not always, these duplication events produced duplicated genes in tandem, head to tail, along the DNA with the most closely related sequences found next to each other. The resulting copies are found in close proximity in the genome in what is known as complexes. In the case of the hemoglobin complex, a whole spectrum of genetic relatedness among its family members is observed reflecting ancient as well as recent duplication events in chronological hierarchical order. Considering the rapid rate at which diversity accumulates in the duplicated sequences, it is likely that gene duplication may facilitate a mechanism that allows for the phenomenon known as punctuated evolution in which dramatic changes in the phenotype occur sometimes leading to reproductive isolation and speciation in a short amount of time. The timescale at which these phenotypic changes take place is not explainable by traditional neo-Darwinian theory that requires geological time to evolve novel characteristics. It is thought that sequences free to mutate in the absence of selection are potentially capable of producing genes with different and new functions that could alter the phenotype in a relatively short period of time, sometimes leading to fast or punctuated speciation (more about punctuated equilibrium, gradualism, and copy number variation is covered in Chapter 5).

The function of the vast majority of our genome remains, at best, unclear. These sequences include spacer DNA located in between genes, highly repetitive sequences, and intronic elements. Recent efforts in bioinformatics by the Encyclopedia of DNA Elements (ENCODE) project initiated in 2007 indicate that most of the DNA previously thought to be nonfunctional is in fact involved in intricate interactions impacting the phenotype [22,23]. This initiative has discovered that more than 80% of our genome is functional (see Chapter 2) [24]. Yet, the actual specific functions of most of these sequences that seem to be functional are still unknown. It is likely that some of our DNA functions as spacer keeping regulatory and structural protein- or RNA-coding genes at a required distance from each other to allow appropriate molecular interactions and gene expression. In instances like this, the actual nucleotide sequence is not subjected to selection, just the distance among DNA elements. Therefore, substitution of nucleotides would be selectively neutral.

Lack of selection pressure on DNA sequences is paramount in the study of phylogenetic relationships. Selection, whether it is natural or artificial (human-driven), acts on certain phenotypes allowing the corresponding variants or alleles to be transmitted, or not, to subsequent generations. Since the fittest organisms enjoy a reproductive advantage, producing more offspring, the DNA

variants responsible for this advantage would preferentially pass on to future generations, likely increasing the frequency of the beneficial alleles in the population. These alterations in allelic frequencies resulting from positive or negative selection on populations may obscure the true phylogenetic relationships among groups of organisms, artificially making them look more or less genetically related. Therefore, selection potentially could mislead in signaling affiliations that do not exist or eclipsing true relationships. An added complication to the issue of selection is that since it acts on the phenotype and most human traits are multifactorial in nature it may be difficult at times to assess if a given gene, out of many, is under natural selection since its overall impact on the trait may be subtle. Especially in the case of minor or modifying genes that exert minimal impact on a given phenotype, their effects on a phenotype under selection pressure could be challenging to ascertain as well as their neutrality or lack of it.

Cultural expressions as markers of ancestry

In addition to molecular markers such as DNA, RNA, and proteins as well as morphometric indices and behavioral characteristics, human creativity has generated a cultural world that oftentimes is population-specific. Take, for example, language and its derivative, writing. Currently, linguistics is routinely being used to ascertain ancestral relationships among human populations with a variable degree of success. A well-studied case is the distribution of the Austronesian language family in East Asia and Oceania [25]. In fact, one of the major lines of evidence supporting the origins of the Austronesian expansion in Southeast Asia, and specifically what is currently the island of Taiwan, is linguistic data. The extreme linguistic diversity of Austronesian languages in the medium-sized island of Taiwan (previously known as Formosa) represents one of the strongest evidence pointing to Taiwan as the source of this diaspora. Furthermore, the linguistic signals left behind as the migrants navigated from island to island are studied to ascertain the routes and timing of the spread (see more details in Chapter 9).

Artistic and folkloric expressions are also signatures of ancestry and relationships among human groups. The Austronesian expansion is again a good example with the Lapita pottery tradition linking the people who participated in this dispersal. Other cultural parallelisms are also detected among Austronesian groups, including dance, musical instruments, and sculptures. A casual observer of Pacific populations would notice the similarities in the Hula-like dances practiced by natives of the Cook, Tonga, Samoan, Hawaiian and French Polynesian archipelagos, and New Zealand as well as the hollow slit drums made by carving the inside of tree trunks. The traditionally undulating and rhythmic hip and arm–hand movements performed at the beat of the drums are characteristic of all of these insular populations. The typical resonating sound of the drums is a hallmark of populations from the Philippines to remote islands of Oceania and their Moai-type statues are artistic expressions linking Polynesian populations.

Domestication and agricultural practices are also informative regarding affinities among humans groups. For example, during this Austronesian trek, humans migrated not only with their families, friends, and neighbors but also with samples of the plants and animals that allowed them to survive in the new land. Although the motivations that drove these people to sail into the unknown open sea are not clear, considering the next island was often out of view, several thousand kilometers away, taking provisions in the trip was paramount. Genetic studies on domesticated dogs, pigs, chickens, and Asian rats, for example, have provided useful data to trace the movement of people from East Asia into the Pacific. The transportation of sweet potato in the other direction, from South America to Polynesia, although a controversial notion, is indicative of the power of human expression as markers of human ancestry and interactions.

Inherited in all of these studies that employ cultural characteristics as markers for generating phylogenies is the possibility of acculturation and not gene flow as the reason for the observed parallelisms between or among populations. The issue of acculturation is a reoccurring theme when considering cultural traits. The practice of agriculture and animal domestication is a well-known cultural characteristic thought to have originated in different parts of the world, at different times, independently. One site was Anatolia in present-day central Turkey. The resulting dramatic change in human existence represented a revolution involving everyday survival to the structure of societal institutions. Humans were not required to practice hunter-gatherer subsistence and had the technology to cultivate the land and raise animals allowing to live in larger groups and coexist

in high population densities in homesteads and cities. From its genesis in Anatolia, the agricultural revolution spread in all directions. Yet, it is its impact in Europe that has received more attention from the scientific community. To date, the issue of how much acculturation or transmission of ideas as opposed to DNA flow by migration and procreation was involved is still debated. The fundamental premise of using cultural characteristics, such as agricultural practices, as markers for assessing phylogenetic relationships is that transmission of DNA and not culture is the cause for the similarities.

Congruency among marker systems

When studying diversity, it is always reassuring to consider the results and subsequent conclusions derived from different marker systems. In instances of congruency of results, it adds weight to the conclusions. If, on the other hand, the different types of data contradict, partially or completely, it would clue the investigators of the potential for additional important variables affecting the outcome of the experiments. Lack of parallelism in experimental outcomes may signal experimental errors in one or more of the marker systems' protocols or fundamental biological differences in the variables that are being tested by the various marker systems. In either case, the assessment of congruency, or the lack of it, is a powerful tool since the process provides for data verification, or, probably more important, it may signal technical and/or scientific problems that otherwise would go undetected. If after exhaustive checking of experimental protocols, it is confirmed that the procedures were correctly performed, then the question is whether both marker systems are testing the same phenomenon. A common source for lack of parallelism is the differences in the molecular clocks of various sequences as well as the impact of selection pressure on some markers. Molecular clocks employ nucleotide or amino acid data to estimate rates of molecular change and time when speciation events took place. For example, let us consider the case of nonparallelism observed between mitochondrial and Y-specific DNA markers in some studies. Since mitochondrial DNA is a signature of maternal inheritance and Y chromosomal markers are indicative of paternal transmission, it is reasonable that sexual biases and sex-specific sociocultural practices would impact these uniparental marker systems differentially. It is unlikely that mitochondrial Eve and Y-chromosomal Adam lived at different times but their genetic clocks tick at different rates. In addition, the two systems exhibit several basic biological variables including different amounts of reiterated sequences and selection pressures that could impact experimental results.

Does junk DNA exist?

Although less used in current scientific literature, the term junk DNA is still employed by some to describe portions of the genome that somehow remain in the gene pool without seemingly performing a function, maintaining themselves selfishly as molecular parasites. According to neo-Darwinian theory, this idea of sequences existing, generation after generation, as cells expend energy in maintaining them, is counterintuitive. This concept is particularly difficult to rationalize when we consider that some consider this portion of the genome to be the vast majority of our DNA. This notion has been recently challenged based on the results of the ENCODE project that have uncovered that the vast majority of our genetic material is functional [26]. Even some repetitive sequences were identified as playing a role. Although a number of recent analyses have challenged the high percentage of functional DNA reported by ENCODE, the conclusions of the initiative have cast serious doubts on the existence of large amounts of junk DNA. For many years, it was argued that only about 1% of our DNA had a clear function. Large areas of our genome are made up of repetitive sequences, some very simple. Some of these sequences are highly reiterated displaying hundreds of thousand of copies such as short tandem repeats, termed satellite DNA. These sequences are referred to as satellite DNA because they sediment away from the nonrepetitive main band DNA during high-speed centrifugation. Some are known as middle repetitive DNA, with thousands of copies. The short and long interspersed sequences (SINEs and LINEs, respectively) are examples of these elements. Examples of these repetitive sequences in humans are the SINE *Alu* and the LINE L1 families of repetitive elements [27]. Although humans possess a number of families of middle repetitive sequences, the *Alu* family of elements is probably the most well known. *Alu*s actually occupy about 10%, by mass, of our genome (the functions of transposons in human evolution are discussed in more detail in Chapter 5). The low copy number reiterated

sequences range from as low as several members to hundreds. Of course, these demarcations in copy number are somewhat arbitrary. In addition, we possess long stretches of DNA separating known coding sequences that are referred to as intergenic DNA.

In our genome, middle repetitive elements such as the SINE *Alu* are reiterated to about 1 million copies per diploid genome. It is not clear how organisms can afford to keep duplicating and maintaining, generation after generation, these sequences without a benefit to the species. In the case of *Alu*s, it is important to keep in mind that these elements have been traditionally considered as useless and junk, and yet they have increased in numbers from one original element to 1 million copies in about 80 million years. *Alu*s are though to derive from a single mutated signal recognition particle gene (the gene that codes for the small RNA moiety of the nucleoprotein complex responsible for internalizing nascent proteins into the lumen of the rough endoplasmic reticulum). They are transcribed from split internal regulatory transcription promoter sequences, but the RNAs produced do not contain open reading frames for the coding of proteins and their transcripts, for the most part, have no known function. How such a mutated sequence was able to expand in the absence of positive selection pressure and in time become about 10% of the human genome? Are we simply naive about the function of these repetitive elements? Although different *Alu* copies have been found to have various types of functions, it is not known what the vast majority of these elements do, or whether they do anything at all. It seems that these reiterated sequences are constantly being duplicated and deleted. It is likely that, currently, they exist in a state of dynamic equilibrium in which no further increment or net change in their numbers is occurring. That is, the number of newly duplicated *Alu*s equals the number of repeats deleted. It is likely that the genetic load on the population for keeping the deleterious copies prevents additional expansion of the elements.

Simple highly reiterated sequences are made up of short segments of DNA a few nucleotides long repeated in tandem, head to tail, in the same orientation, thousands of times (e.g., ATGGATGGATGG . . .). The simplicity and high degree of reiteration of these highly repetitive sequences are such that it is difficult to conceive of a possible function for them other than keeping an appropriate spacing between regulatory and coding sequences to allow, for example, transcription factors to bind and interact properly in space with specific DNA binding sites and other protein factors. Precise, spacing-sensitive molecular interactions between proteins and their DNA binding sites as well as among regulatory proteins are, in some instances, essential for proper control of transcription and RNA processing. These simple sequences are not transcribed. In the past, it has been suggested that these sequences function as sources of genetic material for the creation of novel genes as well as a DNA "garbage disposal" for mutated nonfunctional genes. The simple nature and the resulting mode of mutation and duplication by replication slippage argue against this last possibility.

Since the completion of the Human Genome Project in 2003, we had a rough map of the structure of the human genome. As a map, it provided us with just the DNA sequences and their locations, much like the location of homes within a city or neighborhood. Although it represents a pivotal step in human genome research, clearly the biggest task awaits future generations of investigators in trying to ascertain the structure–function relationships and interactions (including with the environment) of our genetic material. This task will keep researchers busy for decades to come, and will constitute one of the main challenges for medicine and the treatment of human maladies. In other words, now that we have a map of the city, the functions of the different homes need to be assessed, that is, which house is the tailor's home, the mechanic's home, and the grocery store. Results from the Human Genome Project show that only about 23,000 genes were identified in our species, far less than the hundred of thousands previously anticipated. Also, the results of the Human Genome Project indicated that about 90% of our DNA had no known function and only approximately 1.5% encodes for instruction for proteins. As mentioned earlier, in 2007 a consortium of researchers initiated the ENCODE project. This international effort mines existing empirical data from DNA sequences and using advanced bioinformatic tools attempts to ascertain their function. The results obtained point to an unexpected large portion of the genome, previously of no known function, involved in a myriad of complex intergenic interactions. The data illustrate a multitude of regulatory DNA sequences, in regulatory networks, working together to fine-tune proper gene expression of protein-coding genes in time and space, in different tissues during development. It was found that about 4 million regulatory switches, distributed all over

the genome, are responsible for this complex regulatory system of human life. ENCODE estimated that over 80% of the so-called junk DNA in our genome performs some sort of function and it is likely that as our basic knowledge of our genome augments this proportion will increase (see Chapter 2). This discovery is in sharp contrast to the low number of genes previously assessed to be present in our genome. In terms of diversity, ENCODE found that a lot of the variability of our species resides among these regulatory sequences. Also, an important implication of these results to evolutionary biology is that sequences previously thought as neutral could be under selection pressure and their values as genetic markers should be carefully reevaluated.

In light of the stated information, can we now state "junk no more"? The question of whether any junk selfish DNA exists and is maintained with no function remains unanswered since a definitive answer would require a better understanding of the intricate structure and function relationships of our genome to be able to rule out any subtle roles. Of course, nonfunctional sequences are constantly being generated resulting from *de novo* mutations. Thus, it is likely that, since the genetic material is constantly mutating, a portion of nonfunctional DNA is always present and some may persist for a period of time before it is deleted from the genome, especially if it is selectively neutral. Much the same way it takes to dispose of garbage. Also, it is possible that this hypothetical nonfunctional DNA could serve as raw material for evolution, like duplicated sequences, and evolve to serve a role by itself or coevolve with other nearby functional sequences. No doubt this DNA would constitute a genetic load for the population that may be compensated by the benefits of providing for raw genetic material for generating novel genes and providing genetic flexibility in a changing environment. In the context of this discussion, it is important to keep in mind that there are still segments of DNA for which no function is known. Does that mean that they are junk? Or it could simply mean that we do not know their function yet? More on this topic will be explored in Chapter 5.

How genetic diversity is studied?

Humans have been selecting for desired traits in plants and animals during the process of domestication and artificial selection since the late Paleolithic. Mitochondrial DNA evidence indicates that ancient humans started domesticating the gray wolf as early as 33,000 years ago in the Altai region of Central Asia [28]. Domestication and artificial selection by humans intensified during the agricultural revolution in what is now Anatolia about 10,000 years ago. Yet, it was not until the early 1800s that empirical observation and experimentation involving inheritance commenced. Early scientific experiments relied on anatomical characteristics as makers of inheritance. Investigators at the time had no clue on the chemical nature of these factors that seemed to be transmitted to future generations. Mendel conducted his seminal experiments in the vacuum of not knowing what was being transferred to subsequent generations in his pea plants. Physical characteristics are still being used to follow inheritance. Yet, currently, a number of additional markers, expressed at different levels of the phenotype (e.g., molecular, cellular, tissue, and organismal), are employed to study genetic diversity. For example, in addition to anatomical features, investigators routinely score tissue or cellular morphology as well as levels of enzymatic reactions.

Beginning in the mid-1900s, a number of technological advances in biochemistry and immunology allowed for the typing of genetic immunological markers in the form of blood factors such as the A, B, O, and Rh systems. These marker systems proved to be informative but their known functions and lack of selective neutrality compromised their usefulness in phylogenetic studies. More recently, starting in the 1960s, genetic diversity has been visualized by looking at protein polymorphisms. A substantial body of genetic data has been generated in the form of electrophoretic separation of proteins and enzymes in a gel matrix. In this line of investigation, proteins with various net electrostatic charges migrate at different rates within the gel in a closed electric field. Subsequently, protein-specific stains are used to visualize the location of the proteins within the gels. In this way, proteins with different net electrostatic charges, sizes, and shapes, due to their amino acid constitution, will appear on the gel at various distances from the origin, where the extracts were initially deposited (Figure 4.9). Enzymes are particularly suitable for this type of analysis because the location of the variant proteins can be easily seen by coupling the specific enzymatic reaction with a reactant that produces a colored product in the location of the enzymes. The different forms of the enzymes are referred as allozymes or isozymes. There are several limitations to this

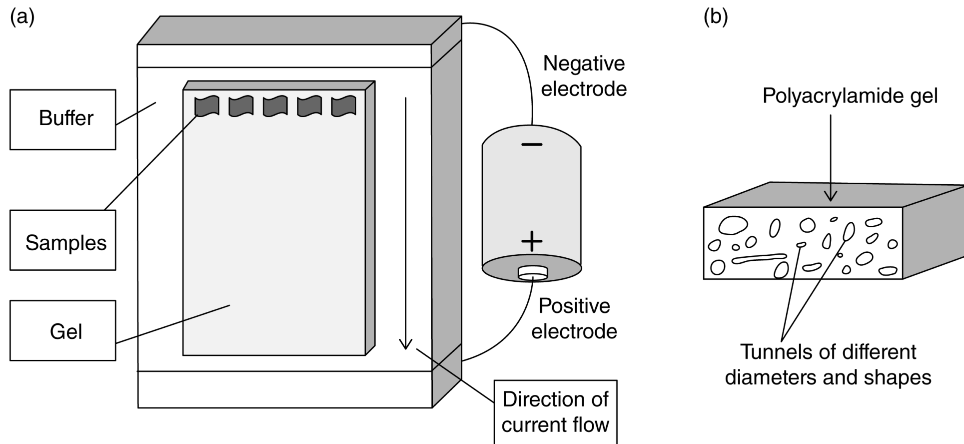


Figure 4.9 Gel electrophoresis. Molecules such as proteins, DNA, and RNA travel with the current (Panel A) through the gel matrix. Different size molecules move through the tunnels in the gel (Panel B) at different rates depending on their size and net electrostatic charge.

methodological approach. First, genetic diversity is being detected several steps (transcription, RNA editing, translation, protein modifications) away from the actual source, the DNA, and as such some variability goes undetected due, for example, to silent mutations (mutations that do not change the amino acids of proteins because several nucleotide triples code for the same amino acid). Also, as proteins with functionality, this genetic diversity is under selection pressure.

In the 1970s, investigators began to mine genetic diversity by looking *directly* at the DNA. Initially, the approach was to specifically modify each of the four nucleotides in separate chemical reactions. In other words, chemicals are used to alter the nitrogenous base of the nucleotide adenine and a different chemical did the same for the nitrogenous base of guanine and so on. These nucleotide-specific modifications make the phosphodiester backbone of the DNA, at the site of the alteration, susceptible to breakage. This selective truncation of sequences produces a ladder of DNA fragments of different sizes that are directly related to the order of specific nucleotides along the DNA molecule. The sequence is read from one end of a gel to the other subsequent to electrophoresis (Figure 4.10) Since this methodology relies on chemicals to specifically modify the DNA, it is referred to as chemical sequencing and is particularly useful for assessing the sequence of short DNA fragments. In this technique, radioactive ^{32}P is used to tag the end of one of the truncated DNA fragments to allow visualization on a X-ray film.

Shortly after the development of chemical DNA sequencing, enzymatic sequencing (Sanger sequencing) was introduced. This approach employs analogs of nucleotides (dideoxy nucleotides or ddNTPs), one for each of the four nucleotides; when incorporated by the enzyme DNA polymerase, these analogs stop additional additions of nucleotides and in doing so truncate the extension of the nascent DNA strand. In this methodology, each nucleotide analog is labeled with a different fluorophore that allows independent detection of

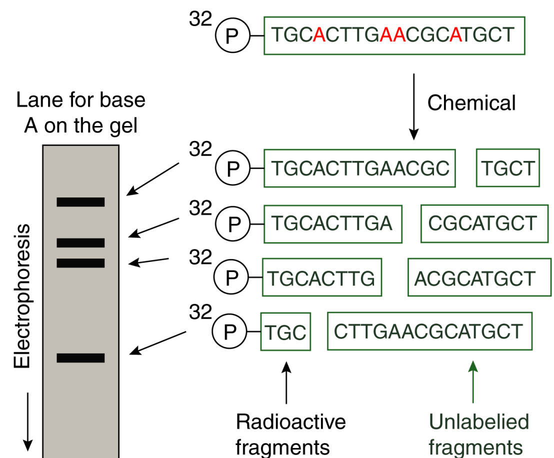


Figure 4.10 Chemical sequencing. (*Source*: Anna Kaksonen. http://wiki.biomine.skelleftea.se/biomine/molecular/index_14.htm. Used under CC BY-4.0, http://creativecommons.org/licenses/by/4.0/.)

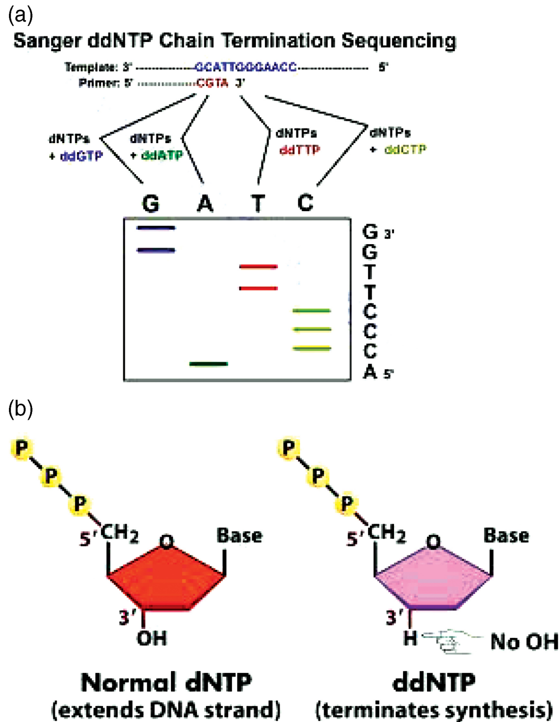


Figure 4.11 Enzymatic sequencing. (*Sources*: (a) Adapted with permission of themedicalbiochemistrypage, LLC. (b) Reprinted from Freeman 2005 [29] with permission of Pearson Education, Inc.) (See the Color Plates section.)

each of the four nucleotides at the end of the truncated strand. Effectively, this procedure, although different in methodology to chemical sequencing, achieves the same outcome, a continuous ladder of DNA fragments that is read as a nucleotide sequence (Figure 4.11). Sanger sequencing was automated and was responsible for the sequencing of the human genome. Currently, a number of novel procedures are being developed as “next-generation sequencing” (NGS) systems, which are explored in Chapter 5.

Epigenetic diversity

Epigenetics is the study of phenotypic differences that are not the result of heritable DNA changes [30]. It includes changes in levels of RNA transcription due to DNA methylation and/or chromatin conformation as well as post-translation modifications of proteins [31]. The term epigenetics can be used to describe the study of stable, long-term alterations in the transcriptional potential of a cell that are not heritable. Thus, short-term changes in metabolism such as increases and decreases in enzyme activity as well as fluctuations in biological feedback mechanisms resulting from hormonal levels are not considered epigenetic characteristics. Perhaps the most obvious type of epigenetic change occurs during cellular differentiation. During development, totipotent stem cells, capable of specializing into all cell types, become the pluripotent cell lines of the embryo, which in turn specialize into the fully differentiated cells of hundreds of tissues. Epigenetic alterations may last for multiple generations even though they do not involve changes in the DNA sequence of the organism and in that sense mimic genetic alterations.

Methylation of DNA is one of the best known mechanisms by which epigenetic changes occur. Nitrogenous bases are known to experience methylations at specific nucleotides (A and C). These locations could be within sequences that code for RNA or proteins as well as in regulatory DNA elements. Generally speaking, methylation of transcription regulatory DNA promoters and enhancers tends to suppress the production of RNA of the nearby structural genes, usually located downstream, under their control. It seems that hypermethylation interferes with the binding of proteinaceous *trans*-acting factors (transcription factors that help to promote gene activity) to the promoter/enhancer elements (DNA sequences responsible for signaling gene activity) interfering with RNA production by the transcription initiation machinery. Therefore, most of the time, hypermethylation of regulatory protein binding sites brings about a decrease in mRNA production that in turn depresses the levels of the corresponding proteins. Hypomethylation of controlling elements usually has the opposite effect, increasing transcription initiation and gene expression. The consequences of methylation of nucleotides within coding reading frames are less clear. In addition to DNA methylation, recently it has been demonstrated that methylation of mRNA could contribute to epigenetic characteristics such as human energy homeostasis. For example, aberrant obesity-associated methylation of mRNA in humans has been demonstrated [32].

Another example of epigenetic mechanism involves the changes in chromatin (DNA–protein complexes). Chromatin in eukaryotes exists as a complex of two types of proteins and DNA. The two kinds of proteins are the histones and non-histones. Within the non-histone category, proteins such as transcription

regulatory factors and proteins involve post-transcriptional processes are included. Histones, on the other hand, are structural proteins that are chemically basic (possess a high proportion of amino acids with positive charges at physiological pH) capable of binding strongly to the DNA and in doing so provide structural support to the double helix. There are five major types of histones: H1, H2A, H2B, H3, and H4. To perform their function of keeping proper chromatin structure, these proteins form octameric structures called nucleosomes made up of two of each of the H2A, H2B, H3, and H4 types. H1 interacts with adjacent nucleosomes linking them together on the DNA to form a unique structure reminiscent of beads on a string. DNA is known to wrap about one and a half times around the octamer. It is known that histones undergo post-translational chemical modifications including methylation, acetylation, and phosphorylation of their amino acids' side groups that alter their interactions with the positively charged backbone of the DNA. Some of these modifications of histones change their three-dimensional conformation making the associated DNA looser and more accessible to the transcriptional machinery allowing RNA production, gene expression, and phenotypic changes. Therefore, this kind of post-translational modification indirectly affects noninheritable phenotypic expression by changing the structure of chromatin.

Certain patterns of DNA methylation and histone modification in humans are known to be responsible for genomic imprinting. In genomic imprinting, only one allele (either the one transmitted by the mother or the one provided by the father) becomes transcriptionally active in the newborn. This is the result of sex-specific methylation occurring in the parents' DNA. In other words, certain genes can be expressed in a parent-of-origin-specific manner. These modifications occur differentially in the germline according to the sex of the individual and persist in the somatic tissue of the next generation. In humans, a number of epigenetic diseases occur only when the affected individuals inherit the deleterious allele from the father. A well-known example is the Prader–Willi syndrome. Conversely, the Angelman syndrome is manifested when in a different gene, located very close in the same region, the mutated allele is transmitted by the mother. Both of these medical conditions are characterized by severe intellectual and developmental disabilities. In order for genetic imprinting to occur, these epigenetic alterations need to be reversed in the germline and then reestablished in a sex-specific mode before meiosis.

All of these types of noninheritable changes can affect the phenotype of individuals and yet this kind of variability not only is useless for the study of phylogenetic relationships but can also be misleading by obscuring the true genetic nature of the traits. It is likely that a number of genetic characteristics, some responsible for human maladies, are governed by sequences subject to genetic imprinting. It is possible that the uniparental restrictions of expression seen in genetic imprinting help alleviate the genetic load on the population.

Review questions and exercises

1. Do you agree with Friedrich Nietzsche's quote at the beginning of this chapter?
2. Do you agree or disagree with the statement "All characteristics possess a genetic and an environmental component, yet the relative degree of their contribution varies with the trait." Explain.
3. Explain the concept of balanced polymorphism. Use a real case scenario.
4. How mutations of somatic cells and germline cells affect survival of individuals and populations?
5. To what degree mutations are random or selective? Provide examples.
6. Contrast between epigenetic changes and imprinting.
7. Defend or argue against the statement "Mutations are not entirely random."
8. Do unifactorial traits exist in humans? Name one case of unifactorial inheritance indicating why it is clearly controlled by a single gene.
9. Explain how the environment can obscure the impact of minor or modifying genes.
10. How natural selection affects the frequency of a specific allele? Use the sickle-cell trait in your explanation.
11. Why duplicated genes promote rapid evolutionary change? Illustrate using a real case scenario.

12 The ENCODE project has uncovered that more than 80% of our genome is functional. Explain why these data have revolutionized our understanding of structure–function relationships in our genome.

13 Cultural characteristics as extensions of human existence and expression are utilized as markers for migration and dispersal. Give examples from linguistics and domestication of plant and animals that have provided information on human migration.

14 Does junk DNA exist? Provide clear cases of DNA without function in our genome.

15 Describe the similarities and differences of chemical and enzymatic DNA sequencing.

16 To what extent the Prader–Willi and Angelman syndromes represent genetic and epigenetic differences?

References

1 Nietzsche FW. 1984. *Human, All Too Human: A Book for Free Spirits*. Translated by Faber M, Lehmann S. University of Nebraska Press, Lincoln, NE.
2 Freese E. 1959. The difference between spontaneous and base-analogue induced mutations of phage T4. *Proc Natl Acad Sci USA* **45**:622–633.
3 Ohta T. 2002. Near-neutrality in evolution of genes and gene regulation. *Proc Natl Acad Sci USA* **99**:16134–16137.
4 Charlesworth BM et al. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**:1289–1303.
5 Dobzhansky T. 1970. *Genetics of the Evolutionary Process*. Columbia University Press, New York.
6 Allison AC. 1956. The sickle-cell and haemoglobin C genes in some African populations. *Ann Hum Genet* **21**:67–89.
7 Kerr JB, McElroy CT. 1993. Evidence for large upward trends of ultraviolet-B radiation linked to ozone depletion. *Science* **262**:1032.
8 Mackie CO, Scriven PN. 2002. Meiotic outcomes of three-way translocations ascertained in cleavage-stage embryos: refinement of reproductive risks and implications for PGD. *Eur J Hum Genet* **10**:801–806.
9 Lee H et al. 2012. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci USA* **109**:2774–2783.
10 Galtier N et al. 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics* **159**:907–911.
11 Kokić T. 2010. Non-randomness of genetic mutations: some philosophical implications. *Nova Prisutnost* **8**:157–175.
12 Klamerth OL. 1972. Molecular mechanisms of spontaneous mutations. *Humangenetik* **16**:59–60.
13 Hodgkinson A, Eyre-Walker A. 2011. Variation in the mutation rate across mammalian genomes. *Nat Rev Genet* **12**:756–766.
14 Azuaje FJ, Ramirez JL, Da Silveira JF. 2007. *In silico*, biologically-inspired modelling of genomic variation generation in surface proteins of *Trypanosoma cruzi*. *Kinetoplastid Biol Dis* **6**:6.
15 Viguera E et al. 2001. Replication slippage involves DNA polymerase pausing and dissociation. *EMBO J* **20**:2587–2595.
16 Scarano E et al. 1967. The heterogeneity of thymine methyl group origin in DNA pyrimidine isostichs of developing sea urchin embryos. *Proc Natl Acad Sci USA* **57**:1394–1400.
17 Deray IJ. 2000. *Looking Down on Human Intelligence: From Psychometrics to the Human Brain*. Oxford University Press, Oxford, UK.
18 Schellenberg GE. 2002. Music lessons enhance IQ. *Nat Neurosci* **15**:511–514.
19 Wielgus AR et al. 2005. Melanin in human irides of different color and age of donors. *Pigment Cell Res* **18**:454–464.
20 Sturm RA, Frudakis TN. 2004. Eye colour: portals into pigmentation genes and ancestry. *Trends Genet* **20**:327–332.
21 Stamatoyannopoulos G. 2001. Molecular and cellular basis of hemoglobin switching. In: Steinberg MH, Forget BG, Higgs DR, Nagel RL (eds), *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management*. Cambridge University Press, Cambridge, UK, pp 131–145.
22 ENCODE Project Consortium. 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**:636–640.
23 ENCODE Explorer. 2012. A series of coordinated scientific reports published in *Nature*, September 5, 2012 online issue.
24 Kellis M et al. 2014. Defining functional DNA elements in the human genome. *Proc Natl Acad Sci USA* **111**:6131–6138.
25 Bellwood PS. 1979. *Man's Conquest of the Pacific: The Prehistory of Southeast Asia and Oceania*. Oxford University Press, Oxford, UK.
26 Pennisi E. 2012. Genomics. ENCODE project writes eulogy for junk DNA. *Science* **337**:1159–1161.
27 Schmid CW, Deininger PL. 1975. Sequence organization of the human genome. *Cell* **6**:345–358.
28 Druzhkova AS et al. 2013. Ancient DNA analysis affirms the canid from Altai as a primitive dog. *PLoS One* 8.
29 Freeman S. 2005. *Biological Science*, 2nd edition. Pearson Prentice Hall, Upper Saddle River, NJ.
30 Bird A. 2007. Perceptions of epigenetics. *Nature* **447**:396–398.
31 Rando OJ, Verstrepen KJ. 2007. Timescales of genetic and epigenetic inheritance. *Cell* **128**:655–668.
32 Guifang J et al. 2011. *N*6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol* **7**:885–887.

CHAPTER 5

Gene and genomic dynamics

Variability is the law of life, and so no two faces are the same, no two bodies are alike, and no two individuals react alike and behave alike under the same conditions, which we know as disease.

—*William Osler* [1]

SUMMARY

Since Darwin's initial theory, advances in our understanding of natural selection and Mendelian genetics have led to neo-Darwinian theory. Until the 1950s, genetic variation was mostly measured by phenotypic differences. In the late 1960s, molecular biology allowed scientists to look at protein and nucleic acid structures as another tool to study variation and evolution. In the mid-1960s and early 1970s, scientists began to question natural selection as the primary force driving evolution. Studies revealing protein variation within populations, the redundancy of the genetic code, and data showing that proteins between species accumulated mutations that did not alter protein function suggested that selection alone could not explain the maintenance of the variation within and between species. Jack King and Thomas Jukes suggested that mutation and genetic drift could explain these phenomena and postulated that although purifying selection would eliminate deleterious mutations, a variety of mutations were selectively neutral and accumulated through drift. This led to the neutral theory of evolution that suggests that some mutations are adaptive while the remaining mutations are selectively neutral, and the forces leading to their persistence are determined by drift.

Today, the study of human evolution requires both molecular and anthropological markers. Anthropological markers provide linguistic, archeological, and other types of data, while molecular markers provide nucleic acid and protein data. By examining DNA from extant populations around the world and from ancient sources found by archeologists, scientists can use contemporary technology to answer questions about our evolution. What migration patterns did our ancestors follow to lead them to a geographic region? What neutral or adaptive genetic changes led to variation between and within populations? How did the genomic rearrangements we see between different primates occur? These questions are not easy to answer, but with modern tools we can begin to understand the genetic changes that led to present-day humans.

There are several advantages to using molecular data to study evolution, including (1) large collections can be made from extant populations, (2) data collected through sequencing the human genome can be used to compare different populations and different species, (3) molecular markers can be used to show direct ancestry, (4) molecular data are quantifiable, and (5) molecular data can reveal the process of microevolution by inferring past events.

Comparisons of genomic data between and within species can provide insight into the mechanisms of genome and species evolution. Genes or DNA sequences that are conserved between species mark important or essential functions for cells, while those that change frequently may be species-specific, redundant, or unnecessary. Genome comparisons can determine the number of nucleotide differences between genomes, the number of amino acid changes, and the ability to compute the rate of change. The rate of change is a useful indicator of selection. Negative selection is indicated by slow or no change, neutral selection by the average rate of change, and positive selection by a fast rate of change (faster than neutral or background level).

The use of modern technology and computational tools in biology has led to a better understanding of human evolution. Recent anthropological and molecular genetic evidence infer that the first humans arose from admixture invarious regions of Africa and then migrated out of Africa around 200,000 years ago. Humans, however, were not the first to migrate out of Africa. They followed other species such as *Homo erectus* (2 mya),

Genomes, Evolution, and Culture: Past, Present, and Future of Humankind, First Edition.
Rene J. Herrera, Ralph Garcia-Bertrand, and Francisco M. Salzano.

© 2016 John Wiley & Sons, Ltd. Published 2016 by John Wiley & Sons, Ltd.

and *Homo heidelbergensis*(0.3 – 0.4 mya). In addition, recent genetic evidence has shown that *Homo sapiens* admixed with Neanderthals in the Middle East, Denisovans, and another unidentified group. Through genetics, linguistics, and archeology, we have been able to follow the migration patterns of humans, and today we define our species and our genetic diversity as a result of ancestry, migration, admixture, and natural selection (see Chapter 4).

Although the commonly held view is that ethnicity and background are a product of a single common ancestor, from a genetic standpoint, we are a product of many ancestors with alleles coming and going from different populations, and a mix of those alleles through recombination.

Stasis, stability through environmental change, is the predominant pattern in the history of species, but punctuated equilibrium produced through differential success has been proposed to have a significant influence on evolution. Punctuated equilibrium was proposed in the 1970s as an alternative to gradualism and because of increasing numbers of genomic examples has become an accepted explanation for some patterns of evolution.

Genome variation is a driving force behind evolution, and with technological advances in sequencing and genome analysis, scientists are beginning to find some of the major genetic alterations that led to modern-day humans. Today, admixed populations can be studied by comparing allele frequencies between populations, thereby inferring migration patterns that led to the admixture. Linkage disequilibrium (LD), which can be generated by genetic drift and selection, is used to study the joint evolution of linked genes and the forces that maintain LD or cause LD to occur. LD is also being used to study human evolution, admixture, and to map genes that are associated with quantitative characteristics. Haplotype maps (HapMaps) of rare populations and maps of common site haplotypes can be useful in reconstructing mutation, recombination, gene conversion, and evidence for natural selection.

Close examination of the genomes of various individuals has revealed increased and decreased numbers of specific nucleotide sequences referred to as copy number variations (CNVs). Using a variety of techniques, investigators have been able to associate many CNVs with a variety of phenotypes, including autism, schizophrenia, resistance to HIV, obesity, malaria morbidity, starch digestion, and steroid metabolism.

The ability to sequence individual genomes in a relatively short amount of time and the ability to look for small genomic variations have led to studies where genomes are compared between samples of healthy and diseased persons. These so-called genome-wide association studies (GWAS) involve the examination of single-nucleotide changes referred to as single-nucleotide polymorphisms (SNPs). SNP frequencies in samples with specific traits or disorders are compared with controls that do not contain the trait or disorder. These studies look for SNPs that occur more or less frequently to establish genetic correlations with the trait or disorder. The development of next-generation sequencing, SNP chips, and the increased HapMap data have made GWAS possible and especially important in looking at genetic markers associated with multifactorial traits, such as Alzheimer's, bipolar disorder, and age-related macular degeneration.

Genomic investigations of chromosomal rearrangements and selfish genetic elements (SGEs) have also provided insight into human evolution. SGEs can affect fitness, genome structure, and sex ratios, and there can be strong selection pressures to control their spread or their action. These elements can compete with nuclear and cytoplasmic components for transmission, and selection can act on the elements to increase their transmission regardless of their effect on fitness. Several investigators have argued that these elements may be an important force driving evolutionary change as evidenced by their increasing role in gene regulation, development, and the evolution of new species.

In this chapter, we will look at some of the molecular genetic forces described above and how those forces have led to our current genomic status and shaped who we are today.

Molecular evidence for punctuated equilibrium and gradualism

Darwin's initial theory of evolution by natural selection was predicated on the belief that organisms changed gradually over time. Although Darwin was unaware of the forces behind variation, he felt that change would occur steadily and gradually lead to new species. The idea of phyletic gradualism was the foundation of speciation until 1972 when Niles Eldredge and Stephen Jay Gould proposed the idea of punctuated equilibrium. Based on the fossil record and the theory of developmental homeostasis, Eldredge and Gould proposed that evolution occurred in bursts (rare and rapid events) and that gradual evolution was seldom observed in the fossil record [2]. They suggested that the fossil record revealed large-scale trends and development over long periods of time, and that most species remain in stasis between episodes of rare and rapid changes, leading to speciation. Eldredge and Gould promoted the idea of punctuated

equilibrium as a modified form of gradualism, with gradualism as the norm followed by rapid speciation events. Gould emphasized that development was a major source for natural selection and that changes could occur and accumulate over time, leading to the dramatic change in patterns sometimes seen in the fossil record [3]. This idea stood in contrast to the saltation theory proposed by Richard Goldschmidt. Goldschmidt proposed generational jumps due to extraordinary macromutation over one or two generations. These mutations would lead to the existence of "hopeful monsters" that could create new species over a few generations [4,5].

The idea of punctuated equilibrium was partially based on allopatric speciation that was promoted by Ernst Mayr [6]. Allopatric speciation is the theory that large populations remain static and that mutations, even if beneficial, rarely reach fixation due to gene flow. Small isolated populations, on the other hand, fix mutations much easier and are exposed to natural selection. If a small population became isolated from the larger group over time, the small population would evolve into a new species. Allopatric speciation partially explains the idea of stasis in large populations, although many other mechanisms have been proposed and are thought to contribute to stasis.

George Gaylord Simpson promoted the idea that paleontology could reveal large-scale trends in evolution (macroevolution) and that population genetics could reveal aspects of microevolution. While most evidence supporting punctuated evolution is attributed to the fossil record, recent studies have also contributed molecular evidence that supports punctuation.

Most examples of molecular punctuated evolution come from viruses. There are a number of examples where viral genomes have changed dramatically in a short amount of time, leading to new viral strains with the ability to infect different hosts. One example is the influenza A virus, which contains two envelope proteins known as hemagglutinin (HA) envelope protein and neuraminidase (NA) envelope protein. Both proteins are highly variable. HA is responsible for viral cell binding, and NA is required for viral replication. HA binds to specific receptors on the cell surface, and changes in HA can alter the cell type the virus binds to; for example, virus can bind to swine cells or through mutation can bind to human cells. HA rapidly evolves changes in amino acid structure to avoid acquired immune detection. This was previously thought to be due to positive selection. However, the H3N2 strain has gone through long periods of neutral sequence evolution (stasis) of HA displacing previous lineages (evidenced by the 1995–2005 stasis period). This was followed by nonsynonymous substitutions in HA fixing antigenically favorable mutations, and rapid displacement of old lineages with new dominant ones. This suggests that new strains can emerge rapidly from low-frequency preexisting strains. The data also showed that viral antigenic evolution was punctuated, and genetic changes had a large effect on antigenicity. Similar results of rapid change have been demonstrated for NA, which has been shown to exhibit punctuated evolution at the amino acid scale.

Webster et al. [7] compared the genetic changes between 56 phylogenies, among a variety of species, inferred from gene sequencing data. They found that rapid genetic change frequently correlated with speciation and that the data were most consistent with a punctuated molecular model.

Kopp and Matuszweski [8] analyzed a number of current models for evolutionary change and assessed the effectiveness of each model in a changing environment. The models took into account genetic adaptation, phenotypic plasticity, spatial adaptation, and interspecies relations. While some models supported a gradual form of evolution others indicated a punctuated equilibrium. The authors argue that the validity of both these forms of evolution makes sense for species in changing environments and that the pace of evolution may depend on the situation. During periods of dramatic climate changes, humans may have evolved more quickly. The ability of humans to learn and their genetic variability may have facilitated their adaptation and survival during rapid environmental changes. However, because some traits that have undergone selection are polygenic (involving many genes) and multifactorial (involving genes and environment), it is very probable that these traits involved gradual evolution. The findings are useful in understanding the various adaptations that humans had to make in changing environments.

One environmental change that may have had a dramatic and rapid impact in human evolution was change in diet. Changes in diet and food availability, due to environmental conditions or migration patterns, represented selective pressures that acted on both biological processes and anatomical features. Genome-wide and single-gene studies have shown that diet is an important evolutionary force that can have effects on gene regulation and expression. This type of environmental change (e.g., animal domestication, agriculture,

disappearance of food sources, and appearance of new food sources) would have led to a rapid response in the biological mechanisms underlying nutritional processes and metabolism. One example is the various mutations that occurred in the lactase gene in populations that had access to milk-producing domesticated animals [9].

Mammalian evolution occurs at different rates based on mating behavior, body size, generation turnover, fecundity, and life span. Those mammals with smaller bodies, faster generation times, high fecundity, and short life spans show faster rates of molecular evolution. Horovath et al. [10] first reported evidence for punctuated equilibrium in the human genome. A thorough evaluation of the 700 kbpericentromeric region (region containing the centromere) on the short arm of chromosome 2, formed from a fusion of two ape chromosomes, is one of the distinct cytogenetic differences between humans and their primate relatives. They found 14 ancestral loci that gave rise to duplicated DNA segments within the 700 kb region on human chromosome 2. They compared these to chimpanzee, gorilla, baboon, macaque, and orangutan. Because of the limited number of pericentromeric regions in the closely related primates, they concluded that the duplications occurred in a burst of activity during a very narrow time period between 10 and 20 mya (corresponding to the split between humans and old world monkeys) as a result of punctuated duplicative transposition. Since duplications have been associated with novel gene development (as discussed in the copy number variation section), this is an import finding and provides evidence for punctuated events at the molecular level.

Evidence for punctuated human genome evolution has been detected in the arrangement of segmental duplications (see CNVs below). Jiang et al. [11] identified 4692 ancestral duplication loci in the human genome, and ordered them into 24 distinct groups of duplication blocks. Their analysis showed that segmental duplications were often arranged around regions of transcriptional activity and primate specific genes, indicating their involvement in positive selection (due to a selective advantage). Their comparisons of segmental duplications in the human genome, with the genomes of chimpanzee and macaque supported a punctuation model of genome evolution.

A distinct feature of human evolution is hominin encephalization. The major driver of this process was increased cranial capacity that is thought to have evolved due to predation, climate, sociality, language evolution, and metabolic demands. Shultz et al. [12] reviewed the arguments for the pressures driving expansion, quantitatively evaluated the time changes in brain size, and compared these to the environmental-based hypotheses. They concluded that the use of both absolute and residual brain size estimates showed that evolution is likely to result from a mixture of both gradualism and punctuated equilibrium. Punctuated changes were observed at approximately 100 kya, 1 Mya, and 1.8 Mya, in addition to gradual intra-lineage changes in *Homo erectus* and *Homo sapiens*. In addition, punctuated changes in brain size were not shown to be temporally associated with changes in paleoclimate instability or long-term trends.

Current theories on the origin of humans point to a discrete event that occurred 100,000–200,000 years ago. Evidence for this time frame is found in (1) fossils found prior to this time period are dramatically different from those of modern humans, (2) fossils that appear to be similar to anatomically modern humans, and (3) a coalescence of mitochondrial DNA around this time. These three observations are considered evidence for punctuated morphological and molecular change. This does not rule out lengthier processes before and after the event but it does suggest that the transition from archaic to modern humans was punctuated.

The controversy over gradualism and punctuation continues today and although there is evidence for punctuated evolution, there is also evidence for gradualism. Most likely both events occurred in combination.

Next-generation sequencing

The Human Genome Project and the subsequent sequencing of thousands of human genomes was a groundbreaking event that changed the way we study human evolution. The sequences resulting from the Thousand Genomes Project and the International HapMap Project have led to our ability to study genes and gene evolution, chromosomes and chromosome rearrangements, and our ability to begin to unravel the complex genetic nature of many diseases [13–15].

The original human genome sequencing (HGS) project was proposed in 1984, and the project was begun in 1990. In 2001, the HGS project announced that they had completed 90% of the sequence, and in 2003 the project was completed. This was a huge undertaking that involved laboratories around the world working on

different regions of the genome. The completion of the project had huge implications in the study of human evolution. The human sequence revealed the complexity of the human genome, the presence of genes that humans shared with flies and worms, and allowed for the comparisons of the human genome with that of other primates [16,17].

The release of next-generation sequencing (NGS) in the late1990s was a major advance in molecular biology and a significant tool for the study of evolution. New techniques significantly lowered the cost of sequencing and have recently increased the speed of sequencing 50,000-fold over techniques developed in the 1970s. Various techniques including pyrosequencing and massively parallel sequencing are currently used, and the most popular techniques start by (1) shearing the DNA into small fragments (50–500 bp), (2) sequencing those fragments, and then (3) aligning the thousands or millions of fragments into the complete sequence using various statistical algorithms. Because the fragments are short and the alignment requires significant overlap, the ability to sequence a genome requires more than one genome copy to be sequenced. (See the following web site and the Sequencing Technology Video for a demonstration: http://technology.illumina.com/technology/next-generation-sequencing.html).

The number of copies sequenced is referred to as the coverage. A high coverage (greater than 30 copies) ensures that there are enough overlapping sequences to cover both copies of the diploid genome. High coverage sequencing is often referred to as deep sequencing that is a measure of the number of times nucleotides in the sequence are read. Low coverage (5–10 copies) leaves gaps that can be filled statistically by other sample reads but with lower reliability. Filling in gaps is done by imputation, a statistical method that uses a reference data set with known SNPs to fill in the missing data. (SNPs are single-nucleotide mutations that can be seen when comparing genomes as discussed below). This is accomplished by using algorithms that look for similar patterns among the test sequence and a completely sequenced reference, with the intention of providing SNPs and structural variants that may have been missed during sequencing [18].

NGS has been used to sequence entire genomes in order to compare individuals within and between populations, and to compare the genomes of different species [19,20]. Analysis of genomes between different species can reveal how species are related and how they evolved. For example, metagenomics (study of genetic material from environmental samples) has been used to study the coevolution of humans and their microbiome. For many years, scientists studied human microbes by growing them on artificial media. Because many of these microbes relied on specific environmental conditions to grow, most of them went undetected by culture-based methods. The development of NGS allowed scientists to uncover many of these microbes by sequencing their 16S ribosomal RNA. Prior to metagenomics, scientists did not realize the number or kinds of microbes that colonized the human body. Today, the human microbiome is thought to consist of more than 10 times the number of human cells, thought to be involved in various types of disease, the normal metabolic processes of the human gut, the protection of human epidermal layers, and the development of the immune system.

Today, NGS is beginning to reveal some of the genes that may be involved in selection and pregnancy. Selective pressures that act on women and the fetus during pregnancy include infections, oxygen deficiency, metabolic disorders, nutrient imbalances, immunological challenges, and physical fitness of the mother. Many of these pressures can threaten the life of the mother and/or the fetus. NGS is beginning to reveal some of the genes under selection in order to predict possible problems during pregnancy, and provide insights into the adaptations to unique diets and environments [21].

In addition, NGS is leading to the identification of genes for childhood diseases. Rare childhood diseases can be debilitating, have lifelong effects on the individual or be lethal. Although these diseases affect a variety of systems, they are most often associated with the central nervous system. NGS along with sequence analysis is leading to the identification of genes involved with childhood disease and a greater understanding of the mechanisms responsible for these heritable diseases. These types of studies are important because pregnancy, childhood, and parental care are all targets of selection [22].

Parental care is significant in enhancing physical fitness, social behavior, and intelligence. Parental care does not come without cost, and the more children, the higher the cost to parents, future pregnancies, and care of other siblings. Parental care can modify the genome of progeny through epigenetic mechanisms that can be transmitted to future generations. Variation in the heritability and practices of human parental care is of growing importance and is a focus of genetic and psychological research [23].

Weber-Lehmann et al. [24] used deep NGS to show nucleotide variation between identical twins. They found five nucleotide differences between monozygotic twins, indicating that the mutations between the twins had occurred after the embryo split and had been carried into the somatic tissue. One of the twins passed on the nucleotide changes to his offspring, showing that cells destined to be germ tissue had also been modified early during embryo development. This was the first example of nucleotide variation in the somatic and germ cells of identical twins, and the study points to a method (other than methylation) that can distinguish between twins.

NGS is also being used to sequence mitochondrial genomes for both forensic and evolutionary analyses. Mitochondrial DNA is much more abundant in the cell than nuclear DNA, and so it is more amenable to sequencing and alignment in samples that have been degraded. Limitations on its usefulness include its strictly maternal inheritance, lack of recombination, and heteroplasmy [25].

Studies interested only in expressed sequences can enrich the DNA sample for those sequences (exome). Exome sequencing has been used to reveal protein-coding sequences involved in disease and is especially useful for examining families that have a history of complex disease [26]. Sequencing the exome of family members may reveal variants responsible for the disease. In evolutionary studies, the variation in protein-coding regions between species may be important in finding genetic relationships and the construction of phylogenetic trees.

Recent advances in genome sequencing promise to lower the cost and speedup the process, leading to an increase in the number of human genome sequences. Some of these techniques include tunneling currents, sequence by hybridization, sequence by mass spectrometry, and nanopore sequencing [18]. One of these techniques (nanopore sequencing, also called "strand sequencing,") sequences both DNA strands in real-time using nanopore proteins attached to a semiconductor chip. An ionic current is passed through the protein pore and the DNA sequence is carried along with the current through the pore. As the DNA passes through the pore, the nucleotides alter the ion current, with each nucleotide blocking ion flow for a specific period of time. The period of current blockage is recorded as a specific nucleotide, and as the DNA strand continues to pass through the pore, the nucleotide sequence is determined [27]. Further explanation of nanopore technology can be found at the following link: https://www.nanoporetech.com.

The future combination of NGS, GWAS, and genomics promises to facilitate our understanding of the molecular mechanisms involved in selection, population divergence, and speciation [28].

Genetic variation

Over the last decade, the size and diversity of genetic data sets has increased tremendously and along with advances in technology this has allowed investigators to explore our evolutionary history and advance medical genetics. As discussed in Chapter 4, genetic variation influences the expression of traits; helps us identify populations, genetic relationships, and migration patterns; and helps predict how individuals may respond to certain medications through the study of pharmacogenetics and pharmacokinetics. Analysis of genetic variation can also help in understanding regions of the genome that have undergone selection, inferences about gene function, and the genomic patterns of variation.

Evolution occurs through the accumulation of genetic variation primarily by the mechanisms of mutation and recombination, resulting in new alleles or new allelic combinations, respectively. New variants can then be under negative selection, positive selection, balancing selection, or neutral selection. Although mutation is a driving force behind evolution and the more mutation that occurs, the faster the rate of evolution, mutation has costs. As discussed in Chapter 4, most mutations result in deleterious effects and very few are beneficial. Mutation occurs through a variety of different mechanisms, and different mutations fall into a variety of classifications. The causes of mutation include replication errors, free radicals, environmental mutagens (UV radiation, X-radiation, etc.), and transposable elements. Mechanisms reducing the free radicals from metabolism and increased integrity of DNA repair can reduce the number of mutations. When comparing nucleotide differences between humans or between humans and other species, the majority of nucleotide changes occur in the nontranscribed DNA.

Mutations can result from base substitutions, nucleotide insertions, nucleotide deletions, expanding nucleotide repeats, and transposon insertions. Base substitutions that cause no change in the amino acid of a protein are referred to as silent mutations, base substitutions that change the amino acid are missense mutations, and base substitutions that result in a stop codon are nonsense

mutations. Many base changes can occur throughout the genome without deleterious effects.

Base substitutions that cause changes in amino acids during translation can be used to study selection. Single-base substitutions leading to no change in amino acid sequence (silent) are referred to as synonymous, while those leading to amino acid changes (missense) are non-synonymous. The ratio of nonsynonymous (Ka) to synonymous (Ks) mutation can be an indicator of selection both when intraspecific or cross-species data are compared. When Ka/Ks = 1, neutral selection is in effect. When between species and within species comparisons are done, and there is a high degree of nonsynonymous mutation between species, and a low number of non-synonymous mutations within species, this indicates positive selection (Ka/Ks > 1). Purifying selection (also known as stabilizing selection) that does not alter adapted phenotypes is Ka/Ks < 1. The Ka/Ks measure can be used with single-nucleotide polymorphisms or with copy number variations [29] as discussed below.

Mutation rates can be measured directly or indirectly. Direct measurements estimate the number of mutations by measuring the number of mutations in a known number of generations, while indirect rates are determined by comparing mutations within or between species, assuming specific divergent times. Estimates of direct mutation rates in humans range from 1.1 to 3×10^{-8} per base per generation. Various studies of mutation rates have found higher levels of mutation in the germline of male mammals compared to females; however, other studies have found considerable variation in mutation rates within and between families regardless of gender.

When comparing the genomes of individuals in a population, base substitutions are detected about every 1 in 800–1000 nucleotides, with an approximate 0.1% nucleotide difference between individuals. These single-base changes are referred to as SNPs. Using modern technology, it is possible to detect millions of SNPs in a single genome. Some of these SNPs occur as haplotype blocks where very little or no recombination has taken place and the SNPs are tightly linked. SNPs can be useful for following ancestry, mapping genes, detecting selective sweeps (reduction of nucleotide variation due to strong positive selection), and in the case of large haplotype blocks are useful in GWAS [30,31].

SNPs and haplotype blocks are useful for evolutionary studies because they originated from a single ancestral haplotype. Figure 5.1 shows how an ancestral haplotype

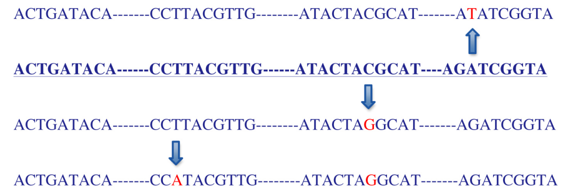


Figure 5.1 Single-nucleotide polymorphism haplotypes and their evolution. The figure represents a region of the genome covering 4000–5000 nucleotides. An ancestral haplotype (underlined and in blue font) can undergo a series of single-nucleotide changes (single-nucleotide polymorphisms—SNPs) indicated by the arrows. The SNPs occurred over evolutionary time and each new sequence on either side of the ancestral represents a new haplotype that can be traced back to the original. Note that changes can occur in different sequences and that once a sequence is changed the chance for reversion to the original sequence is infinitesimal, therefore the initial change is carried from one generation to the next until an additional change occurs. SNPs occur about once in every 1000 nucleotides in noncoding sequences and about once in every 3000 nucleotides in coding sequences. (Reproduced from Hartwell et al. 2008 with permission of McGraw Hill.)

changes over time as different SNPs occur. These different haplotypes can be traced back to their precursors and the original haplotype. This allows geneticists to follow ancestral migration patterns and genetic relatedness based on the SNP frequencies and haplotype blocks.

Copy number variations are another source of genetic variation and are widespread in the human genome. CNVs are duplicated or deleted segments of DNA and range in size from 50 to 100 bases to several megabases. CNVs can involve transcribed and/or nontranscribed regions of the genome and account for a very large segment of genome diversity. The number of base-pair differences in the genome due to CNVs, for example, is 100–1000-fold greater than SNPs. CNVs are inherited from both parents and so result in a variety of combinations within individuals. Some of these combinations can lead to disease phenotypes while others have no significant effect. CNVs have been linked to a variety of phenotypes including steroid metabolism, malaria morbidity, and starch digestion [30].

Recent studies have correlated DNA replication timing with specific mutations. In mammals, the formation of SNPs and CNV deletions have been associated with late-replicating regions, while CNV duplications have been associated with early replication. It is not clear why these differences in SNP and CNV are correlated with DNA replication timing.

Variation, population structure, and effective population size

Because genetic variation is dependent on population size and structure, it is important to define the populations used in various evolutionary studies and the measures of variation between populations. The description of populations is important for collecting and interpreting data accurately. One set of populations used frequently in evolutionary studies are those used in the HapMap Project. The HapMap Project used several populations from different cultures and ethnic groups for their study. These were the Yoruba in Ibadan, Nigeria (YRI); the Japanese in Tokyo, Japan (JPT); the Han Chinese in Beijing, China (CHB); and residents of Utah representing Northern and Western Europe (CEU).

Population geneticists use a variety of statistical tools to measure variation between populations due to genetic structure. One popular measure uses F-statistics, developed by Sewall Wright. F_{ST} (the fixation index) is the proportion of the total genetic variance contained in a subpopulation relative to the total genetic variance. F_{ST} can be measured using SNPs, microsatellites, or allele frequencies. Values of F_{ST} range from 0 to 1 with high values of F_{ST} indicating a high degree of differentiation among populations and a low value indicating that the populations are sharing genetic material through breeding. F_{ST} values for mammals generally range from 0 to 0.25, for randomly chosen pairs in a population.

Other important definitions used in evolutionary genetics are census population size and effective population size. The census population size (N) is the actual number of individuals in the population at a particular time regardless of age or gender distribution. The effective population size (N_e) is a measure that incorporates the variation in the sex ratio of breeding individuals, the number of breeding individuals in different generations, and the offspring number. Effective population size is most frequently used in population genetics because it represents a measure of variation in the population of individuals that will contribute to the next generation. When considering effective population size, it is important to remember that it also reflects the number of gametes that gave rise to the generation and the chromosomes and alleles they carried. The effective population size cannot be greater than the number of gametes that gave rise to the existing adult population, but it can be smaller depending on fluctuations in the population caused by differences in the number of surviving progeny or by changes in the number of potential parents having children. N_e can be used not only to determine variation, but also genetic drift and linkage disequilibrium, as will be discussed below [32,33].

Effective population size can be estimated using sex ratios and variance in the number of progeny produced. Sex ratios are used to measure N_e by determining the number of females (N_f) and the number of males in the population (N_m). Using the number of females and males, the effective population size can be estimated by

$$N_e = \frac{4N_f N_m}{N_f + N_m}.$$

This can be useful when considering human populations that have a lower number of females or males due to cultural reproductive practices. An example is seen in some Asian countries where females are selectively aborted, leading to a higher number of males in the population. Another important measure of effective population size in humans is the measure of variance in progeny production [34]. If the population size is going to remain constant, then each individual must contribute 2 gametes to the next generation. In this idealized population, the variance (V) in the number of gametes is 2 (i.e. $V = 2$), and the number of progeny per couple will be 2, so the effective population size is

$$N_e = \frac{4N - 2}{V + 2}.$$

If the variance is 2, then N_e will be approximately equal to N, while if $V = 0$, then N_e will be approximately $2N$. If a large proportion of the individuals in the population do not have offspring and only a small number contribute to the next generation, then the variance will be large and V will be greater than 2 and N_e is small. This measure is useful when considering human populations that restrict family size or populations that do not have access to contraception or do not use contraception for social or religious reasons.

Recombination and its effect on variation

Meiotic recombination occurs during prophase of meiosis I and results in the mixing of linked genes and their allelic combinations, a force involved in adaptive and non-adaptive evolution. Recombination is an important

feature in mammalian reproduction; in mammals, at least one recombination event per homologous pair is required for proper chromosome segregation. Recombination has a direct effect on mutations that have occurred in the genome and that may or may not be passed down. Additionally, recombination may itself be mutagenic through gene conversion (the non-reciprocal exchange of DNA sequence information). It should be noted that the process of gene conversion can occur both inside or outside of genes. Gene conversion is important for maintaining variability, thereby providing a target for natural selection [35,36].

Recombination is not uniform across the genome and can occur more or less frequently in particular regions. In humans, recombination appears to occur with high frequency in clusters of 1–2 kb regions called hotspots. Hotspots can occur in both allelic and nonallelic regions of the genome. One gene implicated in facilitating allelic hotspot recombination is *Prdm9*. This gene codes for a protein with trimethyltransferase activity and has a zinc finger domain. The protein product recognizes a 13 bp motif that is enriched at human hotspots and is responsible for up to 40% of hotspots in the human genome. Mutations in the *Pdrm9* allele or the 13 bp binding motif reduce recombination activity [37].

In humans, hotspots also correlate with GC content and a 7 bp motif CCTCCCT. This motif is associated with 11% of the hotspots, while another GC-rich motif, CCCCACCCC, is associated with 3% of the hotspots. Hotspots in humans tend to be outside of genes. No hotspots have been found in exons and only a few have been described in introns. This may be due to selection eliminating progeny with hotspots inside of genes. Hotspots also occur in imprinted regions of the human genome that are highly methylated [38,39].

Recombination frequencies in humans also vary by gender, with females having 1.7 times as many recombination events as males. Differences are also seen in the chromosomal regions undergoing recombination when comparing genders. Male recombination tends to occur more frequently near the telomeres, whereas female recombination occurs more frequently closer to the centromere. These differences may be caused by sex-based variation in the double-strand break mechanisms or by sex-specific mechanisms of imprinting that are recognized on a genome-wide scale [40].

Biased recombination events and meiotic drive (selective transmission of certain chromosomes or alleles to the next generation) can affect mutations by passing them on selectively. Recombination hotspots can also play a role in the selective transmission of alleles. In humans heterozygous for a recombinogenic allele and a nonrecombinogenic allele, Jeffreys and Neumann [41,42] showed that meiotic drive favors the transmission of the nonrecombinogenic allele. Myers et al. [40] later showed that meiotic drive against hotspot motifs in primates involved the *Prdm9* allele.

Studies by Winkler et al. [43] found that recombination hotspots in humans are not found at the same positions in the chimpanzee genome even though similar DNA sequences exist in both genomes. Through fine-scale recombinant mapping studies, they suggest that human recombination evolved at a faster rate than in chimpanzees.

Nonallelic homologous recombination (NAHR) is another source of genetic diversity and can result in genomic duplications, deletions, translocations, and inversions (Figure 5.2). NAHR hotspots are different than allelic hotspots and studies have shown that these hotspots can be shared across species. Segmental duplications, regions of similar sequence, and transposable elements serve as regions for NAHR activity. NAHR can often lead to highly deleterious combinations and lethality, and NAHR activity resulting in gene conversion has been implicated in a number of genetic diseases (e.g., neurofibromatosis, Prader–Willi syndrome, and other microdeletion disorders). NAHR is also a source for CNVs [44,45].

Linkage equilibrium and disequilibrium

Linkage equilibrium (LE) occurs when linked alleles associate randomly with each other in a population. LE is similar to the conditions under Hardy–Weinberg equilibrium (HWE) in that it assumes random association of alleles over time. LE differs from HWE in that random association does not occur in a single generation but is dependent on the rate of recombination and the number of generations. LE increases over time and with an increasing number of generations.

Linkage disequilibrium (LD) occurs when there is a nonrandom distribution of linked alleles in a population. For example, assume loci $\boldsymbol{A}$ and $\boldsymbol{B}$ are linked. Taking into account only two alleles at each locus, alleles $\boldsymbol{A}$ and $\boldsymbol{a}$ are linked to alleles $\boldsymbol{B}$ and $\boldsymbol{b}$. We can set the frequency of $\boldsymbol{A}$

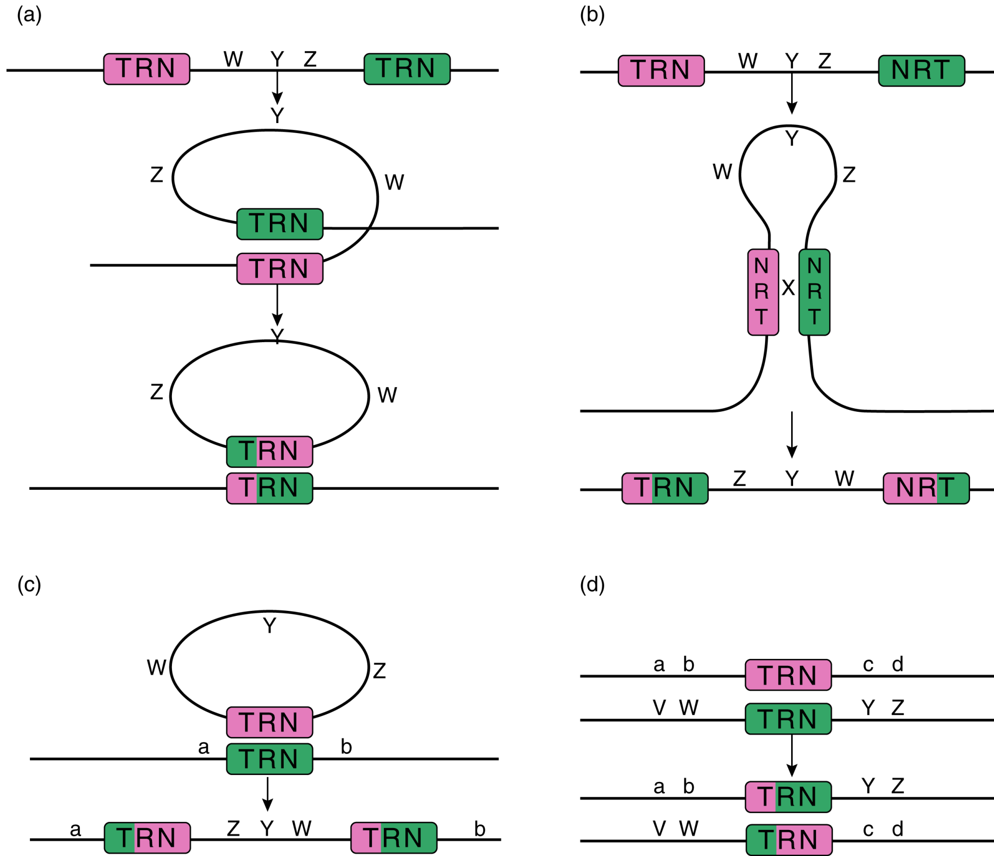


Figure 5.2 Nonallelic homologous recombination and its function in chromosome rearrangements. Figure represents the various types of recombination that can occur in a nonallelic homologous recombination (NAHR) resulting in a deletion, inversion, insertion, and translocation. The figure shows a transposable element (TRN) that can act as a sight for NAHR. (a) Shows a chromosome loop forming as a result of two homologous TRN sequences existing as direct repeats on the chromosome. Recombination between the elements results in a deletion (ring chromosome) of the sequence between the two elements and a hybrid (red–green) transposon. (b) Shows the recombination between two inverted elements along the same chromosome resulting in an inversion of the intervening sequences. (c) Shows a ring chromosome with a TRN element and a nonhomologous chromosome with a TRN element. Recombination between the two TRN elements results in an insertion of segment *ZYW* into the linear chromosome. (d) Shows two nonhomologous chromosomes (*abcd* and *vwyz*) lining up as a result of two TRN sequences on the different chromosomes. Recombination between the TRN sequences results in a chromosome translocation, where *a* and *b* are now linked to *y* and *z*, and *v* and *w* are now linked to *c* and *d*. (See the Color Plates section.)

and ***a*** to *p* and *q*, respectively, and the frequency of ***B*** and ***b*** to *r* and *s*, respectively. If the alleles are in linkage equilibrium, then the arrangement of alleles in coupling [***AB***; ***ab***] and in repulsion [***Ab; aB***] will be random, and the product of the two gametes in coupling ($pr \times qs = pqrs$) is equal to the product of gametes in repulsion ($ps \times qr = pqrs$). If the alleles are in disequilibrium, then the alleles do not associate randomly and the combination of allele frequencies in repulsion does not equal the combination of allele frequencies in coupling. The extent of disequilibrium can be measured by the difference between the two products and is represented by *D*:

$$D = (\text{freq. of } \boldsymbol{AB})(\text{freq. of } \boldsymbol{ab}) - (\text{freq. of } \boldsymbol{Ab})(\text{freq. of } \boldsymbol{aB}).$$

Linkage equilibrium is at maximum when $D = 0$. Based on a frequency of 0.5 for each allele at the different loci, the maximum value of *D* when disequilibrium occurs is 0.25. If allele frequencies at the loci are different, then disequilibrium is not complete. For example, if the frequency of ***A*** is 0.6 and the frequency of ***B*** is 0.5, then not

all of the ***A*** alleles would be linked to ***B***. LD typically involves kilobase regions of linked alleles or SNPs, but because it is a measure of association, LD can also involve unlinked genes that occur in a population [46,47].

The classic measure of linkage disequilibrium (D) is the observed allele frequency minus the expected frequency (if the alleles were randomly segregating). For two loci, each with two alleles (***A*** *and* ***a***) and (***B*** and ***b***), the observed frequency (P) is the number of individuals in the population that have genotype $\boldsymbol{AB} = P_{AB}$. The expected frequency (P') is the product of the two allele frequencies in the population: $P_A \times P_B = (P_A P_B)$. Therefore, the measure of the magnitude of LD is: $D = P_{AB} - P_A P_B$.

Example 5.1

Observed genotypic frequency : $\boldsymbol{AB} = 0.6; \boldsymbol{Ab} = 0.1;$
$\boldsymbol{aB} = 0.2; \boldsymbol{ab} = 0.1.$

Given these genotypic frequencies in the population, you can determine the allele frequency (see Chapter 3).

Allele frequency : $\boldsymbol{A} = 0.7; \boldsymbol{B} = 0.8; \boldsymbol{a} = 0.3; \boldsymbol{b} = 0.2.$

The LD can be computed by taking the observed frequency minus the expected frequency:

$$D = P_{AB} - P_A P_B = 0.6 - (0.7 \times 0.8) = 0.04.$$

D can also be calculated from the observed frequencies:

$$D = (P_{AB})(P_{ab}) - (P_{Ab})(P_{aB}) = (0.6)(0.1) - (0.1)(0.2) = 0.04.$$

Using the classic measure of LD, we can also measure LD as the square of the correlation coefficient between the A and B loci:

$$r^2 = D^2/(P_A P_a P_B P_b) = (0.04)^2/[(0.7)(0.3)(0.8)(0.2)] = 0.048.$$

When $r^2 = 1$, there is complete disequilibrium. A good guide to the usefulness of LD is to use r^2 as a measure, because the sample size required to detect statistically significant LD is inversely proportional to r^2. Additionally, r^2 is dependent on allele frequency, making it useful in association studies.

Under selectively neutral evolution and equilibrium, the value of r^2 is $1/[4N_e\,c + a]$ where N_e is the effective population size, c is the recombination rate, and a is a constant. In the absence of mutation, $a = 1$, while $a = 2$ if mutation is taken into account. Thus, effective population size can be estimated from LD. r^2 is also related to the population recombination parameter ρ (rho). Rho measures the effective population size and recombination rate using the following equation:

$$\rho = 4N_e\,c$$

where N_e is the effective population size and c is the recombination rate. When ρ is large, LD $= r^2$, which is approximately $1/[4N_e c + 2]$.

According to Hardy–Weinberg equilibrium, alleles at the same locus should reach equilibrium within one or two generations of random mating, depending on the distribution of alleles among the sexes. When two loci are linked, the distance between loci is proportional to the amount of recombination (considering no crossover suppression due to chromosome rearrangements, no mutations in the genes controlling recombination machinery, and no mutation at DNA recombination sites). Linkage disequilibrium should decrease with each generation according to the distance between loci. The new value of D after one generation of random mating is

$$D_1 = (1 - r^0)D_0,$$

where D_1 is the new value of disequilibrium, r^0 is the frequency of recombination (not to be confused with r the correlation coefficient), and D_0 is the initial value of D in one generation. Over time, disequilibrium will diminish with each generation:

$$D_t = (1 - r^0)^t D_0,$$

where D_t is the new value of disequilibrium, r^0 is the frequency of recombination, D_o is the initial value of D, and t is the time.

When two genes are unlinked, the value of $r^0 = 0.5$, and when there is complete disequilibrium, $r^0 = 0$. The level of disequilibrium is halved with each generation of random mating. In large populations, LD will decrease over time as recombination occurs. Because recombination breaks down LD over time, large regions of LD indicate recent events, allowing LD to be used to indicate the time that a mutant allele occurred [48].

Forces leading to linkage disequilibrium

The forces leading to LD are selection, genetic drift, and admixture. Selection can favor certain allelic combinations that impart a higher level of fitness. Selection may act on a favorable gene that carries a whole series of linked alleles (that may or may not be favorable), maintaining disequilibrium. A favorable allele linked to neutral or unfavorable alleles is the classic example of genetic hitchhiking where the neutral or unfavorable allele increases in frequency due to the selection of the favorable allele and LD.

The nonrandom association of alleles at different loci, due to genetic coadaptation, may exist with some alleles within a population. This may be due to specific allelic combinations imparting an adaptive advantage to the organism. Disequilibrium will only occur if natural selection favors a particular combination of alleles. The further the linked alleles are from each other, the stronger the forces of natural selection need to be in order to maintain disequilibrium. Natural selection of genes in disequilibrium can also occur among nonlinked loci (on different chromosomes), but this is rare [49,50].

Drift can constantly remove haplotypes from large populations, and subsequent recombination will decrease LD. In small populations, drift can lead to LD by parents passing on blocks of linked alleles that by chance leave more offspring. Additionally, LD can reach an equilibrium value in these small populations. Through drift, this block may increase in frequency in the population [51].

During the first few generations, admixture can especially lead to LD. When two diverse genotypes come together, each chromosome will vary by allele composition and genes along those chromosomes will be linked by ancestry. Because each parent contributes a different set of chromosomes with different allele combinations, admixture leads to LD over very large chromosomal segments. For example, if a population with homozygous genotype ABDE is admixed with a population whose homozygous genotype is abde, the offspring will all show complete disequilibrium in the first generation ABDE/abde. The greater the number and size of LD regions on the chromosome, the earlier the migration event occurred. When recombination occurs in future generations, LD declines and ancestral alleles become fragmented with only very tightly linked alleles remaining in LD [52]. Vernot and Akey [53] used whole genome sequencing to examine the admixture between extant European and Asian populations and Neanderthals. They recovered a substantial amount of Neanderthal sequence, that revealed positive and purifying selection occurred, and found that admixture occurred both before and after the divergence of non-African modern humans.

Because LD decreases over time, admixed populations can be used to determine the time of the migration event by looking at the decay of LD. The use of modern data from admixed populations along with appropriate statistical methods can help to determine the level of admixture remaining and the number of generations since the initial event (e.g., migration). Software programs like GLOBETROTTER, HAPMIX, ROLLOFF, ALDER, and modified versions of principle component analysis examine SNP haplotype blocks and determine the decrease in SNP linkage when compared to specified parental groups [54]. Having more than two groups mixing, gradual mixing rather than a single event, or having similar groups admixing multiple times can complicate admixture-dating methods.

In the case of ALDER software, only one ancestral population is needed to determine admixture history. Loh et al. [55] used data from the Human Genome Diversity Project to validate the use of ALDER, revealing the admixture history of Central African Pigmies, Sardinians, and Japanese.

Hellenthal et al. [56] recently developed a genetic atlas of human admixture using the programs fineSTRUCTURE, CHROMOPAINTER, and GLOBETROTTER. The fineSTRUCTURE program groups populations on the basis of genetic similarities. The CHROMOPAINTER algorithm then colors the chromosomes from the populations according to worldwide group patterns based on shared genetic ancestry. Using these methods, Hellenthal et al. were able to date admixture events corresponding to 4000 years ago (assuming a birth date of 1950 and a generation time of 28 years).

Elhaik et al. [57] used admixture to develop the geographical population structure (GPS) algorithm and demonstrated an accuracy of 83% when placing individuals in their country of origin. The GPS method appeared to outperform the SPA (spatial ancestry analysis) approach that models allele frequencies [58].

Linkage disequilibrium and SNP haplotypes

SNPs are widely used in LD studies because reverse mutation (reversion) in these markers is rare. Less focus has been placed on LD regions containing microsatellites, insertions, deletions, and inversions. The statistical association of SNP variability along the chromosome can be used to estimate the presence and size of LD regions. The statistical significance of LD, which depends on sample size, is different from the magnitude of LD (the size of the LD region). Using an example where there are two SNPs (X, Y), one measure of magnitude can be defined as

$$D = P_{XY} - P_X P_Y$$

where P_{XY} is the frequency of XY being carried by gametes, P_X is the frequency of X, and P_Y is the frequency of Y. Note that this is the same as the measure of D with alleles A and B as shown above. The maximum value of D depends on the SNP frequency. Whether D is negative or positive is arbitrary and depends on the labeling of the alleles.

When considering an order where there are three different SNPs (X, Y, Z), the estimate of LD is

$$D_{XYZ} = P_{XYZ} - P_X D_{YZ} - P_Y D_{XZ} - P_Z D_{XY} - P_X P_Y P_Z.$$

In this case, the D_{XYZ} value is a measure of the three-way interaction of the SNPs and the D values in the equation indicate the pairwise disequilibrium. This can be extrapolated to include more than three linked SNPs.

LD can also be established using SNP tags. SNP tags are unique SNPs that flank a region of high LD (SNP blocks that form haplotypes). These tags flank the LD region and can be used in association studies to help map genes associated with complex disease. The size of the LD regions flanked by SNP tags varies among populations and can be estimated by the statistical association of variation along the chromosome. In some populations, SNP blocks can become fixed. SNP blocks that have a very high frequency in a population and become fixed are referred to as haplogroups.

SNP haplotypes can be determined by pedigree analysis in LD regions. Haplotypes under high LD (assuming no recombination) can be followed in families by looking at the genotypes of the parents and progeny.

Example 5.2

Chromosome Haplotypes

Parent 1		**Parent 2**
CT or CT		CC or GA

Possible Progeny
(CT, GA) or (CT, CC)

Progeny haplotypes or parental haplotypes can be resolved by genotypes. In the example above, the Parent 2 haplotypes can be resolved by looking at the progeny. LD can also be used to trace ancestry when a new mutation arises within a block and is passed on to subsequent generations.

The transmission/disequilibrium test (TDT) described by Ewens and Spielman [59] is a good example of how linkage association can be determined even in the presence of population subdivision and admixture. This test utilizes data from markers transmitted from heterozygous parents to their offspring in an attempt to determine association with disease susceptibility. Assuming two alleles (A and a), and that the heterozygous parents transmit one of the alleles (A) x times and transmit the other allele (a) y times, then $\mathrm{TDT} = (x-y)^2/(x+y)$, which is the approximate χ^2 distribution with one degree of freedom. The test is particularly useful because it separates linked genes from genes associated with a particular phenotype [60].

The development of next-generation sequencing, microarrays, and SNP chips has greatly reduced the time needed to determine SNP haplotypes. Using SNP chips (based on a SNP tag design) that cover 500,000 to 5 million SNPs, scientists have begun to establish major and minor SNP associations with diseases such as Alzheimer's, bipolar disorder, and age-related macular degeneration. The completion of the International HapMap Project and the 1000 Genomes Project (with the CEU, CHB, JPT, and YRI populations) has provided a large amount of global genome information. This has led to the development of software capable of handling large data sets. Software like Haploview [61] can be used to reveal small, tightly linked SNP blocks, and to locate SNP tags. ArchiLD software developed by Melchiotti et al. [62] examines SNP blocks in LD (as defined by r^2) and orders the LD regions into clusters based on their level of LD. The ArchiLD software allows visualization

across highly polymorphic regions containing high-density SNPs and fragmented LD blocks.

Linkage disequilibrium in humans

The pattern of linkage disequilibrium is highly variable in humans on both regional and population levels. LD creates a series of haplotype blocks in the human genome, separated by regions of recombination hotspots. The human genome shows regions of block-like structure with high-LD regions up to 40 kb, separated by short regions of less than 5 kb. LD can be used in association studies to (1) map genes involved in disease and other complex traits, (2) help understand the evolution of the genome, and (3) identify recent positive selection (Figures 5.3 and 5.4).

Sabeti et al. [65,66] assessed the age of core haplotypes (haplotypes at a locus of interest) by the decay of their association with linked alleles. The measure of maintenance or decay is referred to as extended haplotype homozygosity (EHH). Core haplotypes that have a high EHH (retain linked markers) and a high population frequency are assumed to have accumulated at the locus of interest faster than expected under neutral evolution. The investigators used this technique to look at two loci, G6PD and CD40, proposed to carry malaria resistance. In both cases, the loci showed evidence of strong positive selection.

Voight et al. [67] examined SNPs from the HapMap Project and developed a map of recent positive selection by using SNP tags and the Haplotter web tool. They identified approximately 250 signals of recent selection in each of the populations studied. Although they were unable to identify specific phenotypes associated with genes under selection and selective sweeps in progress, the data and approach are useful for studying adaptation and selection and ultimately the phenotypic variation due to these genetic variants.

A classic example of linkage disequilibrium in humans is found in the region of the major histocompatibility complex (MHC), also referred to as the human leucocyte antigen (HLA). First recognized by Little [68] in mice undergoing transplantation experiments, the MHC region in humans is composed of approximately 140 tightly linked genes covering a 3.6 megabase region of chromosome 6. MHC genes are involved in the immune response and are divided into three subgroups: classes I, II, and III. In general, class I genes mediate destruction of host cells displaying an antigen, class II imparts specific immunity to an antigen, and class III includes several secreted proteins involved in the complement system. In humans, the focus is on classes I and II because of their association with autoimmunity and the apparent linkage of certain haplotypes to disease susceptibility. Class I and II genes are highly polymorphic and are expressed in a codominant fashion.

The first description of LD in MHC class II by Daly et al. [69] revealed a 216 kb segment that made up a haplotype block, flanked by recombination hotspots. The high levels of genetic diversity and LD in the MHC have interested evolutionary biologists. Many biologists

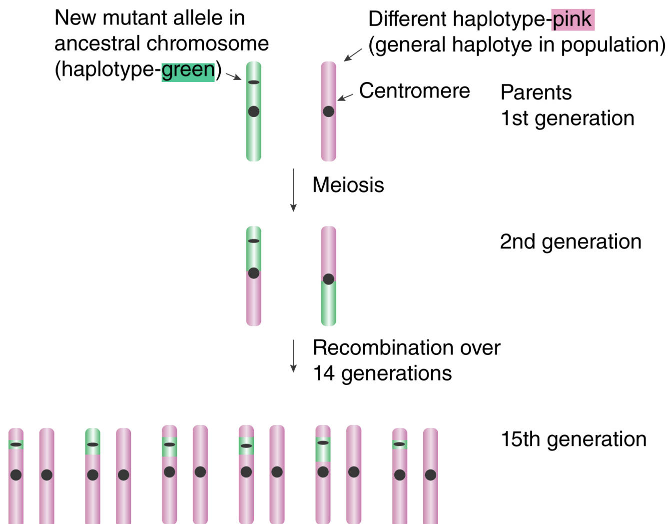


Figure 5.3 Mapping genes using linkage disequilibrium. The figure shows two homologous chromosomes from the original parents (representing two different haplotypes pink and green). A mutation (black line) takes place in one of the original chromosomes (green ancestral haplotype). The pink chromosome represents the haplotype in the general population and the green region the original haplotype where the mutant allele first occurred. After several generations and multiple meiotic recombinations, the mutant allele still carries segments of the ancestral chromosome (green haplotype). The mutation can be mapped by scanning the chromosomes for the ancestral (green) haplotype, which still surrounds the mutation. (Adapted from Lodish et al. 2008 [64] (See the Color Plates section.)

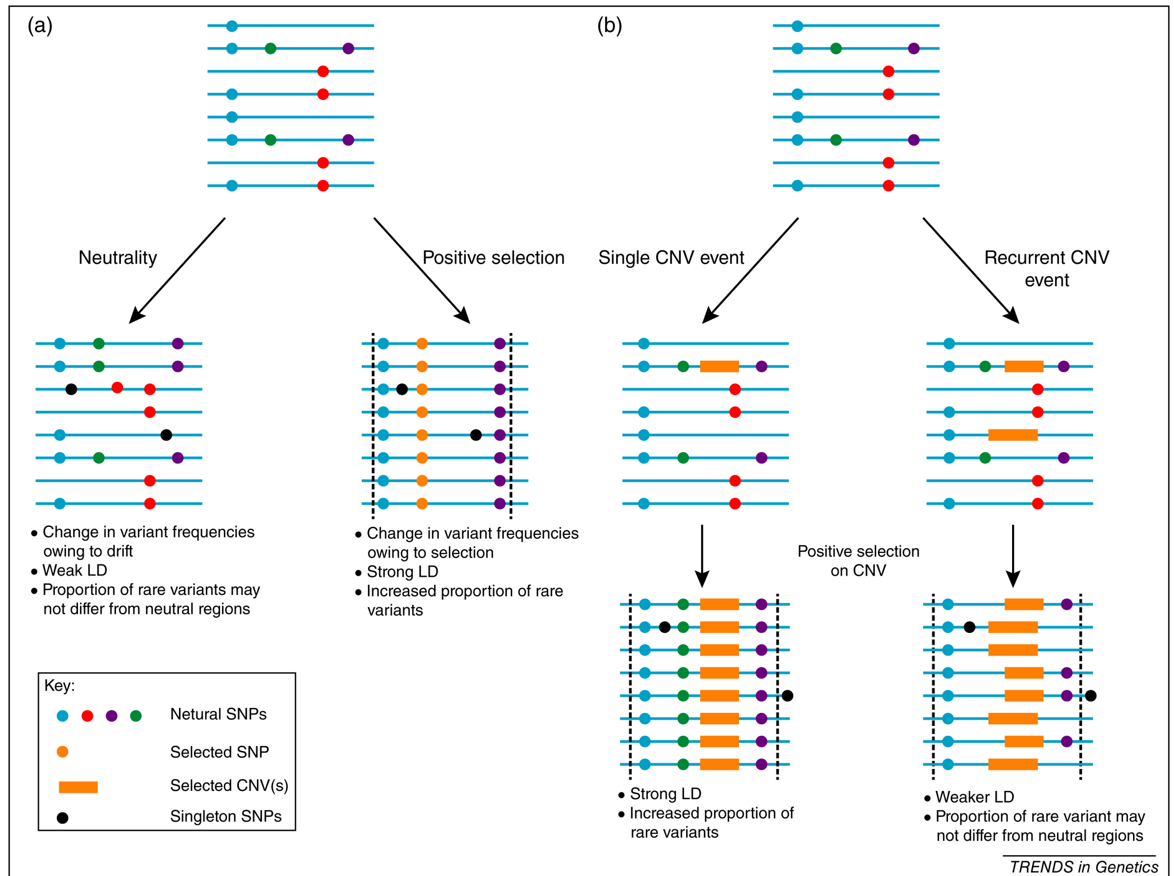


Figure 5.4 Positive selection of SNPs and CNVs. The figure demonstrates how an SNP and a CNV (described later) can undergo positive selection or neutral selection. (a) The horizontal lines represent haplotypes in a population with the circles representing SNPs. In the case of neutral selection, changes in frequency of the haplotypes occur as a result of drift and/or weak linkage disequilibrium. Other SNPs (mutations) may also occur (black circles). In the case of positive selection, an SNP variant (orange circle) increases in frequency and reaches fixation, known as a selective sweep. In this example, fixation occurred rapidly and the population exhibits linkage disequilibrium (LD) with respect to the other SNPs (blue and purple circles). Strong LD is indicated by the dashed black vertical lines. (b) CNVs can occur as a single event (left side) or as two independent (recurrent) events (right side). The left-hand side of Figure 5.4b shows a selective sweep acting on the CNV in strong LD with surrounding SNPs (blue, green, and purple circles). After the selective sweep, neutral mutations can be introduced (black circles). The right-hand side of Figure 5.4b shows a selective sweep between recurrent CNVs and because of independent events between the recurrent CNVs there is weaker LD and more haplotypes, making positive selection more difficult to determine. (Reproduced from Iskow 2012 [63] with permission of Elsevier.) (See the Color Plates section.)

attribute the diversity to balancing selection, where no single allele is most fit but the heterozygote is, thereby leading to allelic variation at the population level. Pathogen coevolution has also been suggested as a mechanism for diversity, where common alleles are under the greatest pressure, thereby driving positive selection of uncommon alleles. If a given pathogen is harmful in the presence of common alleles, new alleles derived by mutation may develop, leading to an enhanced immune response [70]. This situation may be changed by mutations in the pathogen's genome and new changes in the host, in what can be characterized as an arm's race, where stabilization probably never occurs.

Linkage disequilibrium in the MHC varies in a haplotype manner where haplotypes are either ancestral or recombinants of ancestral haplotypes, and are arranged in

conserved blocks of variable size. Variation in the size and boundaries of LD exist in different MHC backgrounds, and the different haplotype backgrounds are important to consider when trying to map disease genes associated with HLA. At least one 60-kb LD SNP region in the MHC II region exists in all populations studied (including Africans), and is a sign of strong positive selection [71].

Understanding disequilibrium variation between populations is important when trying to map disease alleles using GWAS. Li et al. [72] looked at LD variation in three populations from the International HapMap Project and three Asian groups from the Singapore Genome Variation Project. This study focused on LD of SNPs around the LRRK2 locus, previously identified as being associated with Parkinson's disease. These investigators found significant differences in LD SNP patterns between the Caucasian and Asian groups (with the exception of Indians in the Asian group), suggesting that GWAS studies cannot be generalized on a global scale.

Fluctuations in population size can also affect LD and the ability to use LD to map complex traits. Laan and Paabo [73] examined microsatellites on the X chromosome in two populations, the Saami (population from northern Fenno-Scandinavia), which had been in constant population size, and the Finns, whose population had expanded recently. They found higher levels of LD among the Saami, suggesting that populations of constant size may be more suited to mapping complex traits caused by older mutations.

LD mapping has also been used to locate genes involved in behavioral disorders. Using the transmission/disequilibrium test and high-density LD mapping, Hawi et al. [74] were able to show association between two genes (*SLC6A2* and *ADRA1B*) and attention deficit hyperactivity disorder (ADHD). Three SNPs linked to *SLC6A2* were identified and showed nominal association with ADHD. The *SLC6A2* gene encodes a norepinephrine transporter that controls the level of norepinephrine in the synapse. Nominal association with six SNPs was observed with the ADRA1B gene that encodes an adrenaline receptor.

Genome structural variations

Changes in the number of base pairs in the genome are referred to as structural variations (SVs). A significant proportion of the human genome is affected by SVs. These can include large chromosome rearrangements (inversions, translocations, deletions, duplications, aneuploidy, etc.) as well as smaller copy number variations (e.g., segmental duplications, transposable elements, deletions, and inversions). Recently, these structural variations have been recognized as a significant force behind genome evolution.

Chromosome rearrangements typically involve large segments of chromosomes that have inverted, duplicated, deleted, or translocated. These large rearrangements can involve hundreds to thousands of genes and often lead to deleterious outcomes unless they are somehow balanced in the genome.

CNVs are relatively smaller segments of DNA that increase or decrease the number of nucleotides. CNVs can vary in length from a few bases to megabases and result in deletions, duplications, and inversions. CNVs can change the dosage of genes through duplication or deletion, or can affect nontranscribed regions of the genome. Gene and/or chromosome dosage have been recognized as important due to the variety of diseases and developmental abnormalities associated with their duplication or deletion. Duplications and deletions of genes result in various phenotypic effects. The effects of CNVs in nontranscribed regions are less clear, but of increasing interest.

Over the last decade, CNVs have been discovered to be a source of diversity that exists in all mammals. Although the emphasis of CNV investigations has been on screening for disorders, all individuals in the general population have CNVs. In general, the number of CNVs between any two individuals varies by about 0.75%, and studies by Itsara et al. [75] have found between three and seven variants per person (an average of 540 kb pairs per individual). The number of CNVs varies both within and between populations; however, studies have shown that segmental duplication architecture is stratified in different ethnic groups, making it important to study groups of similar ethnicity when looking for CNVs involved with disease. This segmental duplication architecture can result in predisposition or protection from CNV disease in different ethnic groups. CNVs have been associated with changes in gene expression, disease, disease resistance, and sources of new gene expression.

CNV classifications and formation mechanisms

Most CNVs are formed during meiosis; however, studies have shown that monozygotic twins have variable CNVs,

suggesting somatic mechanisms can also lead to CNV formation. There are several mechanisms thought to give rise to CNVs, including NAHR (recombination between segmental duplications), microhomology-mediated break-induced replication, nonhomologous end joining [76,77], and alternative splicing [78].

CNVs are classified as recurrent or nonrecurrent depending on the molecular mechanism that created them. Recurrent CNVs have identical break points and are the result of NAHR. NAHR can occur between small segmental duplications (unique low copy number sequences, 10–100 kb), transposons (unique high copy number sequences, 300–500 bp), and hotspot-mediated sites [79]. Common transposons involved in CNV formation are L1 retrotransposons, *Alu*, and SVA elements (see "Transposable elements" section for definitions). A 14 bp motif highly conserved in *Alu* elements and associated with most SVA elements has been shown to be a common breakpoint for CNV formation, especially in *Alu*–*Alu* recombination [80].

Nonrecurrent CNVs are detected in regions of the genome that lack extensive homology and can result from replication errors and nonhomologous end joining.

CNVs resulting from replication errors occur as a result of breakage induced at replication forks and broken ends that anneal with nearby single-stranded DNA (fork stalling or template switching). The breakpoints at replication forks are sometimes associated with 2–5 bp regions and because of these small homologous regions, the process is known as microhomology-mediated break-induced repair [81]. Nonhomologous end joining also results from breakage and reannealing with a nearby strand, but this can occur outside of replication. Variable number of tandem repeats (VNTRs) have also been implicated in the formation of nonrecurrent CNVs.

When comparing NAHR and VNTR mechanisms of CNV formation, NAHR is seven times more likely to lead to large CNVs than VNTRs, even though both are equally likely to contribute to CNV formation. As expected, large CNVs (>500 kb) are rare and can be deleterious.

Recent studies have correlated DNA replication timing with CNV deletions and duplications. Deletions have been associated with regions undergoing late replication and duplications with early replication. Several mechanisms for these observations have been proposed. Replication involves domains of 400–800 kb where several replicons are activated simultaneously. The distribution of replication origins differs in late versus early replication. The high density of early replication forks may allow for more potential recombination sites within the domains. For example, Lu et al. [82] looked at CNV formation during genome reorganization. They used induced pluripotent cells and compared them to their parental fibroblasts to study how reprogramming impacted CNV formation. They found that a significant number of genomic regions changed replication timing as a result of induced pluripotent reprogramming. Copy number gains accumulated in regions that changed to earlier replication during reprogramming.

A subclass of recurrent CNVs are reciprocal CNVs, which result from a deletion in one segment of a chromosome and a duplication in that same region on the homologous chromosome. Reciprocal CNVs can cause mirrored phenotypes (where opposite clinical features result from deletions or duplications), identical phenotypes, overlapping phenotypes, or unique phenotypes [83].

When CNVs rise to a high frequency in a population, they are referred to as copy number polymorphisms (CNPs). CNPs can exist as biallelic or as multicopy and because most CNPs seem to arise from NAHR, most multicopy CNPs exist in regions with high segmental duplications. CNV sequences that increase or decrease the number of nucleotides between different species are referred to as copy number differences (CNDs) [84].

Methods used to detect CNVs

CNVs can be difficult to detect accurately, but several approaches have been successfully used to detect CNVs associated with disease and evolutionary mechanisms. One of these approaches, array comparative genomic hybridization (aCGH), uses a microarray system to detect both increases or decreases in copy number by measuring the level of fluorescence in a test sample compared to a standard (usually with two copies of the sequence). Cloned sequences are developed as a standard and then labeled with a fluorescent dye, and genomic DNA from the test sample is sheared and labeled with a different fluorescent dye. The samples are then cohybridized to an array spotted with thousands of oligonucleotide probes. Increased copies in the test genome will result in a fluorescence ratio equal to or greater than 3:2, whereas a decrease in copy number will result in a fluorescence ratio equal to or less than 1:2. Once CNVs are revealed, they can

be identified by aCGH array platforms or by next-generation sequencing read depth (Figure 5.5).

One problem associated with the detection of CNVs using microarrays is the use of the human genome reference sample that includes several deletions and duplications, which makes comparisons difficult. For example, if the number of CNVs in the standard is 13 and the test sample has 14, the algorithm used to determine fluorescent level may not be sensitive enough to distinguish the CNV variation from background. Massive parallel sequencing can overcome this problem once it becomes cost effective to analyze large numbers of individuals.

Another method to detect CNVs is the paralogue ratio test (PRT). The PRT uses pairs of PCR primers that simultaneously coamplify the target CNV, as well as a single copy paralogue from a non-CNV region. The amplified sequences vary in length and the number of copies in the CNV region can be detected by fluorescence similar to that described by aCGH. Primers are designed online (see prtprimer.org).

NGS can be used to detect CNVs by sequencing the test sample and then aligning sequences to a reference genome [85,86]. Gaps in alignment reveal deletions, and overlaps in alignment signal duplications. Problems

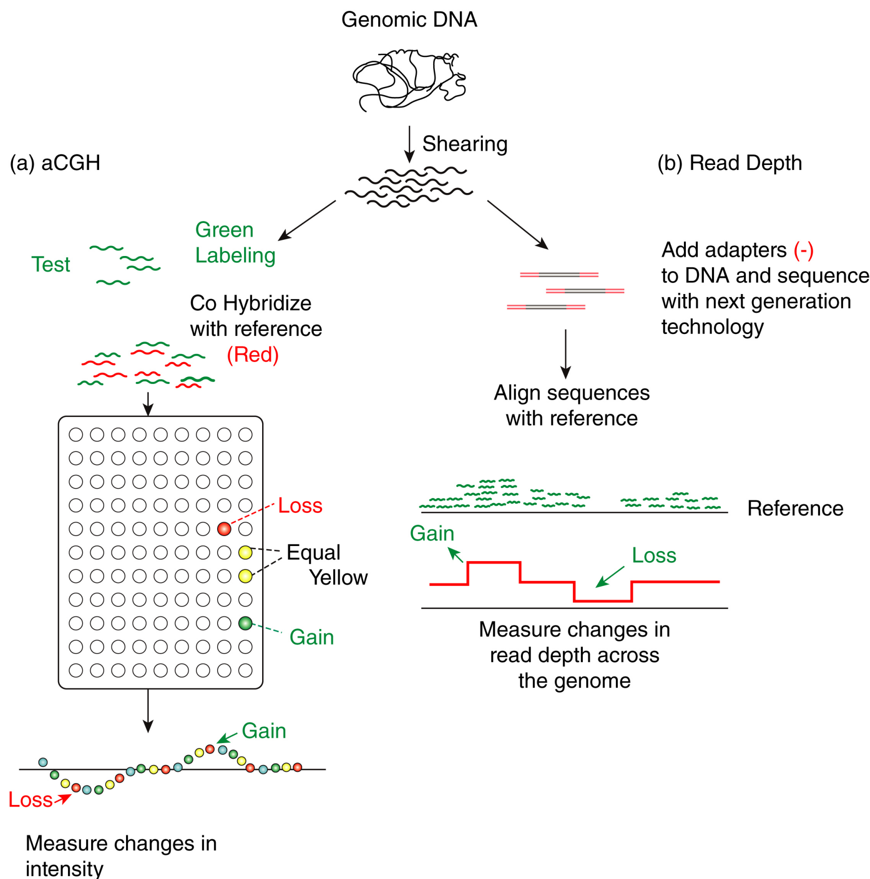


Figure 5.5 Two methods of detecting copy number variations. (a) Array comparative genomic hybridization (aCGH) is performed by shearing the genomic DNA and labeling it with a green fluorescent label. The reference DNA is then labeled with a red fluorescent label and the two DNA samples are cohybridized to an array spotted with thousands of oligonucleotide probes. The signal intensities are detected and used to make CNV calls. Red hybridized regions indicate a loss of copy number, green regions indicate a gain in copy number, and yellow (red–green combinations) regions indicate equal numbers of the sequence. (b) After shearing the DNA, adapters are added in order to use next-generation sequencing. The resulting sequences are aligned to a reference genome to detect increases or decreases in CNVs. (Reproduced from Iskow 2012 [63] with permission of Elsevier.) (See the Color Plates section.)

associated with this technique include (1) highly repetitive tandem sequences and determining the exact number of repeats in the test sample sequence and (2) the detection of break points in duplications. Mills et al. [87] used sequencing data from the 1000 Genomes Project to look for CNVs in 179 individuals. By sequence comparisons, they were able to identify a number of previously unidentified duplications and deletions in various regions of the genome. Advances leading to lower costs for whole genome sequencing are likely to reveal even more CNVs and set the stage for genome-wide association testing.

CNVs associated with human phenotypes

Using a variety of techniques, investigators have been able to associate CNVs with a variety of phenotypes, including autism [88], schizophrenia [89], resistance to HIV [90], obesity [91], malaria morbidity, starch digestion, and steroid metabolism [79].

The detection of disease-associated CNVs is complicated by several processes associated with CNV variants. These include the ability to detect small and rare variations between test samples and controls, the epistatic interactions of CNVs affecting a variety of genes, and the difficulty of assessing clinical phenotypes with variable expressivity and penetrance.

The most direct approach for CNV detection is using pedigree analysis where the proband and the immediate family are scanned. CNVs found in the proband but not other family members are candidates for pathogenicity. On a larger scale, the sliding window or segment-based approach scans the entire genome of controls and affected individuals for overlapping sequences and changes in the frequency of CNVs. Changes detected in one group and not the other are candidates for pathogenicity. This sliding window approach can also be used with SNP combinations to look for changes in the frequency of various combinations of SNPs.

Several studies have focused on the occurrence of duplications and deletions [92,93] in genes or gene pathways thought to be associated with the disease phenotype. These studies focus on the de novo modification of a specific gene or modifications of genes within specific pathways that have previously been associated with the disease [94] (see examples below).

Linkage disequilibrium approaches have also been used to find CNVs associated with disease phenotypes. These studies identify SNPs linked with the disease phenotype in LD and then look for CNVs associated with the SNP haplotype. Many of these CNVs are in LD with the SNPs and can lead to the detection of disease caused by structural variation [95,96] (See Figure 5.4).

Overall, the frequency of deletions associated with disease is higher than that of duplications. This may be due to the ways in which duplications and deletions are formed. Deletions can occur by intrachromatid recombination, interchromatid recombination, and interchromosomal recombination, whereas duplications can occur only by the latter two mechanisms. Deletions are also frequently associated with disease due to haploinsufficient phenotypes or exposing recessive alleles. Below are some examples of phenotypes that have been directly associated with CNVs.

Natural killer (NK) cells are cytotoxic effector cells that respond to viral infection during the innate immune response. The activation of these cells is a response to cell surface receptors including the KIR (killer immunoglobulin-like receptor) family. Individuals with certain genic KIR–ligand combinations are able to activate NK cells in the presence of a variety of pathogens. Pelak et al. [97] performed a genome-wide scan for CNVs in the region encoding KIR and showed that NK cells from individuals with multiple copies of the KIR3DL1 allele inhibited HIV-1 replication by increasing the number of peripheral blood NK cells.

Girirajan and Eichler [98] looked at recurrent CNVs associated with hotspots to identify CNVs associated with autism spectrum disorder. They examined hotspots in 2588 autistic individuals and 580 controls. They identified several recurrent CNVs, including a large duplication at 1q21 along with a deletion at 17q12, associated with autism. Looking at specific gene-disruptive recurrent CNVs, they found six genes enriched in the autism population (DPP10, PLCB1, TRPM1, NRXN1, FHIT, and HYDIN), showing that an imbalance of multiple genes contributes to the disease. Examining the phenotypes, they found that large deletions decreased nonverbal IQ. However, with large duplications, autism severity increased but nonverbal IQ was not affected.

Because noncoding DNA regulatory sequences have been estimated to constitute five to ten times as much of the genome as coding sequences, recent attention has turned toward the nontranscribed region of the genome. Haygood et al. [99] suggested that the noncoding

sequences have dominated human evolution, especially with regards to neural development and function. Investigators have examined CNVs in noncoding regions in order to detect those associated with development and disease. These studies have focused on finding CNVs in spacer sequences, gene transcription enhancer and repressor sequences, transcription factor binding sites, and epigenetic modifications (methylation of certain nucleotides) important in gene regulation and potential targets for evolution. Because most human genes are expressed in a variety of tissues at different times, a mutation in a regulatory sequence could affect a variety of tissues and alter their phenotypes. Enhancers mutate slowly and there is evidence that some are very old. Investigators have also identified CNVs in noncoding regulatory sequences that are important for spatial and temporal gene expression (genes expressed in different tissues and organs at different times). Finding nontranscribed regulator sequences has been a challenge, but new computational methods and gene reporter techniques have detected a variety of cis-regulatory regions (DNA enhancer and promoter sequences). Programs such as predicting regulatory information from single motifs (PRISM) and Genomic Regions Enrichment Tool (GREAT) search for specific motifs or analyze the significance of cis-acting sequences across the genome and not just near gene sequences, respectively [100]. Tuteja et al. [101] used GREAT to identify putative cis-regulatory sequences associated with placenta development and then identified transcription factors (TFs) associated with the placenta that could function as enhancer binding proteins. They found 2216 putative placental enhancers and found 33 known and 17 novel TFs associated with placental function, which plays a critical role in human development.

Microsatellites are repeated nucleotide sequences found in tandem in the genome. Variable-number tandem repeats (VNTRs) and short tandem repeats (STRs) found within promoter sequences (transcription start sites) can modulate gene expression and are highly conserved in the human genome. Genes that contain microsatellites near their transcription start sites are often regulatory genes involved in growth and development, with a few key examples in human brains [102]. There is some evidence that they may work through altering secondary DNA structure in the promoter region [103].

The sonic hedgehog (*SHH*) gene is a good example of a morphogen (substance influencing developmental patterns), with multiple enhancers affecting the temporal and special expression of the gene and giving rise to a variety of phenotypes. *SHH* has enhancer elements 600–900 kb upstream of the gene and within an intragenic intron. The intron-associated enhancer is responsible for expression in the posterior of the limb bud, a region called the zone of polarizing activity (ZPA), and the regulator sequence of this region is called the ZPA regulatory sequence (ZRS). A deletion upstream of the ZRS in humans can lead to polydactyly or syndactyly (additional fingers or the fusions of fingers, respectively). The region that has been deleted (*Lmbr1* upstream of ZRS) may have evolutionary significance since it is also deleted in limbless species of amphibians and in snakes. A duplication in the ZRS region can lead to triphalangeal thumb, resulting in the conversion of fingers to thumbs in the middle of the hand. Interestingly, smaller duplications in this region are associated with more severe phenotypes [79].

Another example of CNVs in noncoding regions is the *SOX9* gene. During mammalian fetal development, a gene called *SRY* (sex-determining region of the Y chromosome) is turned on in XY individuals. *SRY* codes for a transcription factor that upregulates *SOX9* expression in Sertoli cells by binding to a testis-specific enhancer upstream of *SOX9*. The *SOX9* product leads to the differentiation of gonads into testis. Duplications of a 178 kb sequence located in the *SOX9* regulatory region (600 kb upstream of *SOX9* and outside of previously identified enhancer sequences) led to male development in six individuals that were chromosomally 46XX. There are several other CNV modifications within the *SOX9* regulatory region that have led to phenotypes that are much different than those seen as a result of mutations inside the coding sequence of *SOX9* [79].

CNVs and evolution

CNVs can be a source of phenotypic and genetic variability for evolution, by creating paralogs of a gene that can then undergo mutation, and by altering gene dosage. Mutations within paralogs have been shown to act by changing gene expression and changing the coding sequence of a gene. These duplications can have advantageous effects allowing for rapid adaptation, or they can be detrimental due to their

instability and potential disruption of nearby loci. One of the gene families that has undergone extensive copy number increase consists of genes involved in the immune system. These duplications are thought to have added variability to the immune system and, through balancing selection, have allowed increased gene variation over time. The large number of duplications is thought to have allowed for rapid adaptation to new pathogenic challenges.

Because CNVs can overlap gene sequences, they may be candidates for adaptation and selection. Several populations demonstrate the selection of specific CNVs. For example, the salivary amylase gene (*AMY1*) is more abundant in the Japanese population than the Yakut population of Russia. This is thought to be due to dietary pressures faced by the two groups. The Japanese consume a large amount of starch in their diet, thus the need for a large amount of amylase, while the Yakuts are fishermen, reindeer breeders, and hunters [104].

Another example of CNV variation between populations is the apolipoprotein B mRNA-editing enzymes. These are a family of enzymes that have been found to have antiviral activity. The *APOBEC3B* gene (a member of the apolipoprotein B family) has been found to induce G to A transitions in Simian immunodeficiency virus, thereby neutralizing the virus. Kidd et al. [91] found that deletions of *APOBEC3B* are essentially fixed in Oceanic populations (93%), prevalent in East Asians and Amerindians (37–58%), and less prevalent in Africans and Europeans (0.9–6%).

Juan et al. [105] suggest that duplications in late-replicating regions may influence genome evolution. Specifically, they suggest that genes that were duplicated during primate evolution are more commonly found among human genes located in late-replicating CNV regions. These late-replicating CNVs could act as a source for the evolution of new genes. Since many genes may show detrimental effects as a result of duplication, this may explain why housekeeping genes typically occur in early replication regions. Similarly, Schuster-Böckler et al. [106] found that genes encoding proteins associated with protein–protein complexes are less likely to show duplication in humans. They also suggest that protein–protein complex interactions may be sensitive to stoichiometry and are under strong negative selection. This is supported by evidence showing that genes coding for highly interactive proteins tend to be in smaller gene families.

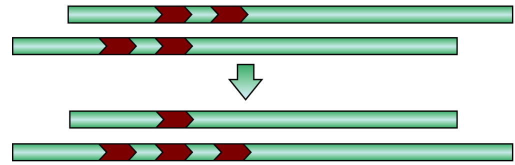


Figure 5.6 CNVs that occur by nonallelic homologous recombination. CNVs misalign on homologous chromosomes. The resulting recombination results in segmental monosomy and segmental trisomy. (Adapted from Fawcett & Hideki 2013 [107]. Reproduced with permission of Elsevier.) (See the Color Plates section.)

The selection hypothesis suggests that both positive and negative selection dictate the distribution of CNVs across the genome. Large duplications and/or deletions (>500 kb) in or near-gene regions are more likely to have deleterious effects. Some large CNVs (> 500 kb) would be eliminated by purifying selection because of their negative association with fitness and their effects on genome structure. CNVs that occur by NAHR that result in highly penetrant phenotypes, and are expressed in a variety of haplotype backgrounds, can be under strong negative selection when the phenotype decreases fitness. CNVs >100 kb are usually created de novo and can result in segmental monosomy or segmental trisomy and because of strong selection pressure may exist for only a few generations (Figure 5.6).

CNV in primates

The study of primate genomes and their comparisons may help us understand the genetic basis of phenotypes. Duplication and deletion comparisons between species have been associated with both loss and gain of traits [108,109]. CNVs and their association with specific genes and/or gene regulatory regions may have had important roles in evolution and our divergence from other primates [110,111]. Many of the CNVs that exist in humans and nonhuman primates exist in regions with high numbers of segmental duplications (SD). African great apes have a predominance of duplications in SD regions. These are not ancient CNVs but instead mark regions of recurrent CNVs due to high levels of SD. Copy number variations that occurred in our ancestors and still persist in us are an indication of positive selection for those sequences.

CNV sequences that increase or decrease the number of nucleotides between different species are sometimes

referred to as copy number differences (CND). Prufer et al. [112] compared the genomes of bonobos, chimpanzees, and humans by Illumina shotgun sequencing. Their data showed that 1.6% of the human genome is more closely related to the bonobo genome than to the chimpanzee, and that 1.7% of the human genome is more closely related to the chimpanzee than to the bonobo. They concluded that about 3% of the human genome is more closely related to either chimpanzees or bonobos than chimpanzees and bonobos are related to each other. They also speculate that, because of the similarities, the last common ancestor may have possessed an assortment of features including those present in the bonobo, chimpanzee, and human.

There are a number of CNDs that persist in all primates, and these are indicators of strong positive selection. CNDs can help uncover ancient variation that has persisted or diminished in our ancestors [113].

McLean et al. [114] identified 510 deletions in noncoding regions that were human-specific, while these same regions were highly conserved (not deleted) in chimpanzees and other mammals. These deletions were enriched in sequences near genes involved in steroid signaling and neural function. One of the human-specific deletions was located in the enhancer of the human androgen receptor (*AR*) gene. The enhancer is responsible for the development of sensory vibrissae (whiskers) and penile spines (small keratinized surface projections on the penis), and the deletion removes these anatomical features. A follow-up study by Reno et al. [115] found that the same *AR* enhancer deletion found in all humans also existed in the Neandertal and Denisovan genomes, demonstrating that the enhancer deletion is a characteristic of the human lineage, and there has been speculation that this may have also been involved with changes in our reproductive behavior. Another deletion, found by McLean et al. [100] in a noncoding region associated with a gene (*GADD45G*) involved in growth arrest, correlated with the expansion of specific brain regions in humans. Similarly, studies have shown that the hydrocephalus-inducing homolog (HYDIN) found in mice has an additional human homolog on chromosome 1. This gene is thought to be associated with regulation of brain size. Deletions of the human HYDIN homolog lead to microcephaly, while duplications lead to macrocephaly.

CYP2D6, a gene involved in metabolism, is another example of a gene that varies in copy number among primates. High copy numbers of this gene increase the metabolism of a variety of drugs, while low copy numbers cause hypersensitivity to certain drugs because of an inability to metabolize them completely. One hypothesis for the selection of multiple *CYP2D6* copies is thought to be its ability to metabolize toxins. As food toxins became less prevalent in the human diet, *CYP2D6* copy numbers dropped in many individuals, most likely due to genetic drift [116].

Chromosome rearrangements and selfish genetic elements

Chromosome rearrangements sometimes fall into the category of CNVs but can also include structural rearrangements that do not increase or decrease copy number. Inversions, reciprocal translocations, and occasionally other translocations do not typically increase gene copy number but can have effects on gene expression through position effects. Many of these rearrangements are caused by NAHR mechanisms, and large rearrangements can be detected by chromosome painting, fluorescent in situ hybridization (FISH), or Giemsa staining (see Figure 5.3). Chromosomal rearrangements are a source for reproductive barriers (low hybrid fitness) and are suppressors of recombination. The most commonly referred to rearrangements involved in human evolution are the formation of the Y chromosome and the variation of karyotypes between primates, especially the formation of human chromosome 2.

It is widely accepted that the human Y chromosome is a result of deletions, mutations, and rearrangements from an ancestral X chromosome. Comparisons of the X and Y show divergence in structure and gene content. The Y chromosome has not only lost genetic material but has also acquired a series of repetitive sequences including SINEs, endogenous retroviruses, and segmental duplications. This is most likely due to the lack of recombination between the X and Y, leading to more mutations in the Y and the presence of mostly male sex-related genes. Rearrangements and mutation rates have resulted in a Y chromosome with very little homology to the X chromosome (only 5% of the Y contains pseudoautosomal regions (recombining regions) confined to the ends of the chromosome). Comparisons of primate Y-chromosomes reveal high sequence divergences between hominoid species; however, there is lower diversity within species. The low within-species diversity

may be due to selection and genetic drift, and related to hemizygosity.

Comparisons of primate karyotypes can also reveal changes that may have led to phenotypic variability through rearrangements. The most obvious difference between primate karyotypes is the chromosome number (humans with 46 and chimpanzees, gorillas, and orangutans with 48). The difference in chromosome number is due to the formation of human chromosome 2 by the terminal fusion of two chromosomes since the divergence from chimpanzees. Further comparisons show very similar banding patterns between species, with the major changes involving paracentric and pericentric inversions of various chromosomal segments, variations in heterochromatin, and G-banding at telomeres of chimpanzee and gorillas that is absent in humans. A reciprocal translocation can also be detected in human chromosomes 5 and 17 when compared to gorillas [104]. These rearrangements may be important in determining phenotypic differences between primates based on position effects. Although primates share a high percentage of genomic sequences, their arrangements are different and those variations can be important in gene regulation.

Transposable elements

Transposable genetic elements (TEs) are ubiquitous among eukaryotes and comprise approximately 45% of the human genome. In many cases, TEs are classified as CNVs, or alternatively as selfish genetic elements (SGEs) because of their autonomous replication and random insertion. TEs vary in size and propagate within the genome through copy and paste, or cut and paste mechanisms. Because TEs are capable of moving, they are a significant source of genetic variation and can be a source of deleterious effects when moving into genes or gene regulatory regions, thereby disrupting gene function (Figure 5.7). TEs have been implicated in disease, genome rearrangements, a source of novel genes and exons, epigenetic silencing mechanisms, cis-acting regulatory elements, and drivers of speciation [117,118]. Most TEs in humans have accumulated enough mutations to prevent them from moving and are sometimes referred to as fossilized TEs. Although many of these elements have lost their ability to move, many still remain transcriptionally active. However, of those that are transcriptionally active, most are not translated. Because TE transposition in humans is often associated with germinal tissues (as opposed to somatic transposition), transposons in the germline will be passed on to the next generation. One possible mechanism associated with increased germline transposition is decreased DNA methylation in these tissues. Since methylation is proposed to downregulate transposition, demethylation of DNA during meiosis could allow for a window of increased activity [119].

The non-long terminal repeat (non-LTR) TEs in the human genome are commonly classified as SINEs (short interspersed nuclear elements) and LINEs (long interspersed nuclear elements). LINEs are retrovirus-like elements that contain their own reverse transcriptase and can contain their own endonuclease. These are the copy and paste type of element that increases in number as they replicate and move. The human genome contains more than 750,000 LINEs. The best studied of these are the L1 retrotransposons. L1 elements are considered the youngest of the LINEs and are the only ones capable of moving. L1 insertions have accumulated at higher rates in bonobos and chimpanzees than in humans [121]. They occupy about 30% of the human X chromosome and have been implicated in X-inactivation.

SINEs are elements that are less than 500 bp in length and do not contain their own reverse transcriptase, thereby relying on LINE-encoded proteins for transposition. The human genome contains more than 1.5 million SINEs located outside of the imprinted regions of the genome [122]. The most common SINEs in primates are the *Alu* sequences (named for their ability to be recognized by the *Alu* restriction enzyme). *Alus* are active nonautonomous retrotransposons occasionally associated with disease in humans. Cell stress such as heat shock and viral infection can result in transcriptional activation of *Alus*.

SVAs are composite nonautonomous retroposons that include SINE, VNTR, and *Alu* sequences. SVAs are seen in primate lineages, are regulated in trans by L1 proteins, and have been associated with some genetic diseases [123,124]. HERVs (human endogenous retroviruses) are more closely related to retroviruses like HIV and have retained their ability to replicate themselves but have lost the ability to leave the cell. There are approximately 4000–5000 copies of HERVs per haploid genome. The SINE-R element was first reported as a retrotransposon derived from an HERV. Inactive transposons like Mariner and MIR are also found in the human genome (Figure 5.8).

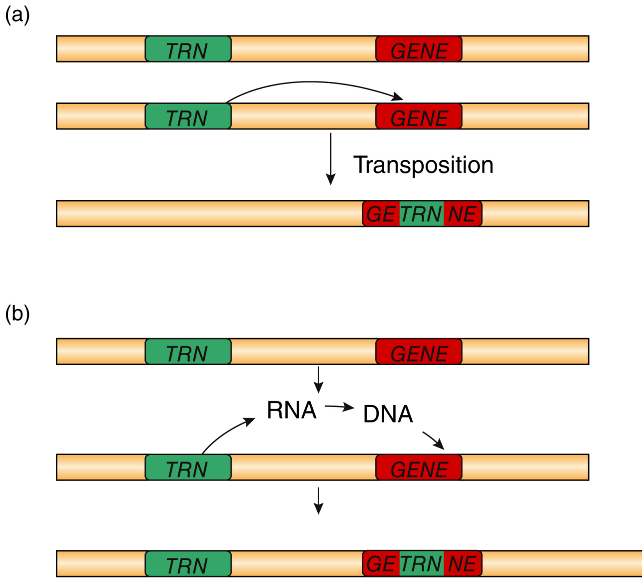


Figure 5.7 Transposition by cut and past or copy and paste mechanisms. Transposons can move by cut and paste mechanisms or copy and past mechanisms and disrupt gene activity. (a) Shows a cut and paste transposon (TRN) and a GENE on the same chromosome. The transposon is excised from its original position and inserts itself into the GENE thereby disrupting the activity of the gene. (b) Demonstrates the copy and past mechanism of transposition. The transposon (TRN) is transcribed and then converted into a new double-stranded DNA transposon. The new transposon is then inserted into the GENE, disrupting its function. Both of these examples show transposition along the chromosome that contains the transposon; however, cut and paste or copy and paste mechanisms can also occur across homologous or nonhomologous chromosomes. (Adapted from Ref. 120.) (See the Color Plates section.)

Population dynamics of transposable elements

When examining the effects of TEs on evolution, it is important to examine their molecular and population dynamics. The dynamics of TE transposition and transmission are important when considering the use of TEs to follow ancestry and the evolutionary significance of TEs in the genomes of similar but distinct species.

Transposition rates and copy numbers of the various types of elements are inconsistent within populations. If we assume there is a TE that has had little or no effect on fitness, called the source element, then in small populations a high level of copy and paste transposition by the source element could significantly alter TE frequency in the population in just a few generations. Although most of the secondary TEs (those generated by the source element) may be lost due to drift, a fraction of these secondary events may persist (approximately $1/2N_e$) and could be subject to negative selection due to reduced fitness, e.g., if one of the secondary TEs is inserted into a gene or regulatory region of a gene.

The disruption of a gene caused by the insertion of a secondary element could affect the continued presence of the source TE. If the secondary TE insertion results in a recessive phenotype, the source TE will continue to be passed on (with the secondary TE) and the secondary TE (recessive allele) will be selected against when homozygous. However, if the secondary element insertion results in a dominant phenotype, then both the secondary and the source element may be eliminated through negative selection. Negative selection is dependent on the effective population size and the selection coefficient (the measure of a genotype's relative fitness in the population). If the source locus can stay below a selection coefficient of $1/2N_e$, it will be neutral and could become fixed assuming the source element can stay below this threshold level. If the effective population size drops, then the transposition frequency per birth can increase. A subsequent higher population size may significantly reduce transposition and TE duplications. These fluctuations in population size and initial transposition events may explain the differences in primate transposition levels. The chimpanzee genome has fewer *Alus* than the human genome, and comparisons show 12 new *Alus* in the human genome and only five in the chimpanzee genome, of which only one is active. The implication is that the last common ancestor between chimpanzees and humans would have had the same number of *Alus* and that humans have generated more *Alus* since their divergence [126].

Secondary elements may themselves become source TEs and the fluctuation in population size and fluctuations in transposition activity (quiescence followed by bursts of activity) will maintain a consistent level of transposition and source elements. Because TEs have

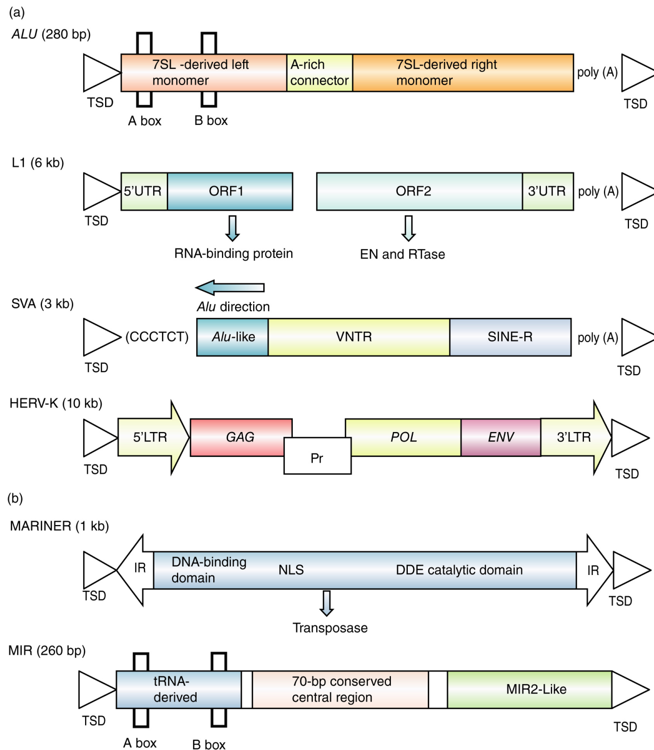


Figure 5.8 Common structures of active and inactive human transposons. (a) The structures of fourhuman transposons (ALU, L1, SVA, HERV-K) that have been active over the past 6 million years.Structures of each transposon vary and segments are identified as: 7SL is a cytoplasmic RNA that functions in protein secretion; TSD, target site duplication; ORF, open reading frame; UTR, untranslated region; VNTR, variable number of tandem repeats; SINE-R, short interspersed repetitive element; LTR, long terminal repeat; GAG, gene for retroviral core protein; POL, gene for reverse transcriptase; ENV, gene for retroviral envelope protein; Pr, protease; (b) Structures of two inactive transposons (Mariner, and MIR). IR, inverted repeats; NLS, nuclear localization structure; DDE, transposase for MARINER; MIR, mammalian-wide interspersed repeat. (*Source:* Mills et al. 2007 [125]. Reproduced with permission of Elsevier.) (See the Color Plates section.)

been successfully used to follow ancestry and the migration patterns of human populations, and because we can identify TEs across species, it is safe to assume that most of these source TEs have remained inactive for long periods in various populations and represent stable allelic markers.

Transposons in human evolution

One reason TEs are useful in the study of evolution is because they act as unique biallelic markers that are inherited by ancestry. For example, when an *Alu* element inserts in the genome randomly, it establishes a unique genetic marker whereby the ancestral state is the absence of the *Alu*. Once the *Alu* has become stabilized in the genome (not capable of moving), the probability that another random insertion occurs in the same region of the genome in another individual is infinitesimal. This allows the *Alu* to be followed from generation to generation and a person that is homozygous for the *Alu* insertion has received both copies by descent. This has resulted in the use of TEs to determine group memberships and their occasional use in forensic analysis.

Ancient stabilized transposon insertions can also be followed from species to species and there are a number of TEs that are specific to primates in addition to those that are specific to humans [127]. Because most TEs become stabilized through mutation, they can be a source of comparison between primates and have been used to determine primate phylogeny, resolving the human–chimpanzee–gorilla trichotomy [128].

Although most TEs are silenced in the human genome, there are a number of TEs that have moved in recent times, including *Alu* and L1. There is some evidence to suggest that *Alu* and L1 transposition is restricted to the germline, with a possible transposition bias in the male germline. Because TEs have been associated with genome rearrangements, disease, a source of novel genes and exons, epigenetic silencing mechanisms, and cis-acting regulatory elements, they have been proposed to have had a strong effect on human evolution and to possibly have led to the rapid divergence between humans and other primates.

One mechanism for the creation of new genes involving TEs is the insertion of TEs into an existing gene sequence. This can be the entire TE or a segment of the TE coding material. A gene that has been created by structural alteration is likely to have a different function than the parent gene. Cordaux et al. [129] discovered a primate chimeric gene resulting from the fusion of a histone methyltransferase gene and the transposase of a TE. They suggested that this type of fusion between TE segments and functional genes may have acted as a mechanism for creating new genes during human evolution.

Chimeric genes can also result from retroposition. This occurs when a gene is transcribed and the RNA product is reverse transcribed to form a DNA sequence that is inserted into or nearby an existing gene. This is known as a retrogene. In humans, many chimeric genes have recruited exons from surrounding sites. It has been estimated that retrotransposition and the formation of these chimeric retrogenes in primates occurs at a rate of 0.01–1 per million years per genome.

Another mechanism for TE involvement in evolution is the ability of L1 and SVA elements to transduce sequence beyond their 3′ ends. When these TEs move, they carry with them sequences beyond their original 3′ ends. If one of these TEs lands in or near an exon, it could transport the exon sequence and be responsible for exon shuffling and the promotion of protein evolution. It has been shown that L1 has this capability *in vivo* [126]. Exon shuffling and exon duplication can also be mediated by TEs through interchromosomal and intrachromosomal NAHR.

Alus are not capable of transduction but have been implicated in gene conversion, alternative splicing, and changes in gene expression and RNA editing. *Alus* are enriched in human genes related to neuronal functions and disease. One example of *Alu* regulating neuronal function is the deactivation of the CMP-N acetylneuraminic hydroxylase gene through gene conversion, leading to its possible involvement in the evolution of the human brain [130].

The microcephalin gene (*MCPH1*) is another example of a gene involved in brain growth that has been affected by TEs. *MCPH1* is a gene that is expressed during fetal brain development. The gene has 14 exons and many of the introns include TE sequences. The 14th exon contains 88% of an *Alu*Y sequence. The *MCPH1* gene sequence contains 57% TEs, including those in the introns. The TEs appear to have been part of the *MCPH1* introns for some time, considering the amount of mutation they have absorbed (determined by mismatches) [131]. Because the gene is functional, it appears that the insertion of TEs played a role in its evolution.

RNA editing comparisons of *Alu* sequences in the brain cells of chimpanzees, humans, and rhesus monkeys showed a higher level of adenosine-to-inosine nucleotide RNA editing in humans [132]. This suggests that increased *Alu* RNA editing may have been adapted by natural selection and may act as an alternative information mechanism for genes in the brain.

A number of studies have examined TEs in the genomes of primates and other vertebrates to better understand their role in evolution. Mills et al. [133] compared the human and chimpanzee genomes for unique TEs and found almost 11,000 TEs that were differentially present in humans and chimpanzees. *Alu*, SVA, and L1 insertions composed more than 95% of the total in both species and about 34% of the insertions were located within genes. The data suggested that there were more transpositions in humans and that the insertions represented species-specific variation that could have contributed to their divergence.

Britten [134] examined the sequences of TEs in vertebrates. He compared only full-length matches and found that of the 2732 TEs, around 1700 matched human DNA. Among all of the TEs, only *Alus* made perfect matches. All of the examples with multiple copies in humans were Y-chromosome *Alus* and are considered to be young *Alus*. There were 655 perfect full-length matches in the human genome and 283 in the chimpanzee genome that are considered recent events. When comparing humans and chimpanzees, 5530 new *Alus* were seen in humans, while 1642 were seen in chimpanzees.

One of the proposed functions of TEs in evolution is their ability to alter gene transcription either as enhancers, repressors, or insulators that block the interaction of enhancers and promoters. Because TEs have transcription regulatory sequences, they have been proposed to alter the rate of transcription or alter transcription factors that regulate gene activity. This occurs as a result of a TE inserting near or inside of a gene and changing the transcription factor binding site, thereby altering the rate of gene transcription, or the gene's tissue-specific activity. Lynch et al. [135] examined the gene regulatory mechanisms of endometrial cells using comparative RNA-Seq and found that more than 1500 genes were expressed in endometrial tissue. About 13% of these genes were within 200 kb of a Eutherian-specific TE called MER20. The genes were regulated by the MER20 that carried progesterone-

induced activity (MER20 element was induced in the presence of progesterone). The genes in the MER20 region were involved in the differentiation of human endometrium. The genes fell into two classes: (1) enhancer/repressor type responding to progesterone and (2) an insulator type that bound insulator proteins. They concluded that MER20 contributed to the origin of the regulatory network that supports pregnancy and the development of the placenta. Years earlier, investigators found an envelope protein from the HERV-W virus, called syncytin, that had been sequestered to act in human placental morphogenesis.

TE sequences have been coopted for the recombination mechanisms involved in immunoglobulin (Ig) gene rearrangements. The Igs are proteins that make up antibodies involved in the adaptive immune system. Antibodies are composed of light-chain proteins and heavy-chain proteins. The genes responsible for antibody production are extremely large and go through a splicing mechanism to form the final light- and heavy-chain transcripts that produce the proteins for antibody construction. For example, the heavy-chain IgG genes are composed of large DNA coding sequences called V, D, and J regions. The heavy- chain locus contains about 250-V regions, 12-D regions, and 4-J regions. These are cut and spliced to eventually form a segment that consists of one V, one D, and one J region (VDJ) that is translated into the heavy-chain protein of IgG. The mechanism for this rearrangement (cutting and splicing), which underlies a vertebrate's ability to recognize a diversity of antigens, relies on two enzymes called Rag1 and Rag2. These enzymes were derived from a TE transposase gene [136].

Copy number variation and TEs have played a role in the formation of the human leukocyte antigen-D-related beta chain (HLA-DRB) genes. Nine different HLA-DRB genes have been described, some of which code for functional gene products while others are pseudogenes with various indels. Several of the pseudogenes (e.g., *DRB2*, *DRB6*, and *DRB7*) appear to be the result of TE intron insertion and CNVs. Several of these pseudogenes are common in chimpanzees and humans but are not seen in other Old World monkeys [137].

Selfish genetic elements in evolution

There are a variety of genetic elements that have been classified as selfish. Selfish genetic elements (SGEs) spread regardless of their effects on the organism. SGEs can affect fitness, genome structure, and sex ratios, and there can be strong selection pressures to control their spread or their action. These elements can compete with nuclear and cytoplasmic components for transmission and selection can act on the elements to increase their transmission regardless of their effect on fitness [138]. Werren [139] and others have argued that these elements may be an important force driving evolutionary change as evidence by their increasing role in gene regulation, development, and the evolution of new species.

Among the SGEs in mammals are the transposable elements that replicate autonomously and alter genome structure, female meiotic drivers, and segregation distorters. During female meiosis, only one cell becomes the oocyte while the other cells do not typically get fertilized and degenerate. This gives a transmission advantage to the chromosomes of the oocyte and any elements that migrate with them to the egg pole. During female meiosis, some centromeric regions may undergo reduced recombination and be preferentially transmitted to the egg, indicating that there may be elements that bias segregation.

Transmission distortion in male gametophytes disrupts or otherwise outperforms their competitors and shows a bias in fertilization. From an evolutionary point of view, selection should be very strong for alleles that segregate preferentially in sperm or eggs. Although no specific female drivers or segregation distorters have been found in humans, several examples exist in mice, and there is some evidence that a more complex system may exist in humans. Zollner et al. [140] examined the genomes of individuals from 148 families and detected some transmission bias among siblings. They concluded that the bias was due to multiple transmission distortion loci (many genes) in the siblings with excess genetic sharing. New genome-wide techniques will certainly uncover more examples of drive in the future.

Genome-wide association studies

The study of complex traits is quantitative genetics, which uses statistical models to determine the contributions of environment and genetics to the expression of traits. Genomic regions that contribute to genetic variation associated with a complex trait are called

quantitative trait loci (QTLs). Phenotypic similarities among individuals can indicate similarities in genetic variance, and recent advances in genome sequencing and SNP chip technology have made it possible to map mutations responsible for complex traits.

A common approach to mapping QTLs in model organisms is to set up crosses of two inbred individuals that differ in the expression of the trait in question. The resulting progeny are then crossed to form the F2 generation. The resulting F2 individuals are examined for the trait and for cosegregation of QTL associated with the phenotype.

Because humans cannot be preferentially inbred and crossed, mapping QTLs in humans is typically done by association studies. Like QTL studies, this approach looks for phenotypic and genetic data, but unlike QTL it uses large populations of unrelated individuals. Related individuals have similar ancestry and similar genealogies, whereas unrelated individuals show different genealogies (even though they may be related in the very distant past by a common ancestor). Using large numbers and many generations of unrelated individuals from natural populations, scientists can collect phenotypic and genotypic data to try to map genes associated with a specific trait. There are two different approaches to these types of studies: (1) investigators can examine the segregation of genes that have been previously associated with the trait or expressed in the tissue or organ associated with the trait, or (2) investigators can examine the entire genome of individuals (genome-wide association testing), which may reveal more cryptic sequences or genes previously unassociated with the trait [141].

In case-control GWAS, the genomes of populations expressing a specific trait are compared to populations not expressing the trait in order to detect genomic differences that may be responsible for the trait. Evolutionary biologists are interested in the genetic factors that led to variation in complex traits broadly defined as those associated with physiology and behavior (height, weight, athletic ability, schizophrenia, intelligence, etc.). These traits are typically multifactorial (regulated by both genes and environment). The development of NGS, SNP chips, and the HapMap data has made GWAS possible [142]. Using SNP chips that cover over one million SNPs, scientists have begun to establish major and minor SNP associations with diseases such as Alzheimer's, bipolar disorder, and age-related macular degeneration [143]. In addition, principal component analysis of GWAS data can reveal population stratification (regional variation in SNP frequencies) and specific geographical differentiation, which may indicate positive selection of specific SNPs.

GWAS look for SNP frequencies in test and control populations in order to establish genetic correlations with a given trait. GWAS identifies high-frequency markers associated with the trait (those markers that are different between the test and control populations). The markers in high frequency associated with the phenotype are selected for further study.

SNP chips are used to screen large populations. The assumption is that SNPs or SNP tags that segregate with the trait will be associated with the gene or closely linked to a gene in linkage disequilibrium (LD), that is, the SNP is part of a haplotype block. Because this method depends somewhat on LD, a low LD weakens the ability of GWAS to detect alleles associated with the complex trait.

The use of SNP chip arrays that are directed toward specific loci thought to be associated with the disease can be an effective approach for GWAS. The Immunochip, iCOG array, and the Metabochip are directed toward autoimmune disease, cancer, and metabolic disorders, respectively. These chips focus on specific phenotypes and can act as a follow-up to mutations detected by GWAS. The use of these focused chips and NGS can help to reveal more variants associated with disease.

Normally, natural selection will increase the frequency of favorable alleles and decrease the frequency of unfavorable alleles. In complex traits, however, with large numbers of alleles, many of the alleles will not be selectively eliminated even though they may contribute to an unfavorable phenotype. For example, complex disease may result from a variety of common deleterious alleles that contribute only a portion of the total phenotype. Individuals in a population may have one or more of the common variant alleles and not express the trait because the individual alleles are selectively neutral. These common variants typically exist inside a gene or gene regulatory sequence and can lead to disease susceptibility. When all (or most) variants are present, the combination can lead to the disease phenotype. GWAS is primarily based on the common-disease common-variant model (CDCV). Common variants are those that exist in high frequency in a population. The CDCV model predicts that common disease variants will be found in all populations that have the disease trait, but that a single common variant will not necessarily trigger the disease.

The CDCV model has proven very successful in most GWAS; however, even in the most successful studies,

there still appears to be "missing heritability" (markers that have been missed by current GWAS practices). One model suggests that the missing heritability may be due to a diversity of different rare alleles that act to drive complex disease. This model is known as the multiple rare variant (MRV) hypothesis. The model suggests that a series of rare incompletely penetrant alleles at low frequency confer disease risk, and that these may lead to a significant proportion of susceptibility. There is some evidence that a proportion of variants revealed in GWAS may be the result of a rare variant, giving rise to what is known as "synthetic association." Under synthetic association, the variant would be a considerable distance from the (common) associated variant [144–146]. Saunders et al. [147] reported the first evidence of a synthetic association contributing to pancreatic cancer susceptibility, and suggested that some GWAS signals may be the result of synthetic association. Current GWAS may skew results by eliminating low-frequency markers. Proponents of CDCV and MRV agree that the controversy can only be settled by whole genome sequencing [148].

Initial GWAS were done on European populations that skewed the tag SNPs toward these populations, and made SNP arrays less effective (attenuated signals due to different allele frequencies and weaker LD) for non-European and ethnically mixed populations.

Genome-wide meta-analysis (GWMA) using thousands of samples is typically used to reveal potential candidates for different phenotypes. Meta analysis combines results from different studies and GWMA relies on the same SNPs displaying consistently across multiple populations. This assumes that the same causal variant is present within the populations, the same LD pattern exists between the causal variant and the assayed SNPs, and the effective blocks at the assayed SNPs are consistent. Meta analysis also requires that the same SNPs be genotyped when looking at different populations sometimes requiring imputation. Imputation is not always possible because it requires appropriate reference panels that are sometimes not available.

A modified meta-analysis approach proposed by Wang et al. [149] is to examine predefined genomic regions (based on statistical significance) and determine the degree of overrepresentation of associated SNPs by measuring the LD between every possible pair of SNPs. The evidence for phenotypic association is the extent of overrepresentation of independent associated SNPs against the total number of independent SNPs in a window of fixed length (the window is a segment of an SNP block). An overrepresentation of statistically significant SNPs in the region would constitute evidence that the region is involved in the phenotype, with the higher overrepresentation indicating stronger evidence for association with the phenotype. This can be done across independent populations by searching the regions to strengthen the association and this method would help to distinguish rare events. This sliding window approach has been used to combine several studies from diverse populations.

Common variants that are found in noncoding regions of the DNA can alter the expression of genes, and thereby alter mRNA levels contributing to phenotypic variation and complex disease. Examining mRNA expression levels can help identify genes controlling complex phenotypes. Expression QTLs (eQTLs) are loci that regulate the expression of mRNA. Classically, specific noncoding regions regulate transcription and mRNA levels of a specific gene and map close to the gene. eQTLs can map close to their regulated gene (local eQTL or cis-eQLT) or can map distantly from their regulated gene—even on another chromosome (distant eQTL or trans eQTL). SNPs can alter the regulatory ability (quantity of transcript) in these eQTLs. Studies have shown that SNPs closely associated with disease are significantly enriched in eQTLs. In these studies, mRNA expression levels are treated as phenotypes that are measured by microarrays and mapped to specific genomic regions. eQTL studies statistically measure the gene expression levels to look for significant correlations between genome variation and transcriptome variation. In some cases, eQTL hotspots have been revealed. An eQTL hotspot is a genetic region associated with altering the expression levels of many genes, and may consist of a transcription factor or several linked loci. eQTLs have been useful in GWAS by revealing some of the noncoding regions involved in complex disease. eQTL studies can reveal CNVs and tissue-specific differences in gene expression between test and control groups [150].

eQTL has been used to study innate immunity in healthy humans. Fairfax et al. [151] used monocytes from 432 healthy Europeans and exposed the cells to g-interferon or lipopolysaccharides (LPS) for 2 and 24 h. They found cis-eQTLs and trans-eQTLs associated with the specific treatments. In another study, Lee et al. [152] used dendritic cells from Europeans, African Americans, and Asians. They exposed the cells to influenza virus, beta-interferon, and LPS and identified 121 cis-eQTLs that responded to one or more of the treatments, and 57

cis-eQTLs that responded to all three treatments. These studies showed an overlap in eQTLs loci previously associated with autoimmune disease.

Massively parallel RNA sequencing (RNA-seq) measures the expression output of each locus and can be used to quantify alternatively spliced transcripts, revealing the mechanisms behind eQTLs. RNA-seq studies combine transcriptional and genotypic data and have greatly enhanced the ability to detect expression variation and noncoding RNA species.

RNA-seq studies use mRNA that is converted into cDNA. The cDNA is treated similar to NGS where it is sheared into small fragments and sequenced in parallel. Computer alignment generates a number of overlapping sequences that can reveal gene structure based on the number and placement of splice junctions. Most of the RNA-seq reads exons and the average number gives an estimate of the activity of the gene output. A combination of exon reads can indicate which exons are expressed more or less often, indicating alternative splicing. SNP variations in heterozygotes can be used for comparisons to determine allelic imbalance in expression levels (where one allele produces more transcripts than the other).

Measuring the contribution of SNPs or other markers associated with a disease phenotype is accomplished using the odds ratio (OR). Odds ratios quantify the relationships between two properties in a given population, that is, whether each individual in a population either has one property or not. This ratio is useful for measuring case-control studies and the strength of association. For example, the properties could be the presence of an SNP (S) and the presence of a disease (D). The OR can be computed with the following method:

For an individual that has S, compute the odds that the individual also has D.

Then:

For an individual that does not have S, compute the odds that the individual also has D.

Then:

Divide the odds from each computation above to determine the odds ratio.

Another common way of quantifying association is by the risk ratio (RR). The RR is determined in a similar way to the OR except using probabilities instead of odds. For example, if we are looking at the total number of individuals in a population that have a specific SNP haplotype (P_S), and we determine the percentage of the population that has a disease trait (D_S) and the percentage of the population that is healthy (H_S), then $P_S = D_S + H_S$. For another population without the SNP haplotype (P_{NS}), and the percentage of those individuals with the disease (D_{NS}) and those that are healthy (H_{NS}), then $P_{NS} = D_{NS} + H_{NS}$. The risk of developing the disease with the SNP haplotype is D_S/P_S and that of developing the disease without the SNP is D_{NS}/P_{NS}. The resulting risk ratio (RR) is

$$RR = (D_S/P_S)/(D_{NS}/P_{NS})$$

$$\text{which can be rewritten as } (D_S\, P_{NS})/(D_{NS}\, P_S)$$

GWAS has been successfully used to find loci, haplotypes, and microRNA (miRNA) associated with complex disease. Hass et al. [153] combined genetics, imaging of the hippocampus, and neurophysiological data to study schizophrenia. They obtained brain scans from 328 individuals and found six SNPs associated with an LD block on chromosome 19, and four SNPs on chromosomes 1, 2, and 10 that were highly correlated with schizophrenia. The six SNPs were associated directly and indirectly with genes involved in hippocampal and brain development, which has been associated with schizophrenia. Three other genes were also identified with unknown function and two genes showed cis-acting variation associated with mRNA expression.

Chen et al. [154] identified 19 common variants associated with breast cancer in African American women, and in a related study, Song et al. [155] used genome-wide haplotype analysis to examine breast cancer risk in African American women. They examined 3016 test individuals and 2745 controls and using a 5-SNP sliding window approach examined over one million SNPs. They found three novel regions on chromosomes 1, 4, and 18 that exhibited moderate affects, and screened previously identified regions on chromosomes 10 and 14 and found moderate haplotype effects.

GWAS cancer studies have revealed changes in miRNA levels associated with a variety of cancers. MicroRNAs are small noncoding RNAs that regulate gene expression through a posttranscriptional process. Previous studies have shown that miRNAs are polycistronic and are functionally related by targeting the same gene or a group of genes in a pathway [156]. These clustered miRNAs have been proposed to play a role in oncogene suppression. Laddha et al. [157] studied the miRNA cluster in a conserved imprinted locus on human chromosome 14q32. The miRNA cluster is known as miR-379/miR-656 and is unique to all placental mammals.

They found a 68% reduction in miRNA cluster activity associated with a glioblastoma, a 61% reduction associated with a kidney carcinoma, and a 46% reduction associated with a breast cancer.

GWAS has also been used to look at mating behavior. A study by Dominque et al. [158] looked at genome-wide genetic assortative mating using non-Hispanic white adults to see if married couples shared more of their genome than nonmarried couples. They examined 1.7 million SNPs and found that spouses were more genetically similar at the SNP level than two individuals chosen at random. They compared these results to prior studies, looking at education selective sorting and found that the genome selective sorting was only one-third the magnitude of the prvious study of education selective sorting. Race/ethnicity was held constant to eliminate bias and because European populations show a low degree of variability in allele frequencies.

GWAS can also examine CNVs and CNV frequencies between controls and test populations to identify CNV involvement with the expression of a trait. Genome-wide oligonucleotide arrays have been developed to detect CNVs in GWAS. Lee et al. [94] found several large deletions and duplications associated with schizophrenia. However, they and others have concluded that there is still a need for better software programs to analyze the CNV data and to improve integrating the data with various research platforms including proteomics, and transcriptome analysis.

GWAS have been successful on a number of levels. Studies have found a variety of markers that had previously not been associated with the phenotype; the comparisons of various studies and populations have shown overlap in the identified markers, indicating a strong connection to the phenotype; and the loci found have been shown to have small cumulative effects on the phenotype.

Concerns over the effective use of GWAS

Although GWAS has been successful in identifying a large number of variants associated with disease, there are still gaps of missing heritability in the screens. Several investigators have proposed multiple rare variants that contribute a small amount to the phenotype, or low frequency of rare markers to explain the missing heritability; however, because of the few cases where rare variants have been shown to be involved, the common variant hypothesis is thought to be most appropriate for GWAS. This is especially true with initial evolutionary studies where common variants lead to pathways and major contributors of complex phenotypes. Common variants associated with complex phenotypes, such as intelligence, weight, and height, can reveal some of the evolutionary mechanisms that may have led to our current state. Further knowledge of the evolution of complex traits will be revealed in GWAS across related species by showing differences in LD, population structure, allele frequencies, and possibly help lead to understanding causation through supplementary biochemical and functional studies.

GWAS have begun to point to genetic associations with some diseases and complex phenotypes; however, the studies do not always correlate strongly with the particular trait. Many times, this is due to population stratification, small sample sizes, variation in the frequency of causative mutations, wide variation in the expressivity of the trait in question, genetic heterogeneity, and insufficient LD when using imputation to fill GWAS gaps. All of these can complicate the analysis. In order for the statistical analysis to generate data without false positives, a minimum of 5000 individuals are needed for correlations in the 99% range. Correlations of 99.9% require 100,000 individuals or more. It is important that GWAS studies be replicated on different populations using different screening technology and different methods of analysis in order to avoid any false positive associations.

Conclusions

The advent of molecular biology significantly changed the way that scientists study evolution. Today, molecular biologists use various molecular markers (SNPs, CNVs, VNTRs, TEs, etc.) and various molecular techniques (NGS, FISH, PRT, etc.) to study human evolution. Molecular biological markers, NGS, and anthropological markers have begun to clarify the contributions of punctuated equilibrium and gradualism to the origin of species. Understanding the mechanisms responsible for recombination, linkage disequilibrium, copy number variation, transposition, and chromosome rearrangements have provided new insights into how human

genome structure evolved and how humans are connected to other primates and other species. A number of important variations including deletions, duplications, insertions, nucleotide substitutions, and transpositions differ between primate species and may explain phenotypic differences. Phenotypic differences between chimpanzees and humans may be based on these variations affecting coding sequences and gene dosage. In addition, these variations may affect noncoding gene regulatory sequences leading to different patterns of gene expression in tissues and organs. Molecular biology and GWAS have led to a better understanding of how genome structure, gene regulation, and metabolism can lead to complex phenotypes. The molecular study of migration patterns and admixture is providing a better understanding of our genetic backgrounds and setting the stage for personalized medicine through the use of pharmacogenetics and pharmacokinetics. Further studies of genomics and proteomics may reveal the mechanisms that led to our current physical and mental state and may provide insights into the future of human evolution.

Review questions and exercises

1 Given the following frequency data, what is the maximum value for D?

$A=0.5$ $a=0.5$ $B=0.5$ $b=0.5$

2 Given the following data, determine D.

Allele frequencies:	$A=0.9$	$a=0.1$	$B=0.5$	$b=0.5$
Genotype frequencies:	$AB=0.5$	$ab=0.1$	$Ab=0.4$	$aB=0.0$

3 a. Using the allele frequencies and the genotypic frequencies below, determine D.

b.

Allele frequencies:	$A=0.9$	$a=0.1$	$B=0.5$	$b=0.5$
Genotypic frequencies:	$AB=0.45$	$ab=0.12$	$Ab=0.39$	$aB=0.04$

c. Determine r^2 for the above.

d. If the recombination rate is determined to be 0.12, determine the new value of D.

e. If the recombination rate is 0.12, determine the effective population size assuming no mutation.

4 Two alleles of two different genes are linked. The genes are G and M and the alleles are G and g, and M and m. The observed genotypic frequency in a population of 700 individuals is:

GGMM = 225;	GgMM = 105;	ggMM = 16
GGMm = 230;	GgMm = 65;	ggMm = 0
GGmm = 59;	Ggmm = 0;	ggmm = 0

a. Determine the genotypic frequencies of GG, Gg, and gg; and MM, Mm, and mm, then determine the frequency of G, M, g, and m.

b. Assuming they are in Hardy–Weinberg Equilibrium, estimate the frequency of GM, Gm, gM, and gm. Calculate the LD between the two markers.

c. Calculate the LD between the markers.

5 Copy number variations are typically identified as phenotypes 0, 1, 3, 4, or more, with 0 identifying individuals who lost both loci (sequences on both chromosomes), 1 identifying individuals with one sequence present on one chromosome, 3 one extra copy of the locus compared to diploid, etc.

A woman that has a CNV phenotype of 3 marries and man with a phenotype of 1. What are the possible CNV phenotypes of their children? Assume no NAHR and the extra locus in the mother is on the homologous chromosome.

6 Explain how GWAS differs from whole genome sequencing when looking for loci involved with complex traits.

7 Explain why GWAS provide correlation and not causation data.

8 a. A family is subjected to microsatellite analysis for three linked loci (A, B, and C). The numbers given for each individual represent the alleles (repeats) for the microsatellites. Determine the haplotypes of the mother and father assuming no recombination.

Father	A(13, 17)	B(10, 19)	C(18, 8)
Mother	A(12, 17)	B(16, 19)	C(8, 8)
Child 1	A(13, 17)	B(10, 16)	C(18, 8)
Child 2	A(17, 17)	B(19, 16)	C(8, 8)
Child 3	A(13, 12)	B(10, 19)	C(18, 8)

b. If recombination occurs in the father between microsatellites A and B, what will be the new haplotypes? Assume that the order of loci is A–B–C.

9 Next-generation sequencing has determined that human and chimpanzee DNA are 98.8% identical. Why are the phenotypes of humans and chimps different?

10 Explain some of the problems involved with using GWAS to study Alzheimer's disease.

11 Explain eQTLs and their significance in looking at the genetics behind complex phenotypes.

References

1 Osler W. 1913. *Way of Life: An Address to Yale Students, Sunday Evening*. Charles C Thomas Publishers Ltd, New York, NY.

2 Gould SJ. 1980. *The Pandas Thumb: More Reflections in Natural History*. W.W. Norton & Company, New York, NY, pp 179–184.

3 Gilbert S. 2000. *Developmental Biology*, 6th edition. Sinauer Associates Inc., Sunderland, MA.

4 Gould SJ. 1977. The return of hopeful monsters. *Nat Hist* **86**:22–30.

5 Gould SJ, Eldredge N. 1977. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* **3**:115–151.

6 Meyer A. 2005. On the importance of being Ernst Mayr. *PLoS Biol* **3**:e15.

7 Webster AJ et al. 2003. Molecular phylogenies link rates of evolution and speciation. *Science* **301**:47.

8 Kopp M, Matuszewski S. 2013. Rapid evolution of quantitative traits: theoretical perspectives. *Evol Appl* **7**:169–191.

9 Luca F. 2012. Diet, changes in gene expression and human evolution. In: *eLS*. John Wiley and Sons, Chichester, UK, pp 1–7.

10 Horovath J et al. 2005. Punctuated duplication seeding events during the evolution of human chromosome 2p11. *Genome Res* **15**:914–927.

11 Jiang Z et al. 2007. Ancestral reconstruction of segmental duplications reveals punctuated cores of human genome evolution. *Nat Genet* **39**:1361–1368.

12 Shultz S. 2012. Hominin cognitive evolution: identifying patterns and processes in the fossil and archeological record. *Phil Trans R Soc* **367**:2130–2140.

13 International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:30–91.

14 Venter JC et al. 2001. The sequence of the human genome. *Science* **291**:1304–1351.

15 The 1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**:56–65.

16 Sulston J, Ferry G. 2002. *The Common Thread: A Story of Science, Politics, Ethics, and the Human Genome*. The Joseph Henry Press, Washington, DC.

17 Dhand R. 2006. The "finished" landscape. *Nature* 1–305.

18 Kilpinen H, Barrett J. 2013. How next-generation sequencing is transforming complex disease genetics. *Trend Genet* **29**:23–30.

19 Sankararaman S et al. 2014. The genomic landscape of Neanderthal ancestry in present day humans. *Nature* **507**:354–357.

20 Dunham I et al. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**:57–74.

21 Brown EA. 2013 Many ways to die, only one-way to arrive: how selection acts through pregnancy. *Trend Genet* **29**:585–592.

22 Boycott KM et al. 2014. Identification of genes for childhood heritable diseases. *Ann Rev Med* **65**:19–31.

23 Kolliker M et al. 2014. Evolution of parental care. In: *The Princeton Guide to Evolution*. Princeton University Press, Princeton NJ, pp 663–670.

24 Weber-Lehmann J. et al. 2014. Finding the needle in the haystack: differentiating "identical" twins in paternity testing and forensics by ultra-deep next generation sequencing. *Forensic Sci Int Gen* **9**:42–46.

25 Walther P et al. 2013. Evaluation of next generation mtgenome sequencing using the Ion Torrent Personal Genome Machine (PGM). *Forensic Sci Int Gen* **7**:632–639.

26 Fu W et al. 2013. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**:216–220.

27 Pennisi E. 2014. DNA sequencers still waiting for the nanopore revolution. *Science* **343**:829–830.

28 Barbujani G, Colonna V. 2010. Human genome diversity: frequently asked questions. *Trend Genet* **26**:285–295.

29 Lynch M. 2010. Evolution of the mutation rate. *Trend Genet* **26**:345–352.

30 Campbell CD, Eichler EE. 2013. Properties and rates of germline mutations in humans. *Trend Genet* **29**:575–584.

31 Hinds DA et al. 2005. Whole-genome patterns of common DNA variation in three human populations. *Science* **307**:1072–1079.

32 Goldstein DB, Chihi L. 2002. Human migration patterns and population structure: what we know and why it matters. *Ann Rev Genomics Hum Genet* **3**:129–152.

33 Hedrick P. 2014. Genetic drift. In: *The Princeton Guide to Evolution*. Princeton University Press, Princeton NJ, pp 307–314.
34 Petrov DA. 2014. Searching for adaptation in the genome. In: *The Princeton Guide to Evolution*. Princeton University Press, Princeton, NJ, pp 466–474.
35 Buard J, de Massy B. 2007. Playing hide and seek with mammalian meiotic crossover hotspots. *Trend Genet* **23**:301–309.
36 Fawcett JA, Innan H. 2013. The role of gene conversion in preserving rearrangement hotspots in the human genome. *Trend Genet* **29**:1–8.
37 Baudat F et al. 2009. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327**:836–840.
38 Kong A et al. 2002. A high-resolution recombination map of the human genome. *Nat Genet* **31**:241–247.
39 Sandovici I et al. 2006. Human imprinted chromosomal regions are historical hot-spots of recombination. *PLoS Genet* **2**:e101.
40 Myers S et al. 2005. A fine scale map of recombination rates and hotspots across the human genome. *Science* **310**:321–324.
41 Jeffreys AJ, Neumann R. 2002. Reciprocal crossover asymmetry and meiotic drive in a human recombination hotspot. *Nat Genet* **31**:267–271.
42 Jeffreys AJ, Neumann R. 2005. Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. *Hum Mol Genet* **14**:2277–2287.
43 Winkler W et al. 2005. Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* **308**:107–111.
44 Munch K et al. 2014. A fine-scale recombination map of the human-chimpanzee ancestor reveals faster change in human than in chimpanzees and a strong impact of GC biased gene conversion. *Genome Res* **24**:467–474.
45 Webster MT, Hurst L. 2011. Direct and indirect consequences of meiotic recombination: implications for genome evolution. *Trend Genet* **28**:101–109.
46 Rannala B, Slatkin M. 1998. Likelihood analysis of disequilibrium mapping, and related problems. *Am J Hum Genet* **62**:459–473.
47 Slatkin M. 2008. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* **9**:477–485.
48 Weiss KM, Clark AG. 2002. Linkage disequilibrium and the mapping of complex human traits. *Trend Genet* **18**:19–24.
49 Barton NH. 2010. Genetic linkage and natural selection. *Phil Trans R Soc B* **365**:2559–2569.
50 Ennis S. 2007. Linkage disequilibrium as a tool for detecting signatures of natural selection. *Method Mol Biol* **376**:59–70.
51 Ohta T. 1982. Linkage disequilibrium due to random genetic drift in finite subdivided populations. *Proc Natl Acad Sci USA* **79**:1940–1944.
52 Moorjani P et al. 2011. The history of African gene flow into Southern Europeans, Levantines, and Jews. *PLoS Genet* **7**: e1001373.
53 Vernot B, Akey JM. 2014. Resurrecting surviving Neanderthal lineages from modern human genomes. *Science* **343**:1017–1021.
54 Pugach I et al. 2011. Dating the age of admixture via wavelet transformed analysis of genome-wide data. *Genome Biol* **12**:R19.
55 Loh PR et al. 2013. Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **193**:1233–1254.
56 Hellenthal G et al. 2014. A genetic atlas of human admixture history. *Science* **343**:747–751.
57 Elhaik E et al. 2014. Geographic population structure analysis of worldwide human populations infers their biogeographical origins. *Nat Commun* **5**:3513.
58 Yang WY et al. 2012. A model-based approach for analysis of spatial structure in genetic data. *Nat Genet* **44**:725–731.
59 Ewens WJ, Spielman RS. 1995. The transmission/disequilibrium test: history, subdivision, and admixture. *Am J Hum Genet* **57**:455–464.
60 Schaid DJ. 1998. Transmission disequilibrium, family controls, and great expectations. *Am J Hum Genet* **63**:935–941.
61 Barrett JC et al. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**:263–265.
62 Melchiotti R et al. 2014. ArchiLD: hierarchical visualization of linkage disequilibrium in human populations. *PLoS One* **9**:e86761.
63 Iskow RC 2012. Exploring the role of copy number variants in human adaptation. *Trend Genet* **28**:249.
64 Lodish H et al. 2008. *Molecular Cell Biology*, 6th edition. Freeman, 202.
65 Sabeti PC et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**:832–837.
66 Sabeti PC et al. 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**:913–918.
67 Voight BF et al. 2007. A map of recent positive selection in the human genome. *PLoS Biol* **4**:e154.
68 Little CC. 1941. The genetics of tumor transplantation. In: Snell GD (ed.), *Biology of the Laboratory Mouse*. Dover, New York, pp 279–309.
69 Daly MJ et al. 2001. High-resolution haplotype structure in the human genome. *Nat Genet* **29**:229.
70 Ahmad T et al. 2002. Haplotype-specific linkage disequilibrium patterns define the genetic topography of the human MHC. *Hum Mol Genet* **12**:647–656.
71 Kauppi L et al. 2003. Recombination hotspots rather than population history dominate linkage disequilibrium in the MHC class II region. *Hum Mol Genet* **12**:33–40.
72 Li H et al. 2013. Patterns of linkage disequilibrium of LRRK2 across different races: implications for genetic association studies. *PLoS One* **8**:e75041.
73 Laan M, Paabo S. 1997. Demographic history and linkage disequilibrium in human populations. *Nat Genet* **17**:435–438.

74 Hawi Z et al. 2013. A high density linkage disequilibrium mapping in 14 noradrenergic genes: evidence of association between SLC6A2, ADRA1B, and ADHD. *Psychopharm* **225**:895–902.
75 Itsara A et al. 2009 Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet* **84**:148–161.
76 Donald FC et al. 2010. Origins and functional impact of copy number variation in the human genome. *Nature* **464**:704–712.
77 Hastings PJ et al. 2009. Mechanisms of change in gene copy number. *Nat Rev Genet* **10**:551–564.
78 Su Z, Gu X. 2012. Revisit on the evolutionary relationship between alternative splicing and gene duplication. *Gene* **504**:102–106.
79 Klopoki E, Mundlos S. 2011. Copy-number variations, noncoding sequences, and human phenotypes. *Ann Rev Genomics Hum Genet* **12**:53–72.
80 Ostertag EM et al. 2003. SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am J Hum Genet* **73**:1444–1451.
81 Hastings PJ et al. 2009. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet* **5**:e1000327.
82 Lu J et al. 2014. The distribution of genomic variations in human iPSCs is related to replication-timing reorganization during reprogramming. *Cell Rep* **7**:70–78.
83 Golzio C, Katsanis N. 2013. Genetic architecture of reciprocal CNVs. *Curr Opin Genet Devel* **23**:240–248.
84 Redon R et al. 2006. Global variation in copy number in the human genome. *Nature* **444**:444–454.
85 Kidd JM et al. 2010 A human genome structural variation sequencing resource reveals insights into mutational mechanisms. *Cell* **143**:837–847.
86 Zhang D et al. 2011. Accuracy of CNV Detection from GWAS Data. *PLoS One* **6**:e0014511.
87 Mills RE et al. 2011. Mapping copy number variation by population-scale genome sequencing. *Nature* **470**:59–65.
88 Glessner JT et al. 2009. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* **459**:569–573.
89 Glessner JT. 2010. Strong synaptic transmission impact by copy number variations in schizophrenia. *Proc Nat Acad Sci USA* **107**:10584–10589.
90 Gonzalez E et al. 2005. The influence of CCL3L1 gene containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307**:1434–1440.
91 Kidd JM et al. 2007. Population stratification of a common APOBEC gene deletion polymorphism. *PLoS Genet* **3**:e63.
92 Conrad DF et al. 2006. A high resolution survey of deletion polymorphism in the human genome. *Nat Genet* **38**:75–81.
93 Hinds DA et al. 2006. Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nat Genet* **38**:82–85.
94 Lee KW et al. 2012. Genome wide association studies (GWAS) and copy number variation (CNV) studies of the major psychoses: what have we learnt? *Neurosci Biobehav R* **36**:556–571.
95 Conrad DF et al. 2010. Origins and functional impact of copy number variation in the human genome. *Nature* **464**:704–712.
96 Cooper GM et al. 2008. Systematic assessment of copy number variation detection via genome wide SNP genotyping. *Nat Genet* **40**:1199–1203.
97 Pelak K et al. 2011. Copy number variation of KIR genes influences HIV-1 control. *PLoS Biol* **9**:e1001208.
98 Girirajan S, Eichler EE. 2010. Phenotypic variability and genetic susceptibility to genetic disorders. *Hum Mol Genet* **12**:849–58.
99 Haygood R et al. 2010. Contrasts between adaptive coding and non-coding changes during human evolution. *Proc Nat Acad Sci USA* **107**:7853–7857.
100 McLean CY et al. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotech* **28**:495–501.
101 Tuteja G et al. 2014. Automated discovery of tissue-targeting enhancers and transcription factors from binding motif and gene function data. *PLoS Comput Biol* **10**:e1003449.
102 Sawaya SM et al. 2012. Measuring microsatellite conservation in mammalian evolution with a phylogenetic birth-death model. *Genome Biol Evol* **4**:636–647.
103 Sawaya SM et al. 2012. Promoter microsatellites as modulators of human gene expression. *Adv Exp Med Biol* **769**:41–54.
104 Jobling M et al. 2014. *Human Evolutionary Genetics*, 2nd edition. Garland Science, New York, NY.
105 Juan D et al. 2013. Emerging methods in protein co-evolution. *Nat Rev Genet* **14**:249–261.
106 Schuster-Böckler B et al. 2010. Dosage sensitivity shapes the evolution of copy-number varied regions. *PLoS One* **5**:e9474.
107 Fawcett JA, Hideki I. 2013. The role of gene conversion in preserving rearrangement hotspot in the human genome. *Trend Genet* **29**:562.
108 Dumas L et al. 2007. Gene copy number variation spanning 60 million years of human and primate evolution. *Genome Res* **17**:1266–1277.
109 Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet* **11**:97–108.
110 James R et al. 2014. tRNA gene copy number variation in humans. *Gene* **536**:376–384.
111 Popesco MC et al. 2006. Human lineage-specific amplification, selection and neuronal expression of DUF1220 domains. *Science* **313**:1304–1307.
112 Prufer K et al. 2012. The bonobo genome compared with the chimpanzee and human genomes. *Nature* **486**:527–531.
113 Marques-Bonet T et al. 2009. A burst of segmental duplications in the genome of the African great ape ancestor. *Nature* **457**:877–881.

114 McLean CY et al. 2011. Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* **471**:216–219.

115 Reno PL et al. 2013. A penile spine/vibrissa enhancer sequence is missing in modern and extinct humans but is retained in multiple primates with penile spines and sensory vibrissae. *PLoS One* **8**:e84258.

116 Gaedigk A et al. 2011. Detection of CYP2D6, SULT1A1 and UGT2B17 copy number variation using multiplex PCR amplification. *FASEB J Suppl* **25**:812.

117 Hedges DJ, Batzer MA. 2005. From the margins of the genome mobile elements shape primate evolution. *Bioessays* **27**:785–794.

118 Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10**:691–703.

119 Suzuki S et al. 2007. Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting. *PLoS Genet* **3**:e55.

120 Atherly AG et al. 1999. *The Science of Genetics*, 1st edition. Harcourt, 592.

121 Marchetto MC et al. 2013. Differential L1 regulation in pluripotent stem cells of humans and apes. *Nature* **503**:525–529.

122 Sen SK et al. 2006. Human genomic deletions mediated by recombination between *Alu* elements. *Am J Hum Genet* **79**:41–53.

123 Bundo M et al. 2014. Increased L1 retrotransposition in the neuronal genome in schizophrenia. *Neuron* **81**:306–313.

124 Raiz J et al. 2012. The non-autonomous retrotransposon SVA is *trans*-mobilized by the human LINE-1 protein machinery. *Nucleic Acids Res* **40**:1666–1683.

125 Mills RE et al. 2007. Which transposable elements are active in the human genome. *Trend Genet* **23**:186.

126 Xing J et al. 2006. Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proc Natl Acad Sci USA* **103**:17608–17613.

127 Jin P et al. 2012. Evolutionary rate of human tissue-specific genes are related with transposable element insertions. *Genetica* **140**:513–523.

128 Mathews LM et al. 2003. Large differences between LINE-1 amplification rates in the human and chimpanzee lineages. *Am J Hum Genet* **72**:739–748.

129 Cordaux R et al. 2006. Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proc Nat Acad Sci USA* **103**:8101–8106.

130 Hayakawa T et al. 2001. Alu-mediated inactivation of the human N-acetylneuraminic acid hydroxylase gene. *Proc Nat Acad Sci USA* **98**:11399–11404.

131 Shi L et al. 2013. Functional divergence of the brain size regulatory gene *MCPH1* during primate evolution and the origin of humans. *BMC Biol* **11**:1–6.

132 Bazak L et al. 2014. A-to-I editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res* **24**:365–376.

133 Mills RE et al. 2006. Recently mobilized transposons in the human and chimpanzee genomes. *Am J Hum Genet* **78**:671–679.

134 Britten RJ. 2010. Transposable element insertions have strongly affected human evolution. *Proc Nat Acad Sci USA* **107**:19945–19948.

135 Lynch VJ et al. 2011. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat Genet* **43**:1154–1159.

136 Jeffreys AJ, Neumann R. 2005. Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. *Hum Mol Genet* **14**:2277–2287.

137 Doxiadis GG et al. 2012. Evolution of HLA-DRB genes. *Mol Biol Evol* **29**:3843–3853.

138 Fishman L, Jaenike J. 2014. Selfish genetic elements and genetic conflict. In: *The Princeton Guide to Evolution*. Princeton University Press, NJ, pp 347–355.

139 Werren JH. 2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc Natl Acad Sci USA* **108**:10863–10870.

140 Zollner S et al. 2004. Evidence for extensive transmission distortion in the human genome. *Am J Hum Genet* **74**:62–72.

141 Ramanan VK et al. 2012. Pathway analysis of genomic data: concepts, methods, and prospects for future development. *Trend Genet* **28**:323–332.

142 Li Y et al. 2010. Critical reasoning on causal inference in genome-wide linkage and association studies. *Trend Genet* **26**:493–498.

143 Craddock N et al. 2010. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* **464**:713–720.

144 Wray NR et al. 2011. Synthetic association created by rare variants do not explain most GWAS results. *PLoS Biol* **9**: e1000579.

145 Dickson SP et al. 2010. Rare variants create synthetic genome-wide association. *PLoS Biol* **8**:e1000294.

146 Orozco G et al. 2010 Synthetic associations in the context of genome-wide association scan signals. *Hum Mol Genet* **19**:137–144.

147 Saunders EJ et al. 2014. Fine-mapping the HOXB region detects common variants tagging a rare coding allele: evidence for synthetic association in prostrate cancer. *PLoS Genet* **10**:e10004129.

148 Cirulli ET, Goldstein DB. 2010. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* **11**:415–425.

149 Wang X et al. 2012. A statistical method for region-based meta-analysis of genome-wide association studies in genetically diverse populations. *Eur J Hum Genet* **20**:469–475.

150 Majewski J, Pastinen T. 2011. The study of eQTL variations by RNA-seq: from SNPs to phenotypes. *Trends Genet* **27**:72–79.

151 Fairfax BP et al. 2014. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**: doi: 10.1126/science.1246949

152 Lee MN et al. 2014. Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science* **343**:1246980.

153 Hass J et al. 2013. A genome-wide association study suggests novel loci associated with a schizophrenia-related brain-based phenotype. *PLoS One* **8**:e64872.

154 Chen F et al. 2011. Fine-mapping of breast cancer susceptibility loci characterizes genetic risk in African Americans. *Hum Mol Genetic* **22**:4491–4503.

155 Song C et al. 2013. A genome-wide scan for breast cancer risk haplotypes among African American women. *PLoS One* **8**:e57298.

156 Lucon DR. 2013. Downregulation of 14q32 microRNAs in primary human desmoplastic medulloblastoma. *Front Oncol* **3**:1–13.

157 Laddha SV et al. 2013. Genome-wide analysis reveals down regulation of miR-379/miR-656 cluster in human cancers. *Bio Direct* **8**:1–14.

158 Dominque BW et al. 2014. Genetic and educational assortative mating among US adults. *Proc Nat Acad Sci USA* **111**:7996–8000.

CHAPTER 6

Human origins and early diasporas

> Man could not stay there forever. He was bound to spread to new regions, partly because of his innate migratory tendency and partly because of Nature's stern urgency.
>
> —*Ellsworth Huntington* [1]

Summary

A number of extinct archaic human groups, such as *Homo erectus* and *Homo neanderthalensis*, are known to have populated Africa, Eurasia, and Oceania. Currently, only *Homo sapiens* or modern humans survive. Although several branches representing the genus *Homo* became extinct, it is not clear whether any of their DNA survives in the gene pool of contemporary humans as a result of introgression.

Environmental and biological forces contributed to the genesis of modern humans in sub-Saharan Northeast Africa about 250,000 years ago. These included dramatic geological changes such as the creation of the East African Ridge and Range that altered climatic conditions and habitat east of the cordillera. In addition, the Saharan pump provided for the climatic pulsations of wet and dry weather in the Sahara and Arabia. These fluctuations in turn created migrational gates that opened during wet expansion episodes and closed during dry contraction periods. Also, changes in mode of locomotion to bipedalism, in diet to include plants and animals, and greater brain size are hallmarks of the transition period between great apes and hominins. In addition, a number of revolutionary genetic changes in the form of mass duplications of certain core DNA elements seem to have played a role in early hominin evolution. The emergence of the first modern humans out of Africa into what is now the Near East about 150,000 years ago provided for possible cohabitation with Neanderthals in the area. For example, modern human and Neanderthal remains have been excavated from the Middle East in cave deposits dating back to 120,000 years ago. Noteworthy is that in spite of potential coexistence in time and space (within the same cave system), modern humans and Neanderthals differed in lifestyle such as seasonal (modern humans) and Neanderthals differed in lifestyle such as seasonal (modern humans). Recently, it is becoming increasingly apparent that although contemporary hominins are represented by a single species, us, in the past, our lineage has been complex with a number of species living in close quarters, likely interacting and possibly interbreeding. As a result, our ancestry does not look like a tree with discrete, separate branches but more like an inter-weaving river with small creeks separating from it only to connect again into the main stream.
Recently, it is becoming increasingly apparent that although contemporary hominins are represented by a single species, us, in the past, our lineage has been complex with a number of species living in close quarters, likely interacting and possibly interbreeding. As a result, our ancestry does not look like a tree with discrete, separate branches but more like an inter-weaving river with small creeks separating from it only to connect again into the main stream.

The *on switch* to humanity

Pennants of recent human evolution

A number of landmarks are usually recognized in the process of recent human evolution. Bipedalism was probably the first characteristic to evolve about 5 million years ago subsequent to the ape–hominin transition period. The main diagnostic trait that signals the beginning of the evolution to upright posture is the gradual switch in the position of the *foramen magnum* from a posterior position to a more anterior location at the base of the skull. At the base of the skull, the cranium rests

Genomes, Evolution, and Culture: Past, Present, and Future of Humankind, First Edition.
Rene J. Herrera, Ralph Garcia-Bertrand, and Francisco M. Salzano.
© 2016 John Wiley & Sons, Ltd. Published 2016 by John Wiley & Sons, Ltd.

Figure 6.1 Oldowan tools. (*Source*: Didier Descouens, https://commons.wikimedia.org/wiki/File%3APierre_taill%C3%A9e_Melka_Kunture_%C3%89thiopie_fond.jpg. Used under CC BY-SA 4.0, http://creativecommons.org/licenses/by-sa/4.0.) (See the Color Plates section.)

atop the vertebral column and allows for balance of the head during bipedal locomotion. Undoubtedly, this evolutionary process and the accompanying bipedalism were pivotal characteristics that facilitated dispersal of hominins within and out of Africa.

In parallel to the morphological change in the position of the *foramen magnum*, the Oldowan toolmaking technology emerged [2]. The stone utensils began to appear about 2.5 million years ago and australopithecines (see below for description) are credited for this development. The evolution of toolmaking continued with the hominins. Although other organisms including birds and primates (e.g., crows and chimpanzees, respectively) are known to make tools, the Oldowan implements are pebble tools that specifically follow a tradition (Figure 6.1).

Hominins also began to be less specific with their diet and gradually became more omnivorous and became increasingly more socially interdependent, creative, and aware of their own existence and of the universe. The development of speech probably had its genesis with early hominins such as *Homo erectus* (*ergaster*) about 2 million years ago. By the time *Homo neanderthalensis* appeared in the fossil record about 250,000 years ago, all anatomical requirements for speech were in place, including a horseshoe-shaped bone in the neck, the hyoid, which allows for the proper interactions between the tongue and larynx to articulate words. Otherwise, we would garble and hoot like chimpanzees. It is likely that language evolved alongside with changes in posture and locomotion occurring congruently. Alongside these changes, the human lineage experienced a remarkable increase in brain size. This increment in brain size was accomplished in a relatively short period of time (2.5 million years) and it led to intellectual capabilities unique and far above what is exhibited by other organisms. Intellectually, modern humans are in a class by themselves. The development of the hominin brain and the necessary coevolution of other organs including the integration to the anatomy and function of other tissues represent a daunting challenge for current evolutionary theory to explain [3].

The East African Ridge

Northeast Africa is considered by most experts as the birthplace of the human lineage [3], although other theories exist [4]. Northeast Africa is the geographic area where hominins separated from the forerunners to the great apes. Furthermore, the fossil evidence indicates that the majority of hominin species originated in East Africa. Curiously, East Africa is geologically being torn apart from the rest of the continent. The splitting process started about 100 million years ago when the African Plate began to move in a northeasterly direction. The driving force for this geological process is the upwelling magna pressing up into the East African crust above. This is taking place in East Africa in the area along Lake Victoria and the other great lakes. The separation of East Africa from the rest of the continent is occurring at a rate of about one inch per year. In time, this movement will generate a new tectonic plate (the Somali Plate) and a new continent will be born. It is stipulated that in approximately 10 million years the separation will be completed [5].

It turns out that these two processes, recent human evolution and the splitting of Northeast Africa into an independent continent, seem to be linked. It is not coincidental that several major events in hominin evolution have transpired in Northeast Africa. The geological episodes described in the previous paragraph triggered a series of environmental changes. In general, for the past 10 million years, Northeast Africa has shifted from a high-humidity forest to arid savanna.

Plate tectonics also created the East African Rift Valley. In turn, the East African Rift Valley has altered the

landscape to a fault graven basin that functions as a depression for narrow, long, deep lakes. In addition, the same pressure from the magna pushing against the crust is creating a mountain chain, the East African Ridge. These mountains run north to south. To the west, the habitat is forest and to the east it is becoming increasingly arid. This trend of an expanding dry environment has created the extensive savannahs that we see today. It has been postulated that these dramatic changes in vegetation and climate played an important role in driving hominin evolution [6]. It is likely that the environmental changes that resulted from these geological events provided for unique selection pressures and evolutionary changes in hominin evolution including bipedalism. In addition, it has been proposed that drastic climate variability east of the Northeast African Ridge and the resulting shifting selection pressures may have contributed to encephalization and migrations out of Africa [7]. This dramatic fluctuation in climate included extreme oscillations of dry and wet weather conditions. For example, the dramatic brain enlargement experienced by hominins about 2.0 million years ago seems to coincide with the creation of deep lakes along the Rift Valley. These rapid changes in habitat along with genetic changes (discussed below) may have provided the setting for punctuated evolution that could explain the revolutionary increase in brain size experienced within the human lineage [8].

Bipedalism

A number of speculations have been put forward to describe the events, forces, and selection pressures that trigger the path to humanity. Bipedalism is generally considered as the pivotal early trait that provided for a series of evolutionary processes that eventually led to modern humans [9]. Upright locomotion started early in hominin evolution. In fact, it preceded the dramatic increase in brain size in hominins (Figure 6.2) as well as the development of stone tool traditions. It is possible that the first facultative biped was *Sahelanthropus tchadensis*, a group that lived about the time that the hominin separated from the great ape lineage, 7 million years ago. Clear indications of bipedalism are evident in *Australopithecus* about 4 million years ago. The most recognizable characteristic that signals upright locomotion is the forward position of the *foramen magnum*, the opening that leads from the cranium to the spinal cord. In addition, other traits that address specific requirements of bipedalism include a bicondylar angle, a reduced or nonopposable big toe, a higher arch of the foot, a more posterior orientation of the anterior portion of the iliac blade, a relatively larger femoral head diameter, an increased femoral neck length, and slightly larger and anteroposteriorly elongated condyles of the femur [10].

Several theories have been proposed to explain the evolution of bipedalism in hominins. A number of them relate to environmental changes occurring at the time (see the previous section). For example, the savanna thesis proposes that a number of geological events including the creation of the East African Ridge and the geologically related East African Rift Range started a series of ecological changes that led to the transformation of the land east of the mountains from forest to grassland. These geological changes gave rise to the East African savannas. It is theorized that arboreal hominins were under selection pressure to evolve anatomical characteristics that would allow them to survive in the new dry plains of Northeast Africa. This transition from a tree-dwelling existence to bipedalism is seen in the anatomy of early hominins such as australopithecines that exhibit a mixture of arboreal and bipedal traits. Specifically, *Australopithecus* had curved fingers that allowed grasping of tree branches and yet they walked upright.

Basically, all of the models that have been proposed depend on the environmental and geological events that changed forest to savannas east of the Northeast African range. For example, it has been suggested that early hominins found it advantageous to adopt an upright

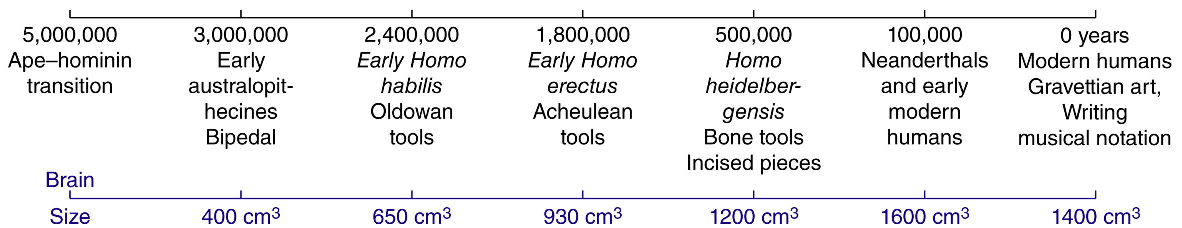


Figure 6.2 Relationship between bipedalism and brain size.

posture to warn and intimidate potential predators. This model is based on aposematism and includes the strategies of being visible and as vociferous as possible to scare attackers away. Along the same lines, it is possible that an erect and vertical posture would be selected for in a grassland environment as a mechanism of vigilance against predators or sexual display to attract mates. Bipedalism also provides for freedom to use the arms and hands for a number of activities that are associated with hominin evolution such as the manufacturing of tools and carrying food and offspring. In a savanna niche, anatomical traits that facilitated these behaviors may have been selected for. Even thermoregulation has been advanced as a selection pressure for upright posture since in an open field grassland, standing vertical reduces body surface area minimizing exposure to sunlight, increasing the distance above the ground favorable for cooler winds and helping in heat dissipation [11, 12].

The human brain

The human brain, with about 86 billion neurons, is a complex and remarkable organ. Its cerebral cortex, a furrowed outer layer of cells, is made up of approximately 10 billion nerve cells, each one interfaced by 100 trillion connections to other neurons, firing at about 10 billion times per second using some 100 neurotransmitters [13]. The cerebral cortex is involved in higher level functions including voluntary movement, integrating sensory information, learning, memory, and individuality. These neuronal activities somehow bring about not only self-awareness but also mindfulness and cosmic awareness. These interactions provide for abstract thinking, emotions, and empathy. Although new discoveries on the mental capacity of different organisms continue to astound us, it is likely that modern humans possess greater intellectual capacity than any other species on this planet.

Subsequent to the adoption of bipedalism in the human lineage, the human brain evolved very rapidly. Then, starting about 2 million years ago, the hominin brain experienced a dramatic rate of increase. And with the start of the Upper Paleolithic or the Stone Age (from 50,000 to 10,000 years ago), the size of the brain augmented even faster. The hominin brain actually tripled in size (from approximately 400 to 1400 cm^3) since *Homo erectus*, in about 2 million years. It is calculated that during this time period about 100,000 neurons and supporting cells increased per generation [13]. In addition to the increment in the number of neurons, the brain experienced a number of significant morphological changes including expansion and convolutions of the cerebral cortex and greater myelination of neurons. Myelination of neurons increases the speed of neural transmission.

A genetic spurt

Given the striking increase in brain size and intellectual capacity, the natural mechanisms that made them possible are still unexplained. In addition to the environmental switches and geological changes that took place in Northeast sub-Saharan Africa at the time of the ape–hominin separation, about 6–7 million years ago, that supposedly provided for selection forces conducive to bipedalism, a number of genetic changes occurred in the human lineage. Mysteriously, great ape and hominin genome evolution is characterized by the appearance and dispersal of core duplicons [14]. These core duplicons are pieces of DNA that have duplicated and dispersed randomly (not in tandem) *only* within the genome of great apes and hominins. About a dozen of these chromosome-specific families of duplicon units have been identified. They carry within them genes responsible for cell proliferation. They are expressed in many tissues, but especially in the brain, particularly in neurons. It is not known what triggered the initial spurt of events and kept their expansion going. It is not surprising to see duplicated pieces of DNA providing for rapid evolutionary change since multiple copies of genes allow for rapid accumulation of mutations due to low-intensity selection pressure. Yet, what is unusual about core duplicons in hominin evolution is their speed of dispersal and their association with genes specifically expressed in neuronal tissue. Current evolutionary theory does not provide an explanation for these observations.

Also a mystery is the mechanism of these dramatic episodes of chromosomal alterations. Core duplicons in hominins that dispersed throughout the genome may represent one of the mechanisms unique to recent human evolution. In addition, the open reading frames in these cassette-type duplications are disproportional (compared with other genes overall) under intense selection pressure and transcriptionally very active, suggesting functional importance. In other words, the genes within core duplicons may represent the proverbial "fast-evolving" human genes.

A specific example of these core duplicon-based amplifications is provided by the formin-binding protein 2 (FNBP2), which is a GTPase activating enzyme encoded

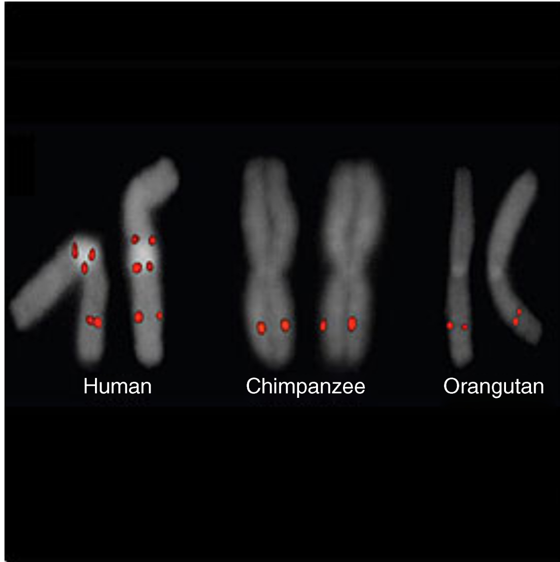


Figure 6.3 Duplication of the FNBP2 in hominin (chromosome 1). (Reproduced from Dennis et al. 2012 [15] with permission of Elsevier.) (See the Color Plates section.)

by a gene known as *SRGAP2*. This enzyme promotes motility and differentiation of neurons as well as synaptic connections. Remarkably, the *SRGAP2* gene has been duplicated several times to generate 23 paralogous loci during hominin evolution (Figure 6.3). Only one copy, the ancestral, is present in other mammals and primates. The first duplication took place about 3.4 million years ago, the second 2.4 million years ago, and the third about 1 million years ago. These duplication events gave rise to the novel *SRGAP2B*, *SRGAP2C*, and *SRGAP2D* loci, respectively. It turns out that *SRGAP2C* is a truncated version of the original gene (*SRGAP2*), inhibits its activity, and promotes neuronal movement. In addition, *SRGAP2C* slows down the aging of neurons and increases neuronal density, the number of neuronal contacts and synapses. It is noteworthy that the duplication that created *SRGAP2C* occurred approximately 2.4 million years ago at the time when cranial expansion dramatically started as seen in species such as *Homo habilis* (likely ancestor to modern humans) and continued in *Homo erectus*. Although these genetic alterations are congruent with increases in brain size, it is baffling why they happened (or at least were retained) so frequently and only in the hominin line. It is likely that these gene duplication events are more in line with punctuated evolutionary changes and not with neo-Darwinism.

Another dramatic core duplicon-type dispersal is illustrated by the DNA element known as DUF1220 [14] (Figure 6.4). DUF1220 is not a gene but a core DNA element reiterated within genes (5–50 copies per gene) that are part of a family of duplicons. The duplication rate of these elements has increased in hominins in comparison with other primates and great apes. For example, nonprimates possess less than 10 copies of DUF1220 while monkeys have about 30, great apes 90–125, and humans approximately 250 copies. This element has been linked to brain size [16]. The steady increment in the number of DUF1220s in primate evolution and their association with increased brain size suggest that DUF1220 has been involved in the evolution of the human brain. Furthermore, a direct proportional relationship between number of DUF1220 copies and amount of gray matter in the cerebral cortex of healthy humans has been uncovered [16].

Another area of interest regarding DUF1220 in humans is its correlation with a number of behavioral and anatomical disorders such as autism, schizophrenia, microcephaly, and macrocephaly [17]. For example, in chromosomal location 1q21, known for its susceptibility to aberrations, high numbers of DUF1220 have been detected. In this locus, microcephaly is linked to the number of DUF1220 repeats. Therefore, it seems that at least some of these core duplicons correlate not only with a revolutionary increase in brain size, but with genetic instability as well. Thus, these fast-expanding elements may represent an evolutionary double-edged sword. On one hand, they may promote rapid evolutionary change, potentially advantageous for the survival of organisms, and on the other hand, the same fast-paced mechanism may go out of control facilitating chromosomal aberrations and pathological conditions related to neuronal tissue.

Early hominins

In all discussions of evolutionary change, it is important to keep in mind that there are no distinct lines of demarcation separating organisms along lineages. Scientists give names to individual fossils from the past to be able to refer to and talk about them in a practical way. We have created artificial lines of separation among organisms and given them names to simplify a very complex evolutionary process. Extinct organisms are part of a

```
EKGKESMAHR LFM---EQQE AEEVNETEED ---TRDEHYL THFHCHDLPD AHQPPSS-TT LMSD---EQE VCSSLGAA

ERNK-PVVLT LFM---EQQE VEGMNETEED ---KLDEDNL TPFHCLDFPD AHQPPSS-TT FMSD---EQE VYSPLSAA

EKEQESLALS LST---EGIN EEKVAEILPD ---SLGEGPE TSSHHYDLTD IHEPFCT-AT ALSD---AHE VRSSLDAA

VEGEEELLLA AS----QKEE AETVKKVLQD ---SVEERYL TPLSLHDLPD AHRLHRS-SA FLFD---DYK FDLAVDGA

EKVQELYAPR -------EVQ KAEEKEVPED ---SLEECAI TCSNSHHPCE SNQPYGN-TR ITFE---EDQ VDSTL-ID

EKVQESCAPR -------EVQ KAEEREIPED ---SLEACAV PCSNSHSPCE SNKPHRN-TK ITFE---EDK VDSTV-ID

DKVQESPAPR -------EAQ KAEEKEVPED ---SLEECVI SCSNSHGPCN SNPPHRN-TK ITFE---EDK ATSTLIVD

EKVQKSSAPR -------EVQ KAEESKVPED ---SLEECAI TCSNSHGPCD SNQPHKN-IK ITFE---EDE VNSTLVVD

EKVQKSSPPR -------EMQ KAEEKEVPED ---SLEECAI TCSNSHGPYD SNQPHRK-TK FTFE---EDK VDSTL-IG

EHLVHKFIP- -------EVQ KAEEKEVPED ---SLEGCAV TCSNSHSPYD SNQPHRN-TK ITFE---EDK FDSGLVVD
```

Key to symbols:

IUPAC nucleotide code	Base
A	Adenine
C	Cytosine
G	Guanine
T (or U)	Thymine (or uracil)
R	A or G
Y	C or T
S	G or C
W	A or T
K	G or T
M	A or C
B	C or G or T
D	A or G or T
H	A or C or T
V	A or C or G
N	Any base
. or -	Gap

Figure 6.4 Alignment of 10 representative DUF1220 core elements. (Adapted from http://smart.embl-heidelberg.de/smart/do_annotation.pl?DOMAIN=DUF1220.) (See the Color Plates section.)

continuum of groups of individuals in the process of evolution, and oftentimes it is not clear whether related specimens are members of the same species or not. Discovered fossils represent only sporadic snapshots of evolutionary lineages. Also, recovered fossils represent only a very fragmentary and limited part of the story. Therefore, the fossil record should not be interpreted as discrete cubicles, stepping stones, or significant landmarks. Fossils represent the part of the evolutionary history that we were lucky to find.

Within the human lineage, the most recent branch includes the hominins. The term hominin refers to organisms, extant and extinct, that include humans and fossil relatives of humans evolutionarily closer to each other than to great apes (chimpanzees and gorillas). Specifically, hominins include modern humans, extinct human species such as *Homo neanderthalensis* and *Homo erectus*, and other genera within the human lineage that originated after the Hominini/Panini rift, including *Australopithecus*, *Paranthropus*, and *Ardipithecus*. Within the evolutionary tree, the Hominini and the Panini tribes split about 6–7 million years ago. The subtribe Hominina, within the hominins, is exclusively made up of species of the genus *Homo*. The genus *Sahelanthropus* dates back to 7 million years ago, close to the time of the Hominini/Panini separation. It may represent a common ancestor to both humans and the chimpanzee, or alternatively, it may represent one of the earliest hominins. Based on the anteriorly located *foramen magnum* at the base of the skull, pointing downward, *Sahelanthropus tchadensis* may have been an occasionally bipedal organism, although dental and facial characteristics point otherwise [18].

Emerging themes and variations in hominin evolution

A general theme within the hominin lineage is the reoccurring finding of individual fossils representing a mixture of ancestral (plesiomorphic) and derived (apomorphic) characters. It is as if differing parts of the hominin anatomy and physiology have evolved at different times in response to various selection pressures. In other words, certain characters of a given specimen exhibit derived traits while other retain the ancestral condition. This phenomenon is clearly seen in extinct species of the genera *Australopithecus* and *Homo*. Illustrations are seen in specimen of *Homo erectus* and *Homo naledi* (see descriptions below). In *Homo naledi,* we see an amalgamation of plesiomorphic and apomorphic metrics including ancestral hip and upper body proportions such as in *Australopithecus* but derived bone taphonomy in legs, feet, ankles and thumb reflecting the genus *Homo*. It is as if some ancestral morphological elements were retained in early hominins as they continued evolving into upright posture, bipedalism and increased brain size. A number of explanations have been posited to explain the observations. For example, it has been proposed that ancestral arboreal traits of early hominins are retentions from tree dwelling ancestors that were either in the process of being lost or were non-functional and selectively neutral, and keep as vestiges, much the same way that the appendix is retained in contemporary humans. Alternatively, the primitive features may have been keep under stabilizing selection forces or even beneficial as an alternative mode of locomotion in transitional evolutionary periods.

Another reoccurring motif that is becoming increasingly evident in hominin evolution is its increasing degree of complexity. The picture of recent human evolution has changed fast and dramatically during the past decade and it is likely that it will continue to do so. As new fossils and genetic discoveries are made, it becomes progressively clear that recent human evolution cannot be drawn so much like a tree with discrete, separate branches but more like an inter-weaving river with small creeks separating only to connect again into the main stream. The increasing number of discoveries of hominin groups living side by side and the medley of ancestral and derived features found together in specimens as well as the various reports of introgression involving extinct and extant human groups, are beginning to pain a complex picture of phylogenetic relationships that resemble a braided river. Specific examples of this emerging general theme is seen in the ancestral and derived anatomical features shared by contemporaneous species of *Homo* (see discussion below on *Homo naledi)* and the indications of interbreeding among contemporary humans and our immediate cousins, Denisovans and Neanderthals, based on shared genetic markers. Furthermore, all of these new discoveries suggest a number of polyphylethic interactions among species in different parts of Africa and Eurasia, from the base of the Hominin lineage to more recent times. It seems that the days of a simple linear evolutionary progression of hominins leading from *Australopithecus* to contemporary humans was rather naiveté.

Australopithecines

Within the hominin lineage is the genus *Australopithecus. Australopithecus* likely originated in East Africa about 4 million years ago and became extinct approximately 2 million years later. During its existence, australopiths spread throughout Africa but failed to migrate outside the continent. A number of distinct groups exhibiting diverse characteristics have been described within this genus, including *A. afarensis*, *A. africanus*, *A. anamensis*, *A. bahrelghazali*, *A. garhi*, and *A. sediba*. As usual, there is controversy on the classification of the various fossils. There are arguments based on whether these groups merit the species category, while some experts believe that the two general types, the robust australopiths and the gracile australopiths, represent different genera. In fact, the robust australopiths, which take their name from their bulky craniodental features, are thought by some investigators to have descended from the gracile type about 2.7 million years ago and that they should be referred to by a different genus, *Paranthropus* [19].

One of the early forms of australopithecines was *A. afarensis*. This group is represented by fossils from about 300 individuals. They existed for approximately 1 million years, 3.85 to 2.95 years ago, in East Africa in what is today Ethiopia, Kenya, and Tanzania. These hominins were capable of bipedal locomotion although their long arms and curved fingers suggest some degree of arborealism. Their diet is thought to have been based on plants and fruits. From this genus, the oldest bipedal footprints have been recovered. Their brain was about 500 cm^3 or one-third the size of an average human. Physically, they possessed a mixture of traits that are associated with apes and humans. For example, they had a flat nose and a robust protruding lower jaw and small canines. There is no data indicating that *A. afarensis* ventured out of Africa.

A more recent form of australopithecine, the gracile *A. africanus*, lived in South Africa from 3 to 2 million years ago. *A. africanus* shared a number of traits with *A. afarensis* including long and robust arms but unlike *A. afarensis, A. africanus* possessed a larger brain ranging from 400 to 500 cm^3. Like *A. afarensis*, *A. africanus* had the foramen magnum at the base of the brain case over the spinal column indicating that the organism was capable of bipedal locomotion. In general, it exhibited a cranium and facial features more similar to modern humans. It is possible that *A. africanus* was a common ancestor to modern humans.

Homo naledi

On October 2013, a collection of about 1550 hominin specimens belonging to about 15 to 18 individuals was discovered by two nibble cavers in a well known cave system in the vicinity of Johannesburg, South Africa. Specifically, the remains were found in an extremely difficult to reach chamber known as Dinaledi in the Star cave system. It is argue that the slender metrics of the two cavers were paramount in the discovery since most humans would not have been able to squeeze through a passage 8 inches wide to reach the diminutive cavity were bones were found. This finding undoubtedly represents one of the most significant revelations of the 20th century to the present in the field of anthropology and human evolution. Most of the remains were unarticulated and except for the bones of an owl, all the fossils were hominins. The first report describing this discovery was published about two years later, a premeditated delay to allow for reliable dating which is still pending.

Although the *Homo naledi* remains have not been dated yet, the orthodoxy postulates that this early homimin group lived 2.5 to 2.8 mya and phylogenetically belongs at the base of the *Homo* lineage. It is thought to be closer to *Homo erectus* than to *Australopithecus*. Males *Homo naledi* (naledi means star in Sesotho language) were about 5 ft tall weighting approximately 100 pounds while females were somewhat smaller and lighter, both parameters within the range of small modern humans. Cranial volume ranged between 560 cm^3 and 465 cm^3, similar to *Australopithecus*. Yet, surprisingly, the skull shape resembles early *Homo sapiens*. The skeleton remains suggest upright posture and bipedal locomotion. The teeth and mandible musculature were small reflecting a diet not requiring heavy mastication. The hands were an interesting collage of ancestral and derived characteristics. It included long curved fingers for arboreal living, an ancestral condition, and long robust thumb for manipulation of objects, a derived trait. It is remarkable that morphometric analyses indicate that *Homo naledi* wrist falls within the parameters of modern humans and Neanderthals and away from great apes, *Australopithecus* and *Homo floresiensis*. Further, metrics of skull and teeth group *naledi* with *erectus*, Neanderthal and contemporary humans, closer to *Homo erectus* than to *Homo habilis*. This sets *Homo naledi* apart from other hominins in an evolutionary position proximal to more recent hominins and to *Homo habilis*, a more ancient group.

The above mentioned characteristics in *Homo naledi* are an interesting potpourri of ancestral and derived traits. It is likely that *Homo naledi*, as many other extinct hominin species, represents a transition stage within recent human evolution; specifically, an intermediate phase between arboreal and terrestrial existence. It seems that *Homo naledi* retained ancestral anatomy and function in the upper limbs while refining the different components of bipedal terrestrial locomotion.

An interesting issue in connection with *Homo naledi* as it relates to the nature of the finding is whether the group practiced ceremonial interment or at least intentional burial. The current inaccessible condition of the Dinaledi chamber and the lack of illumination argue against intentional placement of bodies by members of the same species. Also, no stone tools were found at the site. Personal artifacts usually accompany the dead as part of ceremonial assembles. And yet, the large number of individuals and absence of remains of fauna may be indicative of bodies deposited in the cave after death. The expectations of activity associated with predators dragging bodies into a cave are a diversity of remains from different species typically prayed by carnivores, for example. Currently, the cavity is difficult to reach; yet it is possible that at the time of this species' existence, spaces were not so restricted, especially for individuals like *Homo naledi* with a smaller size and body frame. In other words, geological activity could have made the chamber less penetrable with time. Along this line of thought is the fact that most of the bones were found disarticulated. We can only speculate and envision inanimate bodies being carry and pull through narrow spaces and in the process intentionally or unintentionally becoming disjointed.

Homo erectus

The time range of *Homo erectus* started about 2 million years ago and extended to around 150,000 years from the present. In the human branch of the evolutionary tree, *Homo erectus* possibly descended from *Homo habilis* or *Homo naledi* and led to *Homo heidelbergensis* and ultimately to modern humans. *Homo erectus* possessed more human proportions compared with *Australopithecus*, including longer legs and smaller arms. The cranial capacity of early *Homo erectus* specimens was about 900 cm^3 while late samples averaged 1100 cm^3. It originated in Northeast Africa where it coexisted with *Homo habilis* and other hominins at the beginning of its time range. Unlike the present, in which only one hominin species exists, the fossil record indicates that about 2 million years ago a number of early human groups coexisted and lived in close proximity. At the end of its time period, *Homo erectus* lived side by side with modern humans. There are indications that *Homo erectus* cared for the old and the sick [20].

It is interesting that a possible form of *Homo erectus*, *Homo floresiensis*, was alive as recent as 12,000 years ago in the island of Flores in Indonesia. It was a small organism, only about 3.5 feet with a brain size of 380 cm^3. Since modern humans reached Flores about 45,000 years ago, *Homo floresiensis* and *Homo sapiens* could have lived together on the same island for at least 33,000 years. Its existence is so contemporary that some investigators contest that the Ebu Gogo local traditions and myths of a forest humanoid creature derive from the existence of this group. Yet, the taxonomic status of this fossil has been subject of considerable controversy and recent reports argue that the species is invalid [21,22].

Homo neanderthalensis

Neanderthals are considered by most biologists as a separate species and not as a subspecies or race of *Homo sapiens*. It appeared in Europe about 250,000 years ago. Neanderthals originated in Europe and became extinct in the same continent about 24,000 years ago. Most authorities believe that modern humans and Neanderthals had a common ancestor, *Homo heidelbergensis*, in Africa approximately 350,000 years ago. The African branch of *Homo heidelbergensis* gave rise to contemporary humans, while the European lineage evolved into *Homo neanderthalensis*. The last specimens of Neanderthals lived in what is today southern Iberia, suggesting Neanderthals' migrations occurred within Eurasia, and they were intracontinental.

The average brain size of Neanderthals was about 1600 cm^3 compared with an average of 1400 cm^3 for modern humans. Genetic data indicate that *Homo sapiens* and *Homo neanderthalensis* differ by only about 0.1% of their DNA. Over the last 15 years, various estimates of admixture between these two groups have been reported ranging from 0 to 4%. The data suggesting introgression have been repeatedly contested on the grounds of contamination with contemporary human DNA, faulty comparison algorithms, and/or the possibility that at least some of the DNA in common derive from ancient common polymorphisms (mutations and polymorphisms present in the common ancestor to Neanderthals and modern humans) [23].

Since modern humans undertook at least two incursions into Eurasia, the first one about 125,000 years ago reached the Near East and the second one resulted in a European settlement about 45,000 years ago, humans and Neanderthals cohabitated twice in parts of their ranges. These dates indicate that at least in Europe, these two groups may have coexisted for about 20,000 years. In the Near East, it is possible that they coexisted for a longer period of time. The nature of the interactions between them is not clear, yet considering the options of outright conflict or collaboration, as well as the different degrees of each, it is difficult to imagine living side by side for that amount of time in constant belligerence. However, indications of sharing cave living quarters have been documented in present-day Israel [24].

Several hypotheses have been put forward to explain the extinction of Neanderthals. Authors such as Jared Diamond [25] have proposed a scenario of animosity and outright aggression with *Homo sapiens*, leading to their dwindling and eventual disappearance in their last bastion in southern Spain. Another theory suggests that rapid weather changes with dramatic fluctuations in temperature created ecological conditions unsuitable to Neanderthals. And still a third theory points to admixture as the culprit. In this case, interbreeding with modern humans brought about their extinction as hybrids gradually replaced pure Neanderthals and they were absorbed into the gene pool of an overwhelming larger modern human population. Of course, any combination of these possibilities may have occurred.

Denisovans

In 2008 a pinky bone was discovered in a cave in the Altaic region of south central Siberia, a site notorious for Neanderthal remains. This pinky belonged to a juvenile female. Radiocarbon dating of nearby osseous matter gave an age of 41,000 years from the present. Archaic humans have inhabited the cave as well. Thus, initially the bone was thought to be of a Neanderthal. Subsequently, two teeth were recovered from the same cave. These three items are all that we have of Denisovans. Therefore, little is known of its anatomy and as such it was tentatively classified as *Homo sapiens ssp. Denisova*. In fact, most of what is known of Denisova is just based on DNA analysis.

It is known from genetic evidence that Denisovans split from the modern human lineage about 800,000 ya and from Neanderthals around 600,000 ya, and it is closer to Neanderthal than to modern humans. So modern humans, Neanderthals and Denisovans share a common ancestor. Furthermore, it turns out that Denisovans represent a case in which mtDNA and genomic sequences tell different stories, different timelines. The mtDNA sequences were generated first and indicated that Denisovans diverged before modern humans and Neanderthal did while whole genome analyses suggest that Denisovans separated from Neanderthals more recently after the separation from the modern human lineage. Thus, genomic sequences indicate that Denisovans and Neanderthal share a common ancestor with each other but not with modern humans. Since the initial discoveries, additional mtDNA and genomic sequences have been generated validating the original findings. In other words, mtDNA points to a more ancient lineage.

It seems that Denisovans had a very extensive range encompassing a region from Southeast Asia and Near Oceania to Western Europe. On the western fringe of its distribution, a recent discovery from Atapuerca in Northeastern Spain demonstrates that the Denisovan lineage in Europe is ancient. In a cave complex known as Sima de los Huesos, archeologists found remains of what they thought belonged to a Neanderthal. Considering its anatomical characteristics and the abundance of Neanderthal fossils in the cave, this was a logical classification at the time. It turns out that investigators were quite surprised to find that the fossil (femur) was approximately 400,000 years old and it resemble the Denisovan type, not Neanderthal. Before this discovery, the paradigm was that Neanderthals were found west of the Urals and Denisovans east of the range. Based on this data, the most parsimonious explanation is that the Spanish Denisovan lineage represents an ancient split from the common ancestor of both Neanderthals and modern humans, dating back to approximately one million years ago. Surprisingly, the other end of the Denisova range lies in Southeast Asia and Near Oceania. This conclusion derives from the 3–5% Denisovan DNA detected in some populations from Melanesia, New Guinea and Australia, evidence for not only introgression but a Denisovan presence in the region. With this extensive geographical expanse, it was not unexpected to find that Denisovan exhibit considerable mtDNA diversity. As previously stated, this wide distribution of Denisovan, Neanderthal and modern humans and introgression are indicative of complex interactions among hominins.

As expected, this finding generated a number of questions including how Denisovan got to Spain 400,000 years ago and how it could be related to Siberian fossils

dating to as recent as 41,000 years ago? In addition to the profound and revolutionary nature of these findings, is the technological triumph of sequencing DNA close to half a million years old. Prior to this achievement, the oldest samples accessible to sequencing were 10 times younger. This technological advance in molecular biology technology undoubtedly will open the doors to additional genetic analysis of increasingly ancient groups increasing our understanding of our origins.

The first hominin migrants

Homo erectus

Modern humans were not the first migrants out of Africa. In fact, a number of groups are known to have preceded *Homo sapiens* in reaching Eurasia. Among these early hominin travelers, the fossil record indicates that individuals classified as *Homo erectus*, *Homo antecessor*, and *Homo heidelbergensis* left Africa by the Levant and/or the Horn of Africa about 1.8, 0.8, and 0.6 million years ago, respectively [26].

Some investigators classify some *Homo erectus* fossils in Africa as a separate species, *Homo ergaster*, and consider *Homo ergaster* as the first migrants out of Africa. The African *Homo erectus* (*ergaster*) differs from the Asiatic *Homo erectus* in the shape of the brow ridges and its smaller brain case. It is thought that the arid conditions of Northeast Africa and the Near East presented environmental challenges to these early migrants. It is likely that the harsh environmental conditions and the limited supply of food and fresh water demanded certain physical and mental characteristics to allow survival during the treacherous treks across the two possible routes to the Near East, across the Strait of Sorrows at the Horn of Africa or the Levant.

Probably the earliest form of *Homo erectus* found outside Africa was unearthed in the vicinity and within the town of Dmanisi in the Republic of Georgia, Southwest Asia. The remains of a total of five individuals dating to the same time strata have been recovered since 2000. The latest discovery of a remarkably well-preserved skull was reported in October 2013 [27]. If the dating of these bones is correct and they belong to an early *Homo erectus*, this discovery pushes back the migration of *Homo erectus* some time prior to 1.8 million years ago. Dmanisi's samples possessed a number of archaic features such as protruding brows and salient jaws. The dispersal of *Homo erectus* to Asia and Europe, as attested by the Dmanisi fossils, was rather fast. Representatives of *Homo erectus* reached Java approximately 1.7 million years ago and Atapuerca in Iberia, Spain about 1.2 million years ago [28].

The brain size of the Dmanisi specimens was only about 500 cm^3 or one-third the volume of a contemporary human brain. This cranial capacity is within the range of chimpanzees. *Homo erectus* walked upright. And although it possessed a small brain size it did not compromise its ability to migrate out of Africa and at least reach Southwest Asia. It is significant that this group was able to populate the Caucasus in spite of its brutal winter weather conditions.

In addition to their unexpected old age, the Dmanisi fossils exhibit considerable anatomical heterogeneity. Considering that these remains likely belong to individuals coexisting in time and space and possibly members of a single interbreeding population, *Homo erectus*, as a group, was anatomically quite diverse. Furthermore, it brings to question how much of the variability seen in other hominin fossils justifies separate species classification as opposed to genetic diversity within the gene pool or race differences. Traditionally, the field of physical anthropology has been criticized by some for its lenient tendency to create novel taxa, sometimes based on minimal evidence.

Homo antecessor and Homo heidelbergensis

Homo antecessor with a cranial size of 1000–1150 cm^3 is speculated to be descendent from *Homo erectus* (*Homo ergaster*) in Africa. Fossils of *Homo antecessor* are restricted to Western Europe, specifically Spain, France, and England. Some experts believe *Homo antecessor* to be the ancestral group from which *Homo heidelbergensis* derived, although others suggest that it is simply an early form of *Homo heidelbergensis*. *Homo heidelbergensis* inhabited Africa, Europe, and Asia from about 600,000 to 200,000 years ago. *Homo heidelbergensis* is distinct from *Homo erectus* primarily due to its large brain size and body proportions, which are comparable to contemporary humans. It is theorized that after they left Africa about 300,000 years ago, *Homo heidelbergensis* evolved into *Homo neanderthalensis* in Europe and Denisovans in Asia. It is though that in Africa, they gave rise to *Homo sapiens* around 200,000 years ago [29].

The emergence of modern humans

Most experts are of the opinion that modern humans evolved from an archaic form of *Homo sapiens*, possibly an organism similar to *Homo heidelbergensis* in what is today Northeast Africa. Currently, in the absence of genetic data, the definition of modern humans from Africa is purely anatomical. The oldest fossils of modern humans dating to about 195,000 years ago were discovered between 1967 and 1974 near the Omo River in what is today Ethiopia. They consist of two partially preserved skulls (Omo I and Omo II) and various other bones (e.g., a femur as well as portions of a pelvis and foot). The cranial capacity of the best preserved of the two skulls is about 1400 cm^3. Curiously, the two sets of remains exhibit marked differences. While Omo I looks quite contemporary, Omo II seems more primitive. It is not clear whether the observed differences reflect the coexistence of two subspecies or intraspecies variation. Unfortunately, no DNA data exist from these fossils.

A more recent form of early modern humans from Ethiopia was discovered in 1997 near the village of Herto Bouri in Ethiopia. It is noteworthy that all of the early modern humans in Africa have been found in Northeast Africa in what is today Ethiopia. This has led some experts to suggest that this region may be the cradle of our species [4,30]. The Herto findings include three crania in good condition, the largest of which exhibits a capacity of 1450 cm^3. For comparison, the Herto brain size is about the average for Neanderthal and larger than most living Caucasians (1441 cm^3) and Africans (1338 cm^3), although it is smaller than the average for Asians (1491 cm^3). Radioisotope measurements of the surrounding volcanic layer date between 154,000 and 160,000 years ago. Most experts consider these fossils as an extinct subspecies of modern humans, *Homo sapiens idaltu*. The Herto specimens differ from modern humans in Europe such as Cro-Magnons, in that they exhibit a number of archaic characteristics. These archaic features include large eye sockets, wide cheekbones, prominent brow ridges, sloping forehead, large teeth, a severe post-orbital constriction, and, in general, a more robust skull. No DNA has been extracted from these fossils.

The Saharan pump

The region that is known today as the Sahara and the Arabian Peninsula has experienced pulses of extreme wet and dry conditions. These episodes of fluctuation of available water started about 3 million years ago. Since then, the Saharan region oscillates in this manner transforming forest habitat to barren desert and back to forest approximately every 20,000 years. This transformation happens dramatically within the span of a few centuries. It has been postulated that these fluctuations result from the Earth's wobble, which also takes about 20,000 years. Earth makes a full rotation around its axis during this period of time. It seems that this wobble effect weakens the transport of warm upper waters to the north and cold deep water to the south of the planet promoting arid conditions in the Sahara and Arabian Peninsula area. We are now around 7000 years into a dry cycle. Thus, in about 13,000 years the Sahara Desert should be called the Saharan Jungle with a corresponding biota.

It is thought that these pulsations of extreme wet and dry conditions provided for expansion and contraction episodes of animals and plants in the Saharan and Arabian Peninsula regions. Since a number of the pivotal events to hominin evolution took place in this area of the world, it is likely that these cyclical changes impacted the dispersal of humans out and back to Africa by providing windows of opportunity for dispersals. In fact, it turns out that the recurrent wet periods coincide with known migratory events of various hominin forms including *Homo erectus*, *Homo heidelbergensis*, and early modern humans, in the case of *Homo erectus* and modern humans on more than one occasion. It has been suggested that these wet–dry cycles have been particularly determinant in recent human evolution during the last 200,000 years [31]. Data support the existence of benign conditions with vegetation only at succinct and short time intervals.

The first emigrational event of modern humans into the Arabian Peninsula is thought to have occurred about 130,000 to 100,000 years ago [32] and coincided with a wet period from 120,000 to 110,000 years ago. There is even evidence for ancient rivers flowing through the Libyan and Chad Basins as well as a fresh water passageway bisecting the Sahara [32]. This is also concurrent with a major spread of vegetation in the Sahara and Arabia from 120,000 to 110,000 years ago. This movement of modern humans represents the first out of Africa episode. These migrants were likely the direct descendants of forms similar to Omo and/or Herto described above. The same region was extremely dry prior to 140,000 and after 110,000 years ago.

A second period of movement from Africa to Asia and back to Africa also corresponds to an extreme epoch of wetness. About 50,000 to 45,000 years ago, the Sahara and Arabia were again vegetated overlapping with a period of hominin migration (60,000 to 40,000 years ago) dated by mitochondrial DNA (mtDNA) and Y-specific markers. Uniparental genetic markers also suggest a back to Africa migration within the same wet period [33]. Conversely, an older stretch of time from 75,000 to 45,000 years ago signals drier cooler weather in the Sahara and the Levant that coincides with human extinction events [34]. Two additional dispersals of modern humans out Africa are known, both within the last wet period (25,000 to 4,000 years ago). The oldest, about 18,000 years ago, exported the Epipaleolithic Kebaran culture, the proto-Nostratic language (the ancestor of the Indo-European, Uralic, and Altaic languages of today), the dog, and the bow/arrow to Eurasia. The most recent and final incursion out of Africa took place approximately 8000 years ago and brought about the dissemination of the Afro-Asiatic languages including the Semitic to the Near East, as well as the Berber and Egyptian to North Africa [35].

Early migrations

Primarily two routes are thought to have been the gateway in these dispersals, the Levant and the Horn of Africa across the Strait of Sorrows (the Bab-el-Mandeb Strait). The Strait of Gibraltar that connects Northwest Africa and Iberia has been considered as a third potential crossing; however, except for isolated publications [36], limited evidence exists of its use particularly during pre-Neolithic times.

About 125,000 years ago, during the Middle Paleolithic or Middle Stone Age, modern humans migrated from Northeast Africa to the Near East. Other estimates provide a window for the migration(s) of 150,000 to 130,000 years ago [37]. The Jebel Faya site in United Arab Emirates, Southeast Arabia, for example, has been dated to 127,000 years ago (early Middle Paleolithic) by oxygen isotope data (Figure 6.5). Unfortunately, no fossils have been recovered from the early Jebel Faya location, only stone tools. Interesting, of the three assemblages of stone artifacts found at Jebel Faya, two exhibit striking similarities to contemporaneous implements from Northeast Africa. The third group of tools, on the other hand, was very different. These findings indicate that different stone making traditions existed in close proximity. All together, these data may be indicative of a direct corridor connecting the Horn of Africa and Arabia during the low sea levels of the glacial phase from 180,000 to 125,000 years ago and possibly reflect some degree of interaction between the two regions. It is also possible that the two different styles represent different human groups. Yet, it is not clear whether the Jebel Faya settlement represented long-term occupations, foraging, or transient stays. However, the discovery of modern human remains in Southeast China at the Zhiren Cave, including a mandible dating back to more than 100,000 years ago [38], suggests that these early incursions from Northeast Africa into the Near East may not have died in Arabia but were fertile in the peopling the rest of the world reaching into the Far East. Yet, no genetic data exist indicating a contribution of these early humans to our contemporary gene pool. In addition to Jebel Faya, other locations in the Near East such as Tabun, Ayu al-Buhaira, and Tor Faraj, among others, have been discovered dating back to the Middle Paleolithic (150,000 to 45,000 years ago). Red Sea shore artifacts from Eritrea, on the East African side of the crossing, dating to 125,000 years ago (U–Th mass spectrometry) indicate that early humans occupied these coastal areas and practiced marine food subsistence, suggesting migration across the southern route during the earliest migrations. In recent years, it has become evident that a number of dispersals may have ventured into Arabia during the first out of Africa period.

Neanderthals prevailed

Possible coexistence of modern humans and Neanderthals in the Levant

A series of remarkable discoveries of early modern human remains were made in the Mount Carmel Range in what is today northern Israel. The most notable of these are the modern human fossils in the Skhul (1939) and Qafzeh (1934) sites by teams headed by Arthur Keith and Theodore McCown, and René Neuville, respectively (Figure 6.5). These modern human bones are from the early Middle Paleolithic with dates ranging between 100,000 and 120,000 years ago. Specifically, these digs are located in the western slopes of Mount Carmel near the lower Sea of Galilee and they include other caves such as the Tabun, Jamal, and el-Wad Caves, among

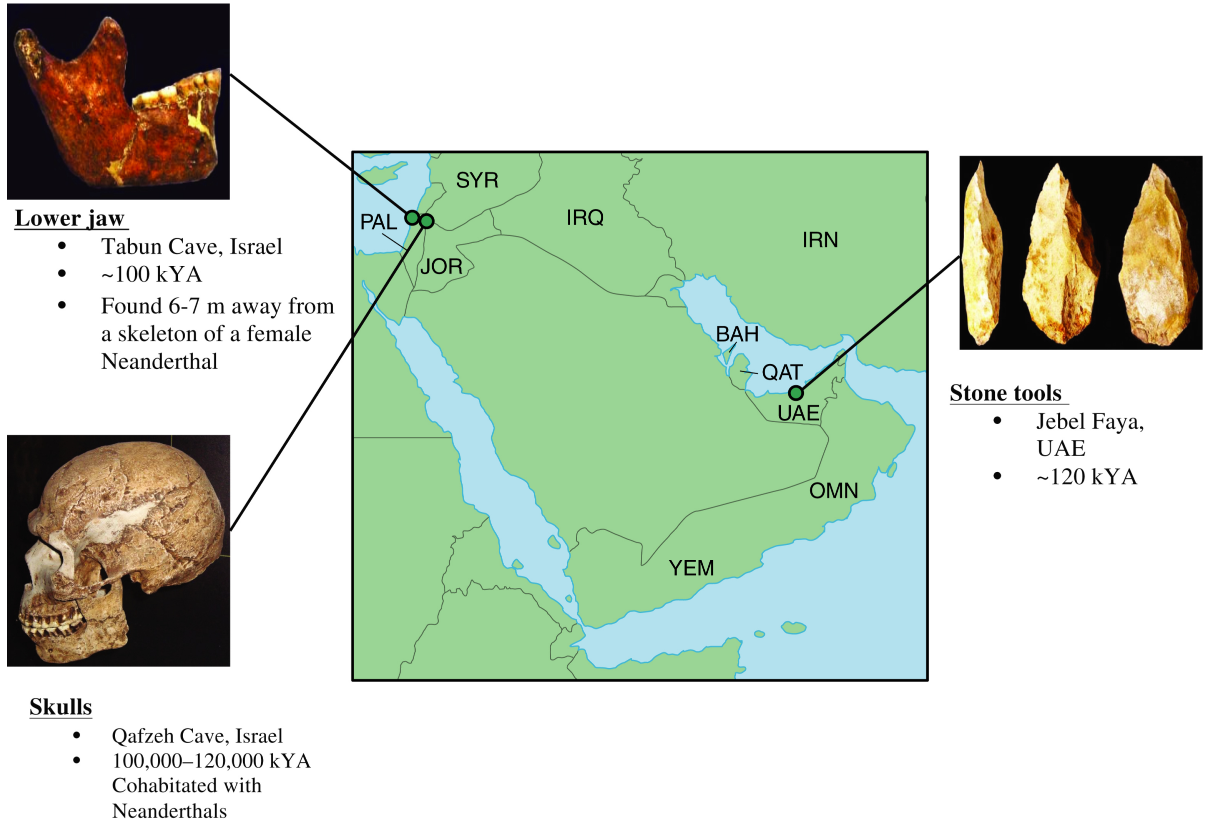


Figure 6.5 Stone tools of Jebel Faya and bone remains from Israel, the latter suggesting potential cohabitation of modern humans and Neanderthals. (*Lower jaw*: Reproduced with permission of www.SkullsUnlimited.com. *Skulls*: Wapondaponda, https://commons.wikimedia.org/wiki/File%3ASkhul.JPG. Used under CC BY-SA 3.0, http://creativecommons.org/licenses/by-sa/3.0, or GFDL, http://www.gnu.org/copyleft/fdl.html. *Stone tools*: Armitage et al. 2011 [39]. Photograph courtesy of AAAS/Science.) (See the Color Plates section.)

others. It is thought that these caves contain the earliest modern human remains outside of Africa.

Of particular interest is that the Skhul and Qafzeh modern human samples belong to the same strata as Neanderthal bones in the adjacent Tabun Cave. Electron spin resonance and thermoluminescence dating techniques indicated that the Tabun Neanderthal samples were contemporaneous with the modern human remains found in Skhul and Qafzeh, both from about 120,000 years ago during the early Middle Paleolithic or the Mousterian. This proximity in time and space of these modern humans and Neanderthal fossils suggests a possible coexistence of these two groups. Since the modern humans of Skhul and Qafzeh disappeared about 80,000 years ago, and Neanderthal sites in the area, such as Kebara Cave, persisted to 61,000 to 48,000 years ago, the latter prevailed longer in the region. It is also possible that Neanderthals inhabited the Near East during two separate occasions, the first from 120,000 to 100,000 years ago and the second from 61,000 to 48,000 years ago.

The modern human specimens from Skhul and Qafzeh are similar to the Omo and Herto fossils from Ethiopia (see above). The modern human remains in the Israeli caves exhibit a mixture of archaic and modern characteristics as the Omo and Herto. Although the Arabian crania possess brow ridges and projecting face reminiscent of Neanderthals and the Omo and Herto specimens, their features in general are less robust. It has been postulated that the Skhul and Qafzeh hominins signal the first migration of modern humans from Ethiopia about 125,000 years ago,

by way of the Red Sea southern route, and that the robust features exhibited by the Skhul and Qafzeh hominins represent archaic *Homo sapiens* characteristics that they share with Omo and Herto rather than Neanderthal features resulting from admixture. In fact, it is possible that due to the low sea level and the short distance between lands at the Red Sea crossing, the Ethiopian and Arabian modern humans were part of a continuous population. On the other hand, since close coexistence still occurred, introgression cannot be dismissed.

Behavioral differences between Neanderthals and modern humans

Habitation

A number of differences have been noted between modern humans and Neanderthals in the Levant. It seems that the two groups used different patterns of occupation of the same dwelling, the caves [40]. From the animal remains found in association with modern human fossils at the Skhul, Qafzeh, and Tabun Caves, evidence for single-season occupation is observed. Only winter prey, mainly herbivores such as gazelles, have been unearthed. It seems that modern humans shifted their living quarters seasonally including camping in the field when the weather was more favorable. On the other hand, the Neanderthal fossils, from about the same time, are found in the context of local animals that were abundant during the entire year. In other words, Neanderthals practiced multiseasonal occupation. These differences represent fundamental behavioral variation between the two groups.

Hunting strategies

The hunting strategies of Neanderthals and modern humans were also different. Related to the more robust anatomy and body proportions, Neanderthals are thought to have practiced a more close-proximity style of hunting. It is theorized that they thrusted and forced heavy spears with triangular stone points into the bodies of nearby large prey [41]. This type of game hunting plan would have necessitated constant replacement of spear points due to the force employed during the actual penetration of the prey. Therefore, the expectations are that a number of point replacements would be found at the caves. This is indeed the case. On the other hand, the weapons associated with modern human remains are more fragile wooden spears made for the purpose of throwing as projectiles at prey from a distance [41].

Burial practices

The expectations of nonritualistic burial are the lack of relic and personal effects related to the remains. In other words, the distribution of artifacts throughout the site is random not showing specificity to the location of the body. Over the years, the contention that Neanderthals performed ritualistic burials has been the subject of heated debate. Most of the sites exhibit random distribution of artifacts. For the example, even the well-known Neanderthal Flower Burial seems to be the result of rodent activity subsequent to death. On the other hand, some of the modern human burials are associated with objects that may have been of symbolic meaning to the dead. For example, animal bones seem to be non-randomly placed on the body of modern humans. Yet, recent archeological data seem to indicate that some Neanderthals cared about their dead [42].

Neanderthals prevailed in the Levant

The initial modern human migrants from Northeast Africa reached the Levant about 125,000 years ago. Yet, after about 80,000 years before the present, they disappeared from the fossil record. It is speculated that modern humans could have retreated back to Africa as a result of increasing cold weather or even competition with Neanderthals. Neanderthals are not seen in the Near East after around 45,000 to 35,000 years before the present. If in fact modern humans overlapped in time and coexisted with Neanderthals in the Levant subsequent to the first migration out of Africa, it is possible that their interactions were not belligerent and that they even mated. Of course, due to potential biological reproductive barriers, fertile offspring may have been compromised. Biological reproductive barriers include physiological or behavioral differences that interfere with interbreeding and cross-fertilization between populations. The modern human bones from the Zhiren Cave in China dating to about 100,000 years ago [38] suggest that the early incursions from Northeast Africa into the Levant may not have died in Arabia but were fertile in the peopling of the world reaching into the Far East with possible interbreeding.

Review questions and exercises

1 Enumerate and discuss the milestones that define the hominin lineage.

2 What geological forces have made the East African Ridge?

3 How the East African Range may have contributed to early human evolution?

4 What selection pressures may have contributed to the adoption of bipedalism in early hominin evolution?

5 Review the evolutionary timeline of the human brain starting with *Australopithecus*.

6 Discuss the genetic changes in the hominin lineage that may explain the rapid evolution of the human brain.

7 Name and contrast the characteristics of hominin types starting with *Australopithecus*.

8 Prior to modern humans which hominins migrated out Africa. Indicate their time and geographical range.

9 Speculate how DNA sequences such as SRGAP2 and DUF1220 could have acted to bring about profound changes in the hominin brain.

10 Explain how the Saharan pump works and how this phenomenon may have help shape hominin evolution.

11 When the first migrations out of Africa took place and which routes modern humans may have used?

12 What evidence exists suggesting cohabitation of the Levant by modern humans and Neanderthals?

13 What behavioral differences were exhibited between modern humans and Neanderthals in the Near East?

References

1 Huntington E. 2009. *The Red Man's Continent: A Chronicle of Aboriginal America*. Standard Publications, Cincinnati, OH.

2 Clark J. 1994. African *Homo erectus*: old radiometric ages and young Oldowan assemblages in the middle Awash Valley, Ethiopia. *Science* **264**:1907–1909.

3 Zollikofer CP, De León MS. 2013. Pandora's growing box: inferring the evolution and development of hominin brains from endocasts. *Evol Anthropol* **22**:20–33.

4 Balter M. 2011. Was North Africa the launch pad for modern human migrations? *Science* **331**:20–23.

5 Haddok E. 2008. Birth of an ocean: the evolution of Ethiopia's Afar Depression. *Sci Am* **29**:12–27.

6 Gibbons A. 2002. Profile: Michel Brunet. One scientist's quest for the origin of our species. *Science* **298**:1708–1711.

7 De Menocal P. 2004. African climate change and faunal evolution during the Pliocene–Pleistocene. *Earth Planet Sci Lett* **220**:3e24.

8 Conroy G. 2000. *Reconstructing Human Origins: A Modern Synthesis*. W.W. Norton & Company, New York.

9 Maslin MA et al. 2014. East African climate pulses and early human evolution. *Q Sci Rev* **101**:1e17.

10 Lovejoy CO. 2005. The natural history of human gait and posture. Part 1. Spine and pelvis. *Gait Posture* **21**:95–112.

11 McHenry HM. 2009. Human evolution. In: Ruse M, Travis J (eds), *Evolution: The First Four Billion Years*. The Belknap Press of Harvard University Press, Cambridge, MA.

12 Sylvester AD. 2006. Locomotor coupling and the origin of hominin bipedalism. *J Theor Biol* **242**:581–590.

13 Edelman GM. 2006. *Second Nature: Brain Science and Human Knowledge*. Yale University Press, New Haven, CT.

14 Muehlenbein MP. 2010. *Human Evolutionary Biology*. Cambridge University, London, UK.

15 Dennis MY, Nuttle X, Sudmant PH, Antonacci F, Graves TA, Nefedov M, Rosenfeld JA, Sajjadian S, Malig M, Kotkiewicz H, Curry CJ, Shafer S, Shaffer LG, de Jong PJ, Wilson RK, Eichler EE. 2012. Evolution of human-specific neural SRGAP2 genes by incomplete segmental duplication. *Cell* **149** (4):912–922.

16 Marques-Bonet T, Eichler EE. 2009. The evolution of human segmental duplications and the core duplicon. *Cold Spring Harb Symp Quant Biol* **74**:355–362.

17 Dumas LJ et al. 2012. DUF1220-domain copy number implicated in human brain-size pathology and evolution. *Am J Hum Genet* **91**:444–454.

18 Milford H et al. 2002. Palaeoanthropology: *Sahelanthropus* or '*Sahelpithecus*'? *Nature* **419**:581–582.

19 Dawkins R. 2004. *The Ancestor's Tale: A Pilgrimage to the Dawn of Life*. Weidenfeld & Nicolson, London, UK.

20 Boehm C. 1999. *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*. Harvard University Press, Cambridge, MA.

21 Eckhardt RB et al. 2014. Rare events in earth history include the LB1 human skeleton from Flores, Indonesia, as a developmental singularity, not a unique taxon. *Proc Natl Acad Sci USA* **111**:11961–11966.

22 Henneberg M et al. 2014. Evolved developmental homeostasis disturbed in LB1 from Flores, Indonesia, denotes Down syndrome and not diagnostic traits of the invalid species *Homo floresiensis*. *Proc Natl Acad Sci USA* **111**:11967–11972.

23 Lowery RK et al. 2013. Neanderthal and Denisova genetic affinities with contemporary humans: introgression versus common ancestral polymorphisms. *Gene* **530**:83–94.

24 Bar-Yosef O et al. 1999. The excavations in Kebara Cave, Mt. Carmel. *Curr Anthropol* **33**:497–550.
25 Diamond JM. 1992. *The Third Chimpanzee: The Evolution and Future of the Human Animal*. Harper Collins, New York.
26 Carbonell E et al. 1999. Out of Africa: the dispersal of the earliest technical systems reconsidered. *J Anthropol Archaeol* **18**:119–136.
27 Lordkipanidze D. 2013. A complete skull from Dmanisi, Georgia, and the evolutionary biology of early *Homo*. *Science* **342**:326–331.
28 Prins HEL et al. 2007. *Evolution and Prehistory: The Human Challenge*. Wadsworth Publishing, Belmont, CA.
29 McDougall I et al. 2005. Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature* **433**: 733–736.
30 McDougall I et al. 2008. Sapropels and the age of hominins Omo I and II, Kibish, Ethiopia. *J Hum Evol* **55**:409–420.
31 Castañeda AS et al. 2009. Wet phases in the Sahara/Sahel region and human migration patterns in North Africa. *Proc Natl Acad Sci USA* **106**:20159–20163.
32 Osborne AH et al. 2008. A humid corridor across the Sahara for the migration of early modern humans out of Africa 120,000 years ago. *Proc Natl Acad Sci USA* **105**: 16444–16447.
33 Olivieri A et al. 2006. The mtDNA legacy of the Levantine Early Upper Palaeolithic in Africa. *Science* **314**:1767–1770.
34 Petraglia MD et al. 2012. Hominin dispersal into the Nefud Desert and Middle Palaeolithic settlement along the Jubbah Palaeolake, Northern Arabia. *PLoS One* **7**:e49840.
35 Černý V et al. 2009. Migration of Chadic speaking pastoralists within Africa based on population structure of Chad Basin and phylogeography of mitochondrial L3f haplogroup. *BMC Evol Biol* **9**:63.
36 Currat M et al. 2010. Human genetic differentiation across the Strait of Gibraltar. *BMC Evol Biol* **10**:237.
37 Walter RC et al. 2000. Early human occupation of the Red Sea coast of Eritrea during the last interglacial. *Nature* **405**:65–69.
38 Liu W et al. 2010. Human remains from Zhirendong, South China, and modern human emergence in East Asia. *Proc Natl Acad Sci USA* **107**:19201–19206.
39 Armitage S, Jasim S, Marks A, Parker A, Usik V, Uerpmann H. 2011. The southern route “Out of Africa”: evidence for an early expansion of modern humans into Arabia. *Science* **331** (6016):453–456.
40 Lieberman DE. 2002. *Neandertal and Early Modern Human Mobility Patterns*. Springer International Publishing, Cham, Switzerland.
41 Shea JJ. 1997. Neanderthals meet modern humans. Modern human origins and Neanderthal extinctions in the Levant. *Athena Rev* **2**:21–32.
42 Rendu W et al. 2013. Evidence supporting an intentional Neandertal burial at La Chapelle-aux-Saints. *Proc Natl Acad Sci USA* **111**:81–86.

CHAPTER 7
Culture

No chimpanzees will ever form a Department of Human Studies.

—*Richard Levins and Richard Lewontin* [1]

SUMMARY

The complexity of our culture, as opposed to those of other animals, is a distinctive aspect separating humans from nonhumans. After a brief characterization of the culture concept, culture origin and development, factors influencing it, and specifically the biology–culture interaction are considered. Language variability is presented as a specific type of cultural evolution. Then, the domestication process of plants and animals is reviewed, as well as art and the concept of free will, morality, and religion. In what sense these developments will influence the biology of our species? The questions raised here will be considered in some detail in Chapters 10 and 11.

Concept

Traditionally, the complexity of our culture has been considered a fundamental distinction between our species and other animals. Yet, despite the use of this term to position us above other species, the term "culture" is not easily defined. We could characterize culture as a complex of beliefs, values, behaviors, and traditions associated with a given population, which are individually acquired through imitation, teaching, and other forms of social learning. This includes knowledge, values, and abilities, which are expressed as behaviors or tools [2].

Human culture presents an astonishing diversity. For example, the number of extant languages has been estimated at about 7000, and a compilation in 1990 listed as many as 3814 distinct ethnographic societies [3]. The primary factor involved in the acquisition of culture is, of course, our cognitive system, ultimately derived by our genome, which generates behaviors and processes. Therefore, any attempt at interpreting cultural variability should consider both biology and sociocultural factors, as well as their interaction, to obtain a truly comprehensive picture.

Origin and development

The origin of culture in humans can be investigated using two main approaches: (a) comparison with other animals and/or (b) the study of paleoanthropological or archeological materials. In relation to the first, the most general concept of culture (social learning) would imply that several hundred vertebrates (fishes, birds, nonhuman primates) would have cultural structures associated with their social interactions. Yet, the phenomena observed in these animals are considerably much simpler than those of humans. If the behaviors found are classified into four classes: (a) tags (shared responses to predators); (b) signals (conventional communication signs); (c) abilities (tool construction); and (d) symbols (which would define an association or a group), humans are the only group to utilize symbols within their cultural structure, while the great apes are the only species to make use of all of the three others. The comparison between the stone tools made by humans, chimpanzees, bonobos, and orangutangs, on the other hand, indicates that an important distinction between humans and nonhumans is the reuse of previously utilized material, suggesting a reflexive conscience and recognition of a symbolic component in their behavior [4].

Genomes, Evolution, and Culture: Past, Present, and Future of Humankind, First Edition.
Rene J. Herrera, Ralph Garcia-Bertrand, and Francisco M. Salzano.
© 2016 John Wiley & Sons, Ltd. Published 2016 by John Wiley & Sons, Ltd.

Box 7.1 How the fossil and archeological records could furnish indications about culture's components.

Culture's wide components	Potential paleobiological manifestations
Learning capacity	Technology and its variation Maybe brain size
Organization and social structure	Density, structure, and distribution of the archeological remains Sexual dimorphism in the fossil hominins Elements of the material culture not ecologically functional
Symbolic reasoning	Maybe brain size Anatomical basis for language Variability in the material culture
Tradition conservation and change	Variation and regional longevity of archeological components

Source: Foley and Lahr 2003 [5].

Box 7.2 Traits that could be used to identify modern human behavior.

1. Burial of dead persons, indicating ritual.
2. Art, ornamentation, and decoration.
3. Ochre's symbolic use.
4. Work in bones and horns.
5. Blade technology.
6. Standardization of artifact types.
7. Artifact diversity.
8. Construction of complex ovens.
9. Organized use of the domestic space.
10. Amplified exchange networks.
11. Effective exploitation of large mammals.
12. Seasonal strategic mobility.
13. Use of unfavorable environments.
14. Subsistence based on fishing and bird domestication.

Source: Henshilwood and Marean 2003 [6].

As for the fossil or archeological records, Box 7.1 furnishes some indications. Three basic elements can be distinguished in the cultural evolution process: (a) individual learning capacity; (b) the organization and social structure of the communities where these people live; and (c) symbolic reasoning, which derives from the interaction of these two variables. In addition, conservation or changes in the traditions that are developed should provide indications of the presence or lack of innovative processes. Identification of these elements in the paleontological or archeological records would involve different anatomical traits of the persons involved, as well as the types of tools used, their style independent of their practical function, and their spatio-temporal distribution.

Signs that would indicate the presence of what has been called modern human behavior would also be important for these interpretations. Box 7.2 indicates 14 items that could be considered. They include tool standardization and diversification; mobility; domestication signs and the hunting of large prey; and evidence for artistic or ritualistic behavior.

Sociocultural development of modern humans involved important changes in structure, modes of subsistence, and ways of living. They included changes from hunter-gatherers to agriculturalists, urban dwellers, and the concomitant modifications that presently involve living in large, industrialized surroundings. Here is not the place for the detailed examination of these changes. A specific point that could be considered would be the emergence and persistence of inequality in pre-modern and modern societies, and whether economic success would have anything to do with our genomes [7,8]. Some suggestions of these authors are basically unacceptable given our present knowledge; see the section on biology–culture interaction.

Factors that could condition cultural evolution

An initial question that could be posited is as follows: "Are evolutionary changes in our genome a cause or a consequence of cultural innovation?" This question is still under debate by researchers. Some argue that a small number of regulatory genes, which led to advanced, complex cognition, would be essential for the beginning of the so-called human revolution. Others, however, maintain that the opposite is true, and that the environment provided by key cultural changes would allow the emergence of genetic variants that otherwise would be forbidden due to natural selection [9]. The question could be answered by supposing that the interaction of the two factors, acting simultaneously on time, would be responsible for our unique characteristics.

Undoubtedly, many factors might have influenced the fantastic development of human culture. Some examples of studies that focused on specific points will be presented here, but this list is not exhaustive.

We begin by considering population size. Since transmission events for complex tasks are many times imperfect, a learner could acquire a better skill than the demonstrator by chance if the number of transmission events is large. This event, in turn, would depend on population size. Since imitation of successful individuals is an important factor in humans, a given individual whose skill improved over the initial demonstrator would become the new demonstrator, starting a process of cultural evolution. This hypothesis was experimentally tested in a sample of 366 men who participated in a dual-task computer game, and it was verified that players in larger groups maintained higher cultural complexity, which would favor cultural evolution [10].

However, of course, population size alone is not enough. Henshilwood and Dubreil [11] proposed that a primary factor, either the reorganization of the temporoparietal areas implicated in theory of mind, perspective taking, and attentional flexibility, or improved connectivity of these regions with the prefrontal cortex, could be responsible for cultural evolution. A complementary view suggested that lithics, entrenched at the base of the sociotechnical system, although strongly conservative, would make changes possible at upper hierarchical levels. Changes external to culture, including ecological or climatic, may also induce cycles of innovation, which would therefore lead to cognitive evolution and the destruction/construction of ecological niches that could create a feedback mechanism of ever-changing development [12].

Cooperation may have also been important for cultural evolution. Hunter-gatherers show extensive cooperation among members of residential units, including (a) food sharing, (b) allomaternal child care, (c) construction and maintenance of living spaces, and (d) provision of other goods. The social structure of hunter-gatherers also involved coresidence of kin and genetically unrelated individuals, leading to large interaction networks that could have been responsible for the development of cumulative culture [13].

Niche construction can be defined as a process by which organisms can alter the ecological environment for themselves, their descendants, and other species. Elements of the system are semantic information, forms of behavior acquired through social learning (for instance, subsistence patterns or social norms), and physical resources, referring to aspects of material culture, such as nutritional resources or tools. Recent developments in social theory have tried to transcend the classic dichotomy between structure (the rules and institutions of societies) and agency (the intentions, motivations, and performances of individuals) [14]. The joint dynamics of cultural transmission, selection, and assortative mating was analytically considered, and it was found that it can lead to cycles of oscillations and stability, as well as to polymorphisms of all cultural phenotypes [15]. An example of niche construction involving an autochthonous American allele, the ATP-binding cassette transporter A1 (*ABCA1*) variant *Arg230Cys*, was reported by Hünemeier et al. [16]. A series of molecular genetic analyses, and the striking correlation between the *230Cys* frequencies and the distribution of maize pollen relics found in nearby places in Mesoamerica (Figure 7.1), suggest that maize domestication was the driving force in the increase of *230Cys* in this region.

It is intuitive that persons ought to be selective with respect to when and who they copy, and that natural selection should favor the deployment of adaptive social learning strategies that would guide reliance on social information. This hypothesis was subjected to an experimental test. It was found that multiple factors, including the number of demonstrators, consensus among them, confidence of subjects, task difficulty, number of sessions, cost of asocial learning, subject performance, and demonstrator performance, all influenced in an adaptive way the use of social information. This experiment provides support for the view that human social learning is regulated by adaptive learning rules [17].

Charles Darwin, as early as the 19th century, was emphatic about the importance of fire for the genesis of modern humans, mentioning that this importance was only surpassed by language. This view is consistent with the fact that no human population has ever been found living exclusively on raw wild food. The advantages of cooked food derive from the fact that cooking consistently increases the energy obtainable from most foods. In addition, it reduces the metabolic work related to digestion, softens the material to be eaten, and makes it less pathogen-bearing [18].

Compared with mammals, primates tend to fall along the slow end of the life history continuum (slow maturation, increased adult body size, late reproduction, high

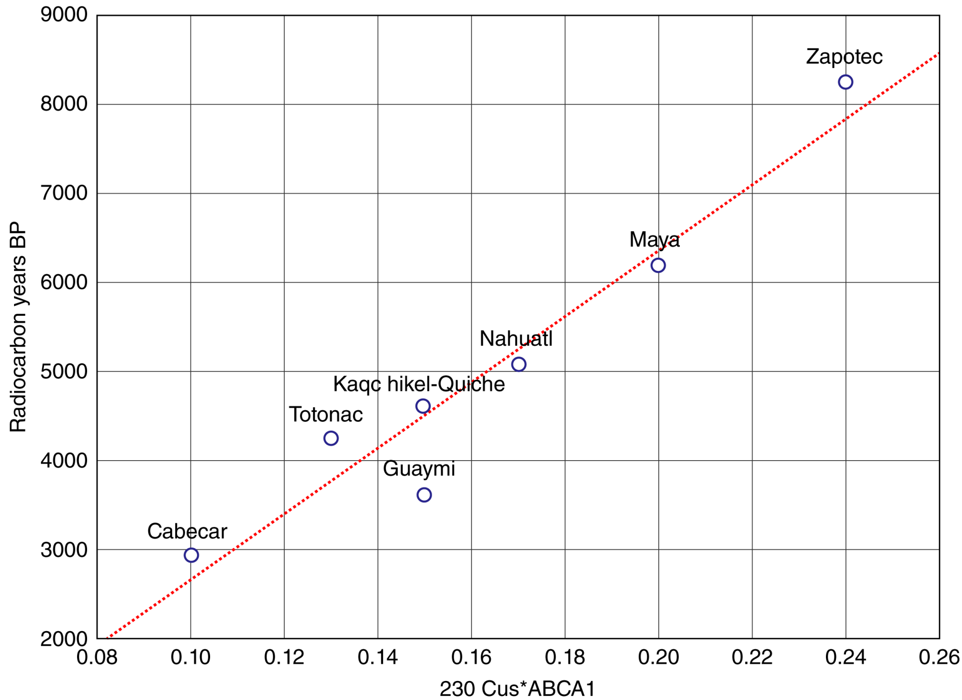


Figure 7.1 *ABCA1*230Cys* frequencies versus radiocarbon ages of maize domestication (*Zea mays* pollen relics). Spearman's rho value = 0.94; $p = 0.002$. (*Source*: Hünemeier et al. 2012 [16].

investment in a small number of offspring, and longer life). Humans, however, are unique among primates in having a mixed-pace life history. They are slow in two variables (life span, age at first birth), but fast in two others (weaning, interbirth interval). The control of fire and consumption of cooked food could have favored the slow component by reducing extrinsic mortality (keeping predators away and reducing infections) and by raising the nutritional value of provisioned food, increasing the value of assistance from older individuals in offspring care. The fast components would also be favored. Earlier weaning would be made possible by cooked foods being softer, more easily digestible, and less pathogen-bearing than raw foods, while interbirth intervals would be reduced due to the energetic advantages of a cooked diet and the provisioning that cooking facilitates, allowing for greater stability in nutritional status of mothers [18].

In terms of phylogenetic anatomical changes, fire control would have influenced (a) increased brain volume, as an allometric compensation for the reduction of gut volume; (b) increased body mass, promoted by reduced mortality due to fire use; (c) reduced molar area, as a result of food being softened by heat; (d) reduced gut volume, due to increased digestibility of cooked food; (e) loss of arboreal adaptations because fire would make possible sleeping on the ground by keeping predators away; and (f) reduced body hair, favored by the extra warmth achieved at night by resting near a camp fire [18].

The human ability to live and cooperate in huge groups through the emergence of large-scale societies should also be considered. A specific model was developed with the central premise that this process evolved as a result of two factors: (1) intense competition between societies, primarily warfare; and (2) geographic features favoring the development of military technologies. The model was simulated considering the Afroeurasian landmass and tested against a data set of the spatiotemporal distribution of large-scale societies of the region between 1500 BC and 1500 AD. Overall, the model explained 65% of the variance in the data. Studies like this one are important to provide quantitative hypotheses about the broad features of cultural evolution [19].

Biology–culture interaction

In 1870, Alfred Russel Wallace (1823–1913) questioned the role of natural selection in the evolution of the human mind. How could the conventional natural selection select the exceptional capacities developed by the human mind? Varki et al. [20] suggested that the explanation, at least in part, would be selection relaxation for the maintenance of the genome's integrity, favoring a plasticity along time to invent, disseminate, improve, and culturally transmit complex behaviors for many generations, with no need to rigidly fix them through genotypic control.

Aspects of human life histories (birth and childhood, puberty, reproductive adult life) are also important. A specific example would be the long postmenopausic life of women, which is postulated to favor the survival of the offspring of their daughters, the so-called "grandmother hypothesis" [21,22].

Behavior can be classified into three types: innate, socially learned, and individually learned. Innate behavior involves the direct expression of gene-codified information. In social learning, information is transferred between socially interacting persons. Individual learning involves transmission of information free of any social influence. Modeling [23] indicates that innate, socially learned, and individually learned behaviors would be favored by selection at short, intermediate, and long intervals, respectively.

Other mathematical models of gene–culture coevolution are listed in Box 7.3. The factors involved are mostly variable, including diet, learning, relationships between persons and groups, language, intelligence, personality, and sex. Genes that were identified

Box 7.3 Mathematical models of gene–culture coevolution.

1. Evolution of learning, social transmission, and culture.
2. Lactase persistence in adult life genes and dairy food.
3. Evolution of language.
4. Evolution of intelligence and personality.
5. Evolution of cooperation.
6. Incest taboos.
7. Sexual behavior and paternity beliefs.
8. Control of the sex ratio.
9. Evolutionary consequences of niche construction.

Source: Laland et al. [2].

Box 7.4 Genes identified as subjected to rapid positive selection and the cultural selection associated with them.

No. of genes	Function or phenotype	Inferred cultural selection
23	Milk digestion, sugar and alcohol metabolism	Domestication, social use of alcoholic beverages
4	Detoxification of plant secondary substances	Plant domestication
31	Immunity, resistance to pathogens	Demographic structure derived from subsistence ways
16	Cold or heat tolerance	Migratory patterns
21	External visible phenotypes	Sexual selection
29	Nervous system functions, vocal learning	Social relations
4	Skeletal development	Sexual selection
2	Maxillary muscles, tooth enamel thickness	Fire use for food cooking

Modified from Laland et al [2].

as being subject to rapid positive selection associated with cultural elements are listed in Box 7.4. The most frequently identified genes ($n = 31$) involved immunity and resistance to pathogens influenced by the demographic structure of their carriers. The next highest number of genes ($n = 29$) were those related to nervous system functions and vocal learning, as a result of interpersonal social relationships. Those genes that condition visible characteristics have been influenced by sexual selection, and the domestication process was important for a series of others.

Cultural evolution develops much more rapidly than biological evolution, and characteristics related to the two types of transmission are listed in Box 7.5. While genetic transmission occurs basically in a vertical way (parent–offspring), cultural transmission can take place in a vertical, horizontal (between persons of the same generation), or oblique (teacher–student) manner. Methods to identify how transmission of cultural traits took place in specific cases were presented by Borgerhoff

Box 7.5 Similarities and differences between genetic and cultural transmissions.

Characteristic	Genetic transmission	Cultural transmission
Information unit	Gene	Meme[a], seme[a]
Information vector	DNA	Behavior and central nervous system
Transmission mechanism	DNA duplication	Imitation, social facilitation, learning, teaching
Variability	Mutations and other types of DNA lesions	Learning errors, innovations
Impact of the change	Mostly deleterious	Variable
Transmission of acquired characteristics	No	Yes
Type of process	Darwinian	Darwinian or Lamarckian
Rhythm	Slow	Rapid

Modified from Danchin et al. [25].

[a] *Meme* would be the unit associated with imitation, while *seme* derives from signal and emphasizes the symbolic nature of culture [26].

Mulder et al. [24]. The authors demonstrated that in certain African groups the inheritance of polygyny could be both vertical and horizontal, but among the U.S. Na-Dene this transmission was more complex. Thus, in a phylogeny, there would be a minimum of 15 additions and 4 losses for a high prevalence of polygyny, and there were indications that the trait would be associated with the type of resource exploitation by men.

A special type of transmission is the one related to the inheritance of material goods between generations, a problem that indirectly is related to unequal wealth and class formation in human society. This question was extensively considered in a set of contributions (1 introduction, 5 articles, and the answers of authors to 10 comments) [27]. These contributions examined three generic categories of wealth: (a) *material*: physical possessions such as land, livestock, and household goods; (b) *relational*: the individual's position in social networks; and (c) *embodied*: strength, practical skills, and reproductive success. Transmission was examined in relation to 40 measures in 21 pre-modern societies classified into four production systems: hunter-gatherers, horticulturalists, pastoralists, and agriculturalists. They found that (a) material wealth is more important in pastoral and agricultural systems; (b) wealth transmission from parent to offspring is markedly higher for material wealth compared with embodied and relational wealth; (c) aggregate wealth is transmitted to a higher degree among pastoralists and agriculturalists; (d) the degree of inequality is greater for material wealth; and (e) the degree of intergenerational transmission of wealth is correlated with wealth inequality.

Language

Two definitions and a classification of human language are provided in Box 7.6. One of the definitions emphasizes the communication process between persons (a social question), and the other the background to the representative and analytical thought (a private question). On the other hand, we can classify language not only in its broad sense, which involves the sensory–motor and conceptual–intentional systems, but also in its narrow sense. The fundamental property of the latter is *recursion*. Recursion generates an infinite array of expressions from a limited number of elements using syntactic rules. These rules are applied to the distribution of words within sentences and the distribution of sentences within discourses by means of some logical relationship. This property of applying syntactic rules to formulate logical, flowing thoughts is only found in humans.

Language structure occurs due to the interaction between three complex, adaptive systems. We acquire language by means of learning mechanisms, which are a

Box 7.6 Human language characteristics.

1. Definitions
   1.1. A specific cultural system, constituted by signals or signs, that serves to foster communication between individuals, mediated by sense organs.
   1.2. An inner component of mind/brain that relates form and meaning, their characterization, and other attributes.
2. Classification
   2.1. Broad sense
      2.1.1. Sensory–motor: to speak, it is necessary to have a fine and quick motor control, as well as elaborate larynx, mouth, face, and tongue movements and respiration, synchronized to a cognitive activity.
      2.1.2. Intentional–conceptual: capacity to acquire and use abstract concepts, directing them in an intentional manner to specific persons.
   2.2. Narrow sense
      Presence of a computational system (syntax) that generates inner representations, mapping them in the sensorial–motor interface by means of a phonological system and in the conceptual–intentional interface by a formal semantic system. Its nuclear property is recursion, that is, the capacity to generate an infinite array of expressions from a limited set of elements.

Source: Hauser et al. [28].

part of our biological constitution. Through learning, language information is transmitted to one or more populations of individuals over time, leading to linguistic universals. The relationship between the learning machinery and the linguistic universals is not trivial, but the process will affect the biological adaptive value of individuals who speak a given language, closing the learning–culture–evolution interaction chain [29].

Young children's language is similar to signing patterns of nonhuman primates: both seem to result from imitation, because they show limited and formulaic combinatorial flexibility. One study [30], however, verified that at least some components of child language follow abstract rules from the outset of syntactic acquisition. This study suggests that children, but not nonhuman primates, use a rule-based grammar.

What might be the function of language? It was reported that Charles-Maurice de Talleyrand (1754–1838), a famous French politician and diplomat, said "Language was invented so that persons could hide their thoughts from one another!" [31]. Locke [32] was less ironic, suggesting that vocal communication has served and serves to signal status and to solve conflicts, promoting collaboration and the sharing of environment's resources. In addition to providing a framework for social interaction, the internal component of language has also been important for the development of a series of abstract concepts, such as number representation and statistical inference [33].

Human language requires an anatomy that is specific for our species, due to the tongue's descent in direction to the pharynx. Language also requires a brain that could freely reorder a finite set of motor activities to form a potentially infinite number of words and sentences. Neural circuits that link cortex regions to basal ganglia and other subcortical structures regulate the motor control including speech production, as well as cognitive processes that include syntax. Dating of the Forkhead Box P2 (*FOXP2*) gene, which regulates the embryonic development of these subcortical structures, indicated that the human form of the gene should have arisen approximately simultaneously with the emergence of anatomically modern humans, and a human anatomy that would suggest a speaking process appears in the fossil record only in the Upper Paleolithic, 509,000 years ago. Neanderthals probably could not speak (see Chapter 2) [34].

It is not easy to estimate the vast array of languages spoken in the world, but a reasonable value situates this number around 7000, classified into 17 families [35]. By far, Chinese is the language spoken by most people (1.1 billion) [36], but variability is the rule. What factors influenced this diversity? A series of 14 articles published in an issue of the *Philosophical Transactions of the Royal Society, Series B* in 2010 analyzed this question in detail [37]. A recent controversy arose, whether phonemic diversity would support a serial founder effect model of language expansion from Africa, as suggested by

Atkinson [38], based on a global sample of 504 languages. His proposal was questioned by Hunley et al. [39], who considered 725 widespread languages. They concluded that phoneme inventories provide information about recent contacts, but their rates of rapid change could not give information about more ancient evolutionary processes. As a matter of fact, there is a dichotomy among linguists, some investigating the historical processes of language evolution, while others are skeptical about the approach due to the linguistic rate of change.

Another interesting result [40], considering the question of sex-specific transmission of language change, suggested that this change in an already populated region would require a minimum proportion of immigrant males, while those with a predominance of female immigrants would represent more ancient settlements. This correlation would be due to causal factors of a social nature, related to marriage rules and male dominance in the familial environment.

Domestication

Domestication is presently being actively investigated, as exemplified by two international conferences whose main results were presented in 2011 [41] and 2014 [42]. Domestication can be defined as a selection process for adaptation to human agroecological niches and, at some point in the process, human preferences [42]. The criteria for identifying domestication differ significantly for plants and animals. Plants rather quickly show distinct morphological changes, while animals are much slower in presenting such developments. In the case of animals, three types of domesticated animals can be identified: (a) commensals, adapted to a human environment (dogs, cats, and guinea pigs); (b) prey, sought for food (cows, sheep, pigs, and goats); and (c) animals targeted for draft and nonfood resources (horses, camels, and donkeys).

As for plants, Box 7.7 presents a selected list of commonly observed traits that may be found due to domestication and posterior diversification. In seeds, traits related to size, number, morphology, and substance changes can be listed, as well as those that influence reproduction. The corresponding traits for roots and tubers are flavor and increased nutritional quality, while in fruits, flavor, size, attractiveness, and other aspects were important.

Agriculture was independently developed in at least 11 different places around the world, and two major chronological periods seem to have been particularly

Box 7.7 Crop traits associated with domestication (stage 1) and diversification (stages 2–4).

Characteristic	Stages			
	1	2	3	4
Seed crop	Larger seeds Thinner seed coat Inflorescence architecture Increased yield potential and productivity	More seeds Pigment change Flavor change Change in starch content Nonshattering seeds	Reduced vernalization Modified hormone sensitivity Synchronized flowering time Dwarfism	Increased yield Improved eating quality
Root and tuber	Flavor change Ability to thrive in modified landscape	Reduced toxicity Abiotic stress tolerance	Hybridization Increased yield	Improved nutritional quality
Fruit	Flavor change Shortened life cycle Softer content	Increased size variation	Improved pollination success	Increased quality and delayed senescence Attractiveness and even ripening

Source: Meyer and Purugganan [43].

Table 7.1 Approximate dates for the appearance of domesticated species in several regions of the world.

Region and organism	Date of appearance (cal BP) (in thousands)
1. Southwest Asia	
1.1. Plants	11.5
1.2. Animals	10.5
2. China	
2.1. Millet	10.0
2.2. Rice	>7.0
3. South Asia	
3.1. Plants	5.0
3.2. Animals	8.0
4. Africa	
4.1. Plants	5.0
4.2. Animals	9.0
5. New Guinea	
5.1. Plants	>7.0
6. Eastern North America	
6.1. Plants	5.0
7. Mexico	
7.1. Corn	9.0
8. South America	
8.1. Plants	10.0
8.2. Animals	6.0

Reproduced from Price and Bar-Yosef 2011 [41] with permission of University of Chicago Press.

important: (a) the transition to the Holocene, 12,000 to 9000 years before present (YBP); and (b) the middle Holocene, between 7000 and 4000 YBP. Table 7.1 gives some selected approximate dates of domesticated species in eight different regions of the world. Although early Holocene plant domestication occurred independently in the Old and New Worlds, early Holocene animal domestication seems to have been restricted to the Near East. All in all, it is calculated that 2500 plant species have undergone domestication and 250 species are considered as fully domesticated [43].

Additional information about the dates of the earliest signs of domestication in both plants and animals, and covering eight world regions, is given in Table 7.2. Not listed there are the dogs, the first organism to be domesticated in Late Pleistocene. For plants, the numbers vary from 11,000 years ago (wheat, Southwest Asia) to 2000 years ago (African rice), while for animals dates are distributed from 10,300 years ago (taurine cattle, Southwest Asia) to 1000 years ago (duck, East Asia).

Table 7.2 Approximate dates of earliest signs of domestication in different regions of the world.

Region and organism	Date of earliest signs of domestication (in thousand YA)[a]
1. Southwest Asia	
1.1. Plants	
1.1.1. Wheat	11.0
1.1.2. Barley	10.5
1.1.3. Pea	10.0
1.2. Animals	
1.2.1. Sheep	9.8
1.2.2. Goat	9.8
1.2.3. Pig	9.7
1.2.4. Cattle (taurine)	10.3
1.2.5. Cat	4.0
2. South Asia	
2.1. Plants	
2.2.1. Rice (indica)	4.0
2.2. Animals	
2.2.1. Cattle (zebu)	8.0
2.2.2. Water buffalo	4.5
3. East Asia	
3.1. Plants	
3.1.1. Rice (japonica)	7.6
3.1.2. Soybean	5.5
3.1.3. Melon	4.0
3.2. Animals	
3.2.1. Pigs	8.5
3.2.2. Silkworm	5.4
3.2.3. Horse	5.5
3.2.4. Bactrian camel	4.5
3.2.5. Duck	1.0
3.2.6. Chicken	4.0
4. New Guinea	
4.1. Plants	
4.1.1. Banana	4.0
5. Africa and South Arabia	
5.1. Plants	
5.1.1. Sorghum	4.0
5.1.2. Rice (African)	2.0
5.2. Animals	
5.2.1. Cattle (taurine)	7.7
5.2.2. Donkey	5.5
5.2.3. Dromedary camel	3.0
6. North America	
6.1. Plants	
6.1.1. Squash	5.0
7. Mesoamerica	
7.1. Plants	
7.1.1. Squash (pepo)	10.0
7.1.2. Maize	9.0
7.1.3. Common bean	3.0

(*continued*)

Table 7.2 (*Continued*)

Region and organism	Date of earliest signs of domestication (in thousand YA)[a]
7.2. Animals	
7.2.1. Turkey	2.0
8. South America	
8.1. Plants	
8.1.1. Peanut	5.0
8.1.2. Cotton	6.0
8.1.3. Coca	8.0
8.1.4. Manioc	7.0
8.1.5. Quinoa	3.5
8.1.6. Yam	5.5
8.2. Animals	
8.2.1. Llama	6.0
8.2.2. Alpaca	5.0
8.2.3. Guinea pig	5.0

Source: Larson G et al. 2014 [42].
[a] YA: years ago.

Hunting and gathering was the primary subsistence strategy for more than 95% of our existence as a species. Why was this apparently successful strategy abandoned in favor of food production? Prime factors could be exogenous or natural (climate change, scarcity of wild species, population pressure) or endogenous or cultural (lesser mobility of farmers favoring the costs of child rearing, accumulation of resources leading to arms investments, and eventually to aggression to neighboring groups). The fact is that the process was gradual, and the two ways of subsistence were maintained (and still are, in some isolated human groups) as a dual strategy. Evidence has been presented [44] that is inconsistent with the hypothesis that at the dawn of farming the productivity of labor in cultivation generally exceeded that in foraging, and indeed suggested the opposite.

Below are examples of some specific recent genomic investigations on a few selected domesticated organisms. We begin with maize, which is a model system in the area of domestication research. The domestication process began around 9000 years ago from *Zea mays parviglumis* (teosinte) and a study [45] examined how this process has reshaped the transcriptome of maize seedlings considering 18,242 genes from 38 maize and 24 teosinte genotypes. They found that 600 genes showed different expression in the two organisms, and they also observed altered coexpression profiles, identifying a subset of genes that were likely targets of selection during domestication.

Banana ranks next to rice, wheat, and maize in terms of its importance as a food plant. Domestication occurred as a series of crossings and selections that should have occurred about 6500 years ago. Most edible bananas are diploid or triploid hybrids from *Musa acuminata* (A-genome) alone or from hybridization with *Musa balbisiana* (B-genome). Studies involving 400 wild and cultivated accessions cultivated in an agronomical center in Guadeloupe were investigated, including additional lineages from Cameroon and Nigeria [46]. The evolution from wild to edible bananas involved seed suppression and parthenocarpy development, and the current global production of more than 100 million tons is based on large-scale vegetative propagation of a small number of genotypes, which derive from only a few ancient sexual recombination events. The danger of agronomical disasters due to new diseases and pests, as well as ecological changes, is high and should be minimized through new crossings and selections, with punctual changes in given genotypes.

Grape, on the other hand, is the most valuable horticultural crop in the world. Archeological data suggest that cultivation of the domesticated grape, *Vitis vinifera vinifera*, began around 7000 years ago in the Near East. A study of 1000 samples from the U.S. Department of Agriculture [47] detected a weak domestication bottleneck that was followed by thousands of years of widespread vegetative propagation. Although substantial genetic diversity has been maintained, the crop faces severe pathogen pressures, and its long-term sustainability will depend on a vigorous research program.

Dogs (*Canis familiaris*) are considered humans' best friends. As was previously indicated, their domestication predated the rise of agriculture, and should have occurred 11,000–16,000 years ago. Genomic data support the notion that dogs are descended exclusively from the gray wolf (*Canis lupus*). They are the only large carnivore ever to have been domesticated, and the resulting process is an amazing array of sizes (from the diminutive 1 kg Chihuahua to the 100 kg Mastiff) and forms. As a matter of fact, dogs far exceed the variation in skeletal and cranial proportions exhibited by the entire carnivore order!

Extensive genomic studies have been performed in dogs, and mention will only be made of selected examples. Wayne and von Holdt [48] provided a useful review of these investigations, emphasizing their two main modes of evolution: (a) fixation of discrete mutations of large effect in individual lineages; and (b) selective breeding for distinct phenotypic or functional attributes ranging from sight and scent hounds to dogs with special abilities for herding, swimming, running, attentiveness, hunting, lethargy, and aggression. The authors also identified a black coat color mutation that evolved in dogs and was afterward transferred to North American gray wolf populations, providing an example of inverse domesticated-to-wild gene flow. Another study [49] analyzed 49,024 autosomal single-nucleotide polymorphisms (SNPs) in 1375 dogs (of 35 breeds) and 19 wolves. None of the so-called ancient breeds derive from regions where the oldest archaeological remains have been found, suggesting that they were formed by geographical and cultural isolation from other lineages, and are not representative of ancient precursors. On the other hand, recent genomic studies in wolves and dogs indicated that the divergence between the two created a much larger pool of genetic diversity in the dog population and a need to reevaluate past hypotheses concerning dog origins is necessary [50].

Horses served to provide food, facilitate transportation, and enhance warfare capabilities. Their domestication occurred around 6000–7000 years ago, and genomic studies suggest that the process involved closely related male animals, since there is virtually no sequence diversity in their Y chromosomes, contrasting with their high diversity in terms of mitochondrial DNA (mtDNA) haplogroups [51].

Goats were one of the first animals domesticated, originating from the wild *Capra aegagrus*, maybe due to the unpredictable availability of wild game over the short and long terms derived from overhunting and ecological conditions in the region. It is probable that foddering and transhumance (moving from one grazing site to another) were resource management strategies integral to the transition from hunting to herding, but these behaviors are not visible using traditional zooarchaeological methods. Stable isotopic analyses, as a direct measure of diet, could, however, provide information about these processes. This technique was used by Makarewicz and Tuross [52], and they obtained evidence that humans provisioned goats with fodder, and mobilized herds to different pastures, as early as 8000 years ago.

Art

Art can be defined as a product that is aesthetically pleasing, with no immediate practical application. Art includes music, dance, ritual, decoration of the body and many other surfaces, and, more recently, written works (novels, poetry, and assays).

Interestingly, many chimpanzees enjoy painting with color, but there is no evidence that this ability is exercised in the wild. The roots of art, however, are undoubtedly linked to our evolutionary development of cognition.

Archeological evidence indicates that some forms of art already existed in prehistoric Africa, which included the use of color, engravings, bone manipulation, and bead making. Undoubtedly, however, there is an unrivalled wealth of archeological European material, indicating the Upper Paleolithic in that continent (around 30,000 years ago) as a period of an authentic revolution manifested in drawings, paintings, and sculptures. Yet, we simply cannot know how much art may have been created using perishable material, or expressed in songs and dances, before that period.

Art involves a coded communication process between the artist and the viewer. Full of a symbolic nature, it implies that the artist is trying to persuade the viewer to see the world in his/her way. Using paintings that have been preserved, it has been possible to detect potential genetic aberrations in some famous artists. For instance, it has been suggested that Vincent van Gogh (1853–1890) and Edgar Degas (1834–1917) may have had a form of protanopia (color blindness) and late-onset macular degeneration, respectively [53]. Paintings and sculptures throughout time have also documented the historic presence of genetic malformations and other conditions (Down syndrome, albinism, and achondroplasia, a form of dwarfism) in many populations [53].

The earliest archeological evidence of art concerns body decoration. Grinding or scraping ochre to produce a powder for use as a pigment was common practice in Africa and the Near East in the Middle Stone Age, and a 100,000-year-old ochre-processing workshop was found in a site at Blombos Cave, South Africa [54]. The practice of piercing teeth, shells, and bones, stringing them singly

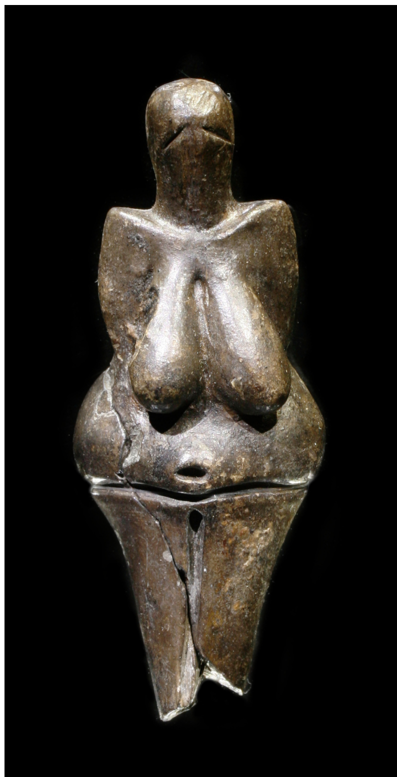
or multiply to make a pendant or necklace, is the oldest known form of personal decoration after body painting [55]. These pieces are thought to have been used by Neanderthals and modern humans in Africa and Europe 80,000 to 50,000 years ago [56,57].

Decorative patterns (parallel lines, nested curves, and zigzag markings present in bone implements and rocks) are also very ancient, and were made by both Neanderthal and modern *Homo* artists, as observed in Africa, Middle East, Europe, and Australia [55]. The cognitive activity underlying pattern making is complex, involving planning and intention; it may have symbolic meanings, but may also be examples of more prosaic forms of representation.

Modifications of suggestive forms (tree trunk contours, earth lumps) to create images may be involved in the origin of three-dimensional art. Fossil material found at prehistoric sites far from their presumed site of origin may also indicate that Neanderthals and early *Homo* were able to recognize these fossils as similar or identical to their living forms, attributing to them talismanic or other symbolic functions.

The richness of Upper Paleolithic finds in Europe has always been a source of amazement among archeologists. Especially impressive are cave paintings. Fertility is the dominant theme in both sculptures and cave wall paintings. The prototypical sculpture from the Upper Paleolithic is the Venus figurine (Figure 7.2). On the other hand, the usual interpretation of figures of a human body with an animal head is shamanism, although it may also represent a god who was master of animals.

Let us make a relatively large leap in time and go to Polynesia 2500–3000 years ago. The region was



(a)

(b)

Figure 7.2 Venus figurines. (a) Venus of Dolni Vestonice, the oldest known ceramic, ~27,000 BP. (*Source*: Petr Novák, https://commons.wikimedia.org/wiki/File%3AVestonicka_venuse_edit.jpg. Used under CC BY-SA 2.5, http://creativecommons.org/licenses/by-sa/2.5.) (b) Venus of Hohle Fels, a limestone sculpture, ~24,000 to 22,000 BP. (*Source*: Matthias Kabel, https://commons.wikimedia.org/wiki/File%3AVenus_von_Willendorf_01.jpg. Used under CC BY-SA-3.0, http://creativecommons.org/licenses/by-sa/3.0/.) (See the Color Plates section.)

colonized by one cultural group that then radiated in a series of related societies, which subsequently developed in relative isolation. Over time, different types of canoes were used for transportation between islands, and they have been carefully described, in 1936–1938, by two scholars. Based on this information, a study was designed involving the presence–absence data for 134 design traits, 96 functional and 38 symbolic, distributed over 11 island groups [58]. The objective was to test whether the rates of change in the functional structures were similar to or different from those of the symbolic designs. The question was important, considering that voyaging-related mortality by these means was potentially on the order of 50%. As expected, the rate of change of the functional characteristics was much slower than that of the cultural features.

In another investigation [59], the history of two musical instruments, the cornet (a soprano brass wind instrument) and the Baltic psaltery (a plucked stringed instrument), was investigated. Valved cornets appeared some 180 years ago, and along the time a large number of changes have been documented. The data consisted of 600 constructional descriptions. The earliest physical evidence of Baltic psaltery, on the other hand, dates to the 10th to early 11th century. The information considered consisted of nearly 100 organological descriptions and the geographical distribution of the distinct forms. Key cornet innovations involved (a) valve number, (b) shifting of the second valve slide and valve alignment, (c) changing of the bell exit position and bell placement, and (d) alteration of the bell shape ("trumpetization"). Evolutionary trees showed a reticulate and complex pattern. On the other hand, for the psaltery data, geography and the linguistic relatedness of players seemed to be the main factors determining the patterns, but again extensive reticulations occurred. The authors concluded that traditional phylogenetic analyses have rather limited application for unveiling cultural phylogenies of material things.

Other developments related to music are the following: (a) The construction of a Darwinian music engine consisting of a population of short audio loops. The sounds were submitted to the consideration of 6931 consumers, who rated the loops' aesthetic qualities. Submitted to 2513 generations of change, there was at the beginning an evolution to aesthetically pleasing chords and rhythms. Later, however, evolution slowed, probably due to problems of transmission fidelity. The experiment demonstrated the creative role of consumer selection in shaping the music we listen to. However, other factors should probably be considered in terms of music evolution, such as preferences of others and interpopulation variability [60]. (b) The search for fractals in the music of Johann Sebastian Bach (1685–1750) [61]! Fractals are typically described as exhibiting self-similarity, namely, the part looks like the whole, and the whole looks like a part. In Bach's *Cello Suite No. 3*, patterns of long and short notes within measures reappeared as patterns of long and short phrases at larger scales [61].

We close by mentioning Laura Splan's idea of creating doilies according to the shapes of viruses' molecular structures. Using a computerized machine embroidery process, she produces decorative motifs on the basis of DNA, RNA, protein spikes, capsids, and lipid envelopes that are pleasing to sight, and could be scientifically instructive [62].

Free will, morality, and religion

At birth, the brain of a child contains not less than 100 billion neurons, each forming on average 1000 synapses. Along his/her life, the majority of these neurons are lost, and the properties of the remaining neurons and their connections result from both genetic and life history events, as well as some degree of stochasticism. It is clear, therefore, that our behavior is not rigidly determined, but that it is influenced by these other factors and/or events. These facts indicate that the notion of free will ("We do what we want to do.") has to be considerably reformulated. As a matter of fact, Cashmore [63] asserted that a belief in free will is nothing other than a continuing belief in vitalism (dualism of body and mind), a notion discarded over 100 years ago. On the other hand, Heschl [64] asserted that "man is completely pre-programmed with respect to his attainable knowledge." The solution to these contrasting views is given by Brembs [65] who pointed out that neurobiology shows indications of a general organization of brain function that incorporates flexible decision making on the basis of complex computations negotiating internal and external processing. This property has obvious evolutionary advantages over others, leading to an unpredictable behavior for competitors, prey, or predators. This freedom is independent of consciousness. Our conscious efforts to make a decision have nothing to do with the degree of freedom we may have.

Box 7.8 The three levels of morality.

Level	Description	Comparison (humans/apes)
1. Moral sentiments	Capacity for empathy, tendency for reciprocity, fairness, and ability to harmonize relationships.	There are parallels between human and nonhuman primates.
2. Social pressure	Rewards for cooperative actions, and punishment for the noncooperative.	In nonhuman primates, pressure is less systematic and less concerned with society objectives as a whole.
3. Judgment and reasoning	Internalization of others' needs and goals. Self-reflective moral judgment.	In nonhuman primates, others' needs and goals may be internalized to some degree, but similarities stop at this level.

Source: de Waal [68].

The relationship between brain structure and psychology was tested by Kanai et al. [66] who verified in 90 volunteers that political liberalism was associated with increased gray matter volume in the anterior cingulate cortex, whereas greater conservatism was associated with increased volume of the right amygdala, as evaluated through magnetic resonance imaging. Although these results do not determine whether these regions play a causal role in the formation of political attitudes, they indicate the need to consider both brain structure and psychological mechanisms that may lead to specific behaviors.

The concept of morality is essentially related to our sense of right and wrong, or good and evil. We have the obligation to secure the well-being of persons by acting positively on their behalf and maximizing the benefits that can be attained. The opposite would be to treat persons in an inappropriate way, ignoring their interests, or treating them as mere instruments in self-benefit.

Is morality an exclusively human condition? Ayala [67] answers affirmatively, proposing three necessary conditions for this attribute: (a) the ability to anticipate the consequences of one's own actions; (b) the ability to make value judgments; and (c) the ability to choose between alternative courses of action. According to Ayala, the ability to anticipate the consequences of one's own actions is the most fundamental of the three conditions. It involves the connection between means and ends, anticipating the future and forming mental images of realities not present or not yet in existence. The second condition, to advance value judgments, implies the notion of perceiving certain objects or deeds as more desirable than others. The third relates to our ability to mentally explore alternative courses of action, acting in accordance with our conscience.

de Waal [68], on the other hand, believes that our moral capacity evolved from rudiments previously existent in primates and other mammals. He postulated three levels of morality, as indicated in Box 7.8. Capacity for empathy, tendency for reciprocity, fairness, and ability to harmonize relationships would occur in both human and nonhuman primates. It is only on levels 2 (social pressure for cooperative actions, punishment for the noncooperative) and 3 (internalization of others' needs and goals; self-reflective moral judgment) that distinctions would occur.

The neural basis involved in moral judgements was investigated using transcranial magnetic stimulation, to disrupt the neural activity in the brain's right temporoparietal junction. It was found that this action led the persons to be morally more permissive, especially in situations of attempted harms (i.e., actors who intended, but failed to do harm) [69].

Closely associated with morality is religion. Its universality (all human societies have some type of belief in supernatural, that is, noncorporeal beings) posits an evolutionary dilemma. This derives from the fact that religion adherence results in significant cost in time and energy, and celibate is mandatory for many of religious leaders, thus potentially working against its frequency. Religion can be conceptualized as a unified system of beliefs and practices uniting a given community, and has connections with myth, ritual, taboo, symbolism, morality, altered states of consciousness, and belief in noncorporeal beings. Explanations for its maintenance vary, including (a) promotion of group solidarity; (b) favoring

of frugal practices, which could lead to higher longevity; (c) relationship with cure in certain conditions (the efficacy of shamanism healing is probably related to the neurophysiology of altered states, such as eliciting of endogenous opioid peptides); and (d) reinforcement for stable unions in nuclear families (humans are the only pair-bonded primate with significant paternal investment that lives in large multimale groups; religion would disfavor cuckoldry). Still, many questions remain; for instance, in modern societies, why religious beliefs and practices show stability in the United States, while they are in significant decline in Western Europe? [70].

Conclusions

It is therefore clear that the parallel and interacting biological and sociocultural evolutions have provided a very special characteristic to our species. In what sense, however, will our prodigious sociotechnological development affect the biology of our species? Pessimists adhere to the "survival of the unfit" principle due, for instance, to the progress in medical practices. Yet, modern medicine in all probability could not compensate for millions of genetic, regulatory, physiological, neurological, and anatomical functions present in our body [71]. Other threats may include self-enhancing artificial intelligence of an ill-conceived type, human-made pathogens, and the unforeseeable consequences of the ability to manipulate the genome, molecular structures, and the matter of life itself [72]. These powers can be used for bad or good use, and the optimistic view is that good sense will prevail.

Review questions and exercises

1 Give your own definition of culture.

2 Are evolutionary changes in our genome a cause or a consequence of cultural innovation?

3 In what way sexual behavior and paternity beliefs would influence our species genetic variability?

4 What is the future of the present dazzling language diversity? The number of languages would stabilize, increase, or decrease?

5 How would you explain the strange relationship between humans and dogs, which is leading to an unprecedented increase in the world dog population?

6 Imagine yourself alone in the midst of a luxurious tropical forest. How would you survive?

7 How would you explain the return of decorative motifs (tattooing, ear and lip piercing) that until recently were considered devices or customs restricted to "primitive" people?

8 What would be the consequence of the notion that there is no free will for the judiciary system?

9 Are you a religious person? Give the reasons for a positive or negative answer.

References

1 Levins R, Lewontin R. 1998. In: Goodman AH, Leatherman TL (eds), *Building a New Biocultural Synthesis: Political–Economic Perspectives on Human Biology*. University of Michigan Press, Ann Arbor, MI, pp xi–xv.
2 Laland KN et al. 2010. How culture shaped the human genome: bringing genetics and the human sciences together. *Nat Rev Genet* **11**:137–148.
3 Price D. 1990. *Atlas of World Cultures*. Sage, London.
4 Davidson I, McGrew WC. 2005. Stone tools and the uniqueness of human culture. *J R Anthropol Inst* **11**:793–817.
5 Foley R, Lahr MM. 2003. On stony ground: lithic technology, human evolution, and the emergence of culture. *Evol Anthropol* **12**:109–122.
6 Henshilwood CS, Marean CW. 2003. The origin of modern human behavior: critique of the models and their test implications. *Curr Anthropol* **44**:627–651.
7 Bowles S et al. 2010. The emergence and persistence of inequality in premodern societies. *Curr Anthropol* **51**:7–126.
8 D'Alpoim Guedes J et al. 2013. Is poverty in our genes? *Curr Anthropol* **54**:71–79.
9 Fisher SE, Ridley M. 2013. Culture, genes, and the human revolution. *Science* **340**:929–930.
10 Derex M et al. 2013. Experimental evidence for the influence of group size on cultural complexity. *Nature* **503**:389–391.
11 Henshilwood CS, Dubreil B. 2011. The Still Bay and Howiesons Poort, 77–59 ka: symbolic material culture and the evolution of the mind during the African Middle Stone Age. *Curr Anthropol* **52**:361–400.
12 Andersson C et al. 2014. An evolutionary developmental approach to cultural evolution. *Curr Anthropol* **55**:154–174.

13 Hill KR et al. 2011. Co-residence patterns in hunter-gatherer societies show unique human social structure. *Science* **331**:1286–1289.
14 Kendal J et al. 2011. Human niche construction in interdisciplinary focus. *Philos Trans R Soc B* **366**:785–792.
15 Creanza N, Feldman MW. 2014. Complexity in models of cultural niche construction with selection and homophily. *Proc Natl Acad Sci USA* **111**(Suppl 3): 10830–10837.
16 Hünemeier T et al. 2012. Evolutionary responses to a constructed niche: ancient Mesoamericans as a model of gene–culture coevolution. *PLoS One* **7**:e38862.
17 Morgan TJH et al. 2012. The evolutionary basis of human social learning. *Proc R Soc B* **279**:653–662.
18 Wrangham R, Carmody R. 2010. Human adaptation to the control of fire. *Evol Anthropol* **19**:187–199.
19 Turchin P et al. 2013. War, space, and the evolution of Old World complex societies. *Proc Natl Acad Sci USA* **110**:16384–16389.
20 Varki A et al. 2008. Explaining human uniqueness: genome interactions with environment, behavior and culture. *Nat Rev Genet* **9**:749–763.
21 Mace R. 2000. Evolutionary ecology of human life history. *Anim Behav* **59**:1–10.
22 Pettay JC et al. 2005. Heritability and genetic constraints of life-history trait evolution in preindustrial humans. *Proc Natl Acad Sci USA* **102**:2838–2843.
23 Aoki K et al. 2005. The emergence of social learning in a temporally changing environment: a theoretical model. *Curr Anthropol* **46**:334–340.
24 Borgerhoff Mulder M et al. 2006. Cultural macroevolution and the transmission of traits. *Evol Anthropol* **15**:52–64.
25 Danchin E et al. 2004. Public information: from nosy neighbors to cultural evolution. *Science* **305**:487–491.
26 Hewlett BS et al. 2002. Semes and genes in Africa. *Curr Anthropol* **43**:313–321.
27 Bowles S et al. 2010. The emergence and persistence of inequality in premodern societies. *Curr Anthropol* **51**:7–17.
28 Hauser MD et al. 2002. The faculty of language: what is it, who has it, and how did it evolve? *Science* **298**: 1569–1579.
29 Kirby S et al. 2007. Innateness and culture in the evolution of language. *Proc Natl Acad Sci USA* **104**:5241–5245.
30 Yang C. 2013. Ontogeny and phylogeny of language. *Proc Natl Acad Sci USA* **110**:6324–6327.
31 Dennett DC. 1996. *Kinds of Minds*. Basic Books, New York.
32 Locke JL. 2001. Rank and relationships in the evolution of spoken language. *J R Anthropol Inst* **7**:37–50.
33 Beaver T, Montalbetti M. 2002. Noam's ark. *Science* **298**:1565–1566.
34 Lieberman P. 2007. The evolution of human speech: its anatomical and neural bases. *Curr Anthropol* **48**:39–66.
35 Pennisi E. 2004. Speaking in tongues. *Science* **303**:1321–1323.
36 Graddol D. 2004. The future of language. *Science* **303**:1329–1331.
37 Steele J et al. 2010. Evolutionary approaches to cultural and linguistic diversity. *Philos Trans R Soc B* **365**:3781–3923.
38 Atkinson QD. 2011. Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* **332**:346–349.
39 Hunley K et al. 2011. Rejection of a serial founder effects model of genetic and linguistic co-evolution. *Proc R Soc B* **279**:2281–2288.
40 Forster P, Renfrew C. 2011. Mother tongue and Y chromosomes. *Science* **333**:1390–1391.
41 Price TD, Bar-Yosef O. 2011. The Origins of Agriculture: New Data, New Ideas. *Curr Anthropol* **52**(Suppl. 4): S161–S510.
42 Larson G et al. 2014. The modern view of domestication: special feature. *Proc Natl Acad Sci USA* **111**:6139–6197.
43 Meyer RS, Purugganan MD. 2013. Evolution of crop species: genetics of domestication and diversification. *Nat Rev Genet* **14**:840–852.
44 Bowles S. 2011. Cultivation of cereals by first farmers was not more productive than foraging. *Proc Natl Acad Sci USA* **108**:4760–4765.
45 Swanson-Wagner R et al. 2012. Reshaping of the maize transcriptome by domestication. *Proc Natl Acad Sci USA* **109**:11878–11883.
46 Perrier X et al. 2011. Multidisciplinary perspectives on banana (*Musa* spp.) domestication. *Proc Natl Acad Sci USA* **108**:11311–11318.
47 Myles S et al. 2011. Genetic structure and domestication history of the grape. *Proc Natl Acad Sci USA* **108**: 3530–3535.
48 Wayne RK, von Holdt BM. 2012. Evolutionary genomics of dog domestication. *Mamm Genome* **23**:3–18.
49 Larson G et al. 2012. Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc Natl Acad Sci USA* **109**:8878–8883.
50 Freedman AH et al. 2014. Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet* **10**:e1004016.
51 Achilli A et al. 2012. Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proc Natl Acad Sci USA* **109**:2449–2454.
52 Makarewicz C, Tuross N. 2012. Finding fodder and tracking transhumance: isotopic detection of goat domestication process in the Near East. *Curr Anthropol* **53**:495–505.
53 Bianchi NO. 2014. *Genética y Arte*. Asociación Argentina para El Progreso de las Ciencias (e-book, http://www.aargentinapciencias.org/).
54 Henshilwood CS et al. 2011. A 100,000-year-old ochre-processing workshop at Blombos Cave, South Africa. *Science* **334**:219–222.
55 Morriss-Kay GM. 2010. The evolution of human artistic creativity. *J Anat* **216**:158–176.
56 Zilhão J et al. 2010. Symbolic use of marine shells and mineral pigments by Iberian Neandertals. *Proc Natl Acad Sci USA* **107**:1023–1028.

57 d'Errico F et al. 2009. Additional evidence on the use of personal ornaments in the Middle Paleolithic of North Africa. *Proc Natl Acad Sci USA* **106**:16051–16056.

58 Rogers DS, Ehrlich PR. 2008. Natural selection and cultural rates of change. *Proc Natl Acad Sci USA* **105**: 3416–3420.

59 Tëmkin J, Eldredge N. 2007. Phylogenetics and material cultural evolution. *Curr Anthropol* **48**:146–153.

60 MacCallum RM et al. 2012. Evolution of music by public choice. *Proc Natl Acad Sci USA* **109**:12081–12086.

61 Ornes S. 2014. Science and culture: hunting fractals in the music of J. S. Bach. *Proc Natl Acad Sci USA* **111**:10393.

62 Horne R. 2014. Science and culture: dangerous doilies. *Proc Natl Acad Sci USA* **111**:11571.

63 Cashmore AR. 2010. The Lucretian swerve: the biological basis of human behavior and the criminal justice system. *Proc Natl Acad Sci USA* **107**:4499–4504.

64 Heschl A. 1998. *The Intelligent Genome: On the Origin of Human Mind by Mutation and Selection*. Springer, Berlin.

65 Brembs B. 2011. Towards a scientific concept of free will as biological trait: spontaneous actions and decision-making in invertebrates. *Proc R Soc B* **278**:930–939.

66 Kanai R et al. 2011. Political orientations are correlated with brain structure in young adults. *Curr Biol* **21**:1–4.

67 Ayala FJ. 2010. The difference of being human: morality. *Proc Natl Acad Sci USA* **107**(Suppl 2): 9015–9022.

68 de Waal F. 2006. *Primates and Philosophers: How Morality Evolved*. Princeton University Press, Princeton, NJ.

69 Young L et al. 2011. Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgements. *Proc Natl Acad Sci USA* **107**:6753–6758.

70 Sosis R, Alcorta C. 2003. Signaling, solidarity, and the sacred: the evolution of religious behavior. *Evol Anthropol* **12**:264–274.

71 Powell R. 2012. The future of human evolution. *Br J Philos Sci* **63**:145–175.

72 Bednarik RG. 2013. From human past to human future. *Humanities* **2**:20–55.

CHAPTER 8
Health and disease

Fortunately, there is a machine that accurately integrates both genetic and environmental risk: the human body itself.

—*Daniel G. MacArthur and Monkol Lek* [1]

Clues to cancer cures may reside in whales.

—*Stephen C. Stearns* [2]

SUMMARY

Health can be characterized as a state of well-being, with the absence of its opposite, disease. Categories of diseases were considered, and the question of why, after thousands of years of evolution, we still get sick is considered in the framework of Darwinian medicine. Parent–offspring conflict is but one example, but pathogen and more specifically infectious organisms' evolution have also been examined. Details of the pathogenic process are then considered, such as mutagenesis, teratogenesis, and gene action. The question of the relationship between reproductive fitness and health is examined, as well as inbreeding. Violence can be viewed as a distinctive pathology, and aspects of it are individual and intergroup aggression. Cancer was singled out as one of our main medical problems, but as human life is extended in time, degenerative diseases assume increasing importance. Changes in lifestyles are also important, but how can we reach the desired target of personalized medicine? Moreover, how could we deal with the toll of genetic disease? This question involves (a) detection, (b) counseling of affected persons and/or their families, and (c) finally, treatment. The challenge of genomic medicine for the 21st century involves not only technological progress, but also the universal access to the tools for a healthy life.

Hopes and reality

All of us want a fully healthy life, free from the problems of disease and aging. Unfortunately, this is not possible, and one of the main reasons for our present condition is related to evolution. Several aspects of what has been called Darwinian medicine will be considered in this chapter, as well as environmental damage to our gene pool, and information about gene action in pathological states. Our demographic structure and reproductive life are also important in the consideration of these questions, especially when degenerative diseases are examined. Violence is one extreme form of pathological behavior and merits discussion in evolutionary terms. Of special importance, of course, are improved methods of prevention, diagnosis, and treatment of hereditary diseases. The fantastic progress in molecular and bioinformatic tools promises considerable advancements in the management of all these problems, but scientific progress alone is not enough. What is needed are policies and implementation of social measures leading to a universal access to a healthier life.

Concept of health and methods of study

Health can be defined as a state of well-being, free of physical disease or pain. Conversely, disease is a condition characterized by discomfort, infirmity, illness, or other unsound condition.

Of course, the word disease designates a complex of different traits, for which a classification is given in Box 8.1. For convenience, we can initially separate

Genomes, Evolution, and Culture: Past, Present, and Future of Humankind, First Edition.
Rene J. Herrera, Ralph Garcia-Bertrand, and Francisco M. Salzano.
© 2016 John Wiley & Sons, Ltd. Published 2016 by John Wiley & Sons, Ltd.

Box 8.1 Disease classification.

1. Infectious
Caused by the entrance, growth, and multiplication of micro- or macroorganisms. The condition may or may not be contagious, in the latter case some special method of transmission or inoculation of the agent being required.

2. Noninfectious, genetic
A complex of conditions that can be classified as follows:

2.1. Chromosome abnormalities
Identified by classical microscopic techniques.

2.1.1. Numerical changes

2.1.1.1. Euploid
Variation in an exact chromosome multiple (polyploidy, e.g., triploidy, 3*N*).

2.1.1.2. Aneuploid
Extra or missing isolated chromosomes ($2N+1$, $2N-1$, etc.).

2.1.2. Structural changes

2.1.2.1. Deletions
Absence of a given chromosome region.

2.1.2.2. Duplications
Duplication of a given chromosome region.

2.1.2.3. Translocations
Transference of a region between chromosomes.

2.1.2.4. Inversions
Two-break event with reunion in an inverted way.

2.1.2.5. Isochromosomes
Exact duplication of a whole arm due to centromeric misdivision.

2.1.2.6. Ring chromosomes
Two-break event in a chromosome, with union of the broken ends.

2.2. Genomic changes

2.2.1. Variable number of tandem repeats
Abnormal numbers due to replication errors. A special case are the dynamic mutations; instability in a region may lead to intergeneration unstable repeat numbers.

2.2.2. Imprinting
Certain genes are marked (imprinted) with their parental origin; this is an example of an epigenetic mechanism, a heritable change in gene expression that occurs independently of a DNA modification, which may lead to diverse clinical symptoms.

2.3. Single-gene defects
Due to mutations in a given locus. Different mutations in a locus may lead to similar clinical pictures.

2.4. Multifactorial
Traits that are determined by two or more factors, often multiple genes interacting with each other or with the environment.

Sources: Strachan and Read 2011 [3]; Rosenberg and Rosenberg 2012 [4].

two blocks of characteristics that describe disease: (a) infectious diseases or (b) noninfectious, genetic diseases. With regard to genetic diseases, four main categories exist: (a) chromosome abnormalities, identified by classical cytogenetic techniques; (b) genomic changes, involving variation in the number of tandem repeats, or in the way they have been inherited (with different imprinting marks); (c) single-gene defects, which are due to mutations in a given locus; and (d) multifactorial disorders, a set of conditions determined by two or more factors, be they multiple genes or a combination of genetic and environmental agents. The process by which these genetic changes occur and the results of these types of abnormalities are further discussed in Chapters 4 and 5.

Genetic diseases can presently be studied by modern methods, and the promises and problems related to advances in genomic medicine were aptly reviewed by

MacArthur and Lek [1]. They examined in detail eight sequencing technologies either commercially available or recently announced, noting that they differ along four primary axes: (a) cost per base; (b) sequencing throughput per hour; (c) read length; and (d) accuracy. However, of course, advances in DNA sequencing approaches alone are not enough. Converting raw genomic data into clinically actionable information is a big challenge. Some of the problems are the accuracy of the massive amount of data generated, avoiding false positives associated with susceptibility to diseases; the health care situation derived from providers and payers; protection of patient privacy; and the need for the adequate collection of medical data and history of life events. Anyway, genomic approaches are already transforming medicine, especially in the areas of cancer diagnosis and treatment, genetic diagnosis of rare diseases, and prevention of adverse drug reactions [1].

An example of a recent investigation included exome sequencing of 70 genes from 2204 African Americans and 4313 European Americans, enrolled in the National Heart, Lung, and Blood Institute Exome Sequencing Project. Of the 70 genes assessed, 39 genes were related to conditions of newborn screening, 17 genes were associated with age-related macular degeneration, and 14 genes influenced drug response. From these 70 genes, a total of 10,789 variants were identified. After appropriate filtering, 399 validated pathogenic variants were pinpointed. The mean number of risk alleles per person was 15.3. Although this is useful information, one of the problems with this technology is the need to develop guidelines and policies related to the return of data to the research subjects, recognizing that they may or may not want to receive this kind of information, especially in relation to conditions for which there is no prevention or cure [5].

Darwinian medicine

Why is it that in an organism such as humans, which is the product of billions of years of the action of natural selection, disease has not been eradicated? The answer can only be given in an evolutionary context, which is the central focus of evolutionary or Darwinian medicine. Box 8.2 lists four messages, three themes, and three insights that illustrate well the perspectives generated by this area of research. First, natural selection is not natural perfection; its result may lead to compromises that affect health. In addition, we live in an environment that is very different from individuals who existed in our early evolutionary history and, therefore, misadaptations are not surprising; finally, host–parasite relationships can be related to an arms

Box 8.2 Perspectives generated by evolutionary medicine.

1. Messages
- **1.1.** Organisms are formed by compromises shaped by natural selection to maximize reproduction, not health.
- **1.2.** Many diseases occur due to the mismatch of our bodies to modern environments.
- **1.3.** Pathogenic agents evolve much faster than humans; therefore, infection is unavoidable.
- **1.4.** Common hereditary diseases are generally due to the interaction of multiple, genetic and environmental factors. This fact leads to a complex pathogenesis and difficult cures.

2. Themes
- **2.1.** Pathogens rapidly develop resistance to antibiotics, while cancers quickly evolve resistance to chemotherapy.
- **2.2.** Pathogenic agents develop strategies to circumvent host defenses, and degrees of virulence are shaped by natural selection to maximize transmission.
- **2.3.** Human genetic variability that increases disease resistance may have costs, and those that increase vulnerability, benefits.

3. Insights
- **3.1.** Humans coevolved with a normal community of symbiotic bacteria and parasitic worms. With their elimination by public health measures or antibiotics, our immune system can react in an unfavorable way, including developing autoimmune diseases.
- **3.2.** Imperfect vaccines, which do not completely eradicate the pathogen, could lead to an increase in pathogen virulence.
- **3.3.** Conflicts of interest among relatives may lead to mental diseases.

Source: Stearns et al. 2010 [6].

race, in which changes in one organism almost immediately evoke a correspondingly adaptive response in the other. Further sources of situations leading to health problems are given in Box 8.2.

Parent–offspring conflict

Parent–offspring conflicts are instances where there can be benefit to one and detriment to the other. This can have evolutionary consequences. The theory of inclusive fitness postulates that the adaptive coefficient of a given allele should be measured not only by its carrier's own fitness, but also by his/her effects on the fitness of his/her relatives. These relationships can lead to diverse evolutionary fates, and to different pathological conditions.

An instructive example is given by preeclampsia, clinically defined as the combination of pregnancy-induced hypertension and proteinuria. A proposal was made [7] that preeclampsia is maladaptive for mothers but adaptive for fetuses, since one of its main results is the increase of maternal blood flow to the placenta, leading to a higher supply of nutrients to the fetus. The longer the gestation is prolonged, the less are the risks to the fetus of being delivered prematurely, but the greater the risk for the mother, due to necrosis and hemorrhages.

Father–mother conflicts, when there are different evolutionary interests between the two, are most easily detected in cases of imprinting disorders. Imprinting involves the methylation of DNA by one parent and not the other. Both parents are capable of imprinting certain chromosome regions but one imprint (methylation pattern) is passed on paternally and another different imprint (methylation pattern) is passed on maternally. A classical example of a maternal–paternal conflict of imprinting concerns the differing effects of deletion or duplication of a single imprinted gene on chromosome 15. When the gene is expressed without the normal paternal inhibition (methylation pattern) (Prader–Willi syndrome), the mother's interests are expressed without restraint, the child feeds poorly, and is easy to care for; when the opposite pattern occurs, the gene is expressed without the normal maternal inhibition (Angelman syndrome), the child wants to suckle frequently, and is difficult to care for. The two types of patients also develop different mental disorders in adulthood [7].

Pathogen history

Infectious pathogens are among the strongest selective factors influencing human populations, migrations, and cultural practices. Colonization of new environments, increased population density, contact with disease vectors (domesticated animals), or those coexisting with us (rodents, sparrows) are among some of the agents responsible for differences associated with host–pathogen relationships. An individual's susceptibility to infectious disease is clearly influenced by the genomes of both members of this dialectical system. There are indications that agriculture influenced, in a negative way, the health of many of their practitioners, and modern medicine, while basically improving public health, can also lead to undesirable consequences, for example, the decreased gut microbiome diversity in residents of developed countries, which may negatively influence mucosal immune responses [8].

Harper and Armelagos [9] established a list of pathogens that, according to genetic evidence, infected humans prior to (Box 8.3) or after (Box 8.4) the establishment of agriculture. Those that presumably infected humans before agriculture included five viruses, four bacteria, and four parasites (Box 8.3). Herpes simplex viruses seem to be the oldest of these organisms, with subtypes having diverged millions of years ago. As for bacteria, *Mycobacterium tuberculosis* is estimated to have emerged in East Africa around 40 million years ago, and to have spread around the world with ancient human migrations. Lice (*Pediculus humanus*) are obligate macroparasites; head and body lice are probably not genetically distinct, and their coexistence with humans is probably connected with clothing, likely invented some 100,000 years ago.

Box 8.4 lists five viruses, two bacteria, and two parasites that did not regularly infect humans until the advent of agriculture, according to genetic data. Smallpox resulted from agricultural practices and the close association with animals. Although these associations continue to exist, modern medicine has virtually eradicated it. Only a century ago this virus was responsible for mortality rates of up 30%, and the disease was described thousands of years ago in historical records from China, India, and Egypt. Today, it has been almost completely eradicated. Because the smallpox virus DNA genome is highly conserved, with an extremely low mutation rate, simple genetic makeup, and reliance on humans as its

Box 8.3 Pathogens that infected humans prior to agriculture according to genetic evidence.

1. Viruses
- **1.1.** Epstein–Barr virus (causes infectious mononucleosis, associated with some types of cancer).
- **1.2.** Hepatitis G virus.
- **1.3.** Herpes simplex viruses 1 and 2.
- **1.4.** Human papillomavirus (causes genital warts and cervical cancer).
- **1.5.** JC virus (causes progressive multifocal leukoencephalopathy in immunosuppressed individuals).

2. Bacteria
- **2.1.** *Bordetella pertussis/B. bronchiseptica* (cause whooping cough).
- **2.2.** *Borrelia burgdorferi* (causes Lyme disease).
- **2.3.** *Helicobacter pylori* (causes gastric ulcer).
- **2.4.** *Mycobacterium tuberculosis*.

3. Parasites
- **3.1.** *Pediculus humanus* (lice).
- **3.2.** *Schistosoma mansoni* (causes schistosomiasis).
- **3.3.** *Taenia saginata* and *Taenia solium* (tapeworms).
- **3.4.** *Toxoplasma gondii* (causes toxoplasmosis).

Source: Harper and Armelagos 2013 [9].

only host, these characteristics probably facilitated its eradication [8].

Not listed by Harper and Armelagos [9], leprosy (caused by *Mycobacterium leprae*) and cholera (caused by *Vibrio cholerae*) began infecting humans around 10,000 and 5000 years ago, respectively [8]. Leprosy was endemic in Europe with a prevalence of 10–40% until the 16th century, but afterward declined rapidly. Currently, it remains a major public health burden in India, China, and South America. The genetics of leprosy susceptibility differs between populations. Two loci (*NOD2* and *CYLD*) associated with susceptibility to this disease in Han Chinese show no evidence of positive selection (and thus association with the disease) in East Asians [8]. On the other hand, cholera is a deadly disease with historic mortality rates as high as 50%, and the disease is still common in Bangladesh and other underdeveloped countries. Host genetic factors seem to strongly influence susceptibility to this condition, especially genes that are encoding potassium channels involved in cyclic AMP-mediated chloride secretion, and those involved in nuclear factor-κB signaling [8].

Box 8.4 Pathogens that did not regularly infect humans until the advent of agriculture according to genetic evidence.

1. Viruses
- **1.1.** Hepatitis C virus.
- **1.2.** Human immunodeficiency virus (HIV-1, causes AIDS).
- **1.3.** Measles virus.
- **1.4.** Rotavirus A virus.
- **1.5.** SARS coronavirus.
- **1.6.** Smallpox virus.

2. Bacteria
- **2.1.** *Shigella sonnei* (causes shigellosis or bacillary dysentery).
- **2.2.** *Yersinia pestis* (causes bubonic plague/Black Death).

3. Parasites
- **3.1.** *Plasmodium falciparum* (causes malignant malaria).
- **3.2.** *Trichinella spiralis* (causes trichinosis).

Source: Harper and Armelagos 2013 [9].

Evolution of infectious diseases

The host–parasite interaction has many facets, which could be mainly classified as (a) virulence, (b) antibiotic resistance, (c) host resistance or tolerance, and (d) emerging diseases [2].

Virulence can be defined as the degree of host morbidity and mortality caused by a given pathogen. It is obvious that too much virulence is bad for the pathogen because it kills the host, and thus the pathogen itself.

Pathogen transmission can be vertical (passed from parent to offspring) or horizontal (passed from host to host). In the first case, if the infection allows the hosts to survive at least until they reproduce, the result would be the transformation of the pathogens into avirulent commensals. The panorama when transmission is horizontal is different. In this case, there is a virulence–transmission trade-off, with two opposing pressures: first, competition within the host's body, which selects for rapid population growth of the pathogen via use of host resources; and second, successful transmission, which requires that the host survives long enough to allow the pathogen to infect other hosts. The optimal virulence for the pathogen is then at an intermediate level, which can still cause serious harm to the host.

A modern medical technology has resulted in antibiotic resistance that has become a huge medical problem. In 2004, resistant bacteria acquired in hospitals killed more than 90,000 persons in the United States [2]. Tuberculosis, which some time ago was thought to no longer be a problem, has reemerged due to the evolution of drug resistance and the increase in susceptible hosts due to the HIV/AIDS pandemic. It is calculated that the treatment of a patient with resistant tuberculosis costs 10 times more than to treat a patient with nonresistant tuberculosis [2].

Most bacterial antibiotic resistance does not arise from *de novo* mutations occurring during treatment, but from horizontal transfer of resistant genes that evolved from bacteria/fungi relationships long before antibiotics were developed, found in the natural environment. Another source derives from the commensal nonpathogenic bacteria that live in our microbiome. Bacterial resistance genes can move horizontally on plasmids, in viruses, and from direct uptake of DNA released from dead bacterial cells. They can combine to form gene cassettes that confer resistance to multiple antibiotics and are transferred as a unit.

The human immune system has evolved into a powerful tool for the development of resistance or tolerance to pathogens. The evolution of the innate immune system (the initial broad response to pathogen attack) and that of the adaptive immune system (the specialized system designed to attack specific pathogens) are both important to life of vertebrate hosts. However, pathogens have also evolved to evade or suppress the antibodies formed against them by the adaptive immune system. One pathogen evasion strategy is variation in its antigenic surface molecules. One mechanism is an inducible system that reacts to immune attack by increasing the mutation rate influencing multigenic families. A second evasion strategy is suppression of the host immune system. This can be achieved by co-opting the host's inhibitory receptors and inducing suppression by mimicking host molecules.

Almost all pathogens causing emergent diseases come from animal reservoirs, the majority being viruses, mostly RNA viruses. Examples are Ebola, probably resident in bats and transmitted from infected primates killed for meat consumption; H1N1 influenza, acquired from infected pigs; and HIV, a member of a large group of simian immunodeficiency viruses, acquired when chimpanzees were killed for food [2].

DNA damage, mutagenesis, and teratogenesis

We are constantly being exposed to external noxious agents that can damage DNA, and another major threat is endogenous chemical attacks and errors arising during the normal functioning of DNA. Box 8.5 lists three sources of external insults and five sources of internal damage. Ionizing radiation, ultraviolet radiation, and environmental chemicals are the external sources, while internal chemical events of different types can also affect DNA. Two main consequences of this damage can result. The first is cell death, and the second is germinal or somatic mutation. The effects of cell death vary according to the place where the event is taking place, while germinal mutation can lead to hereditary diseases, and somatic mutation can lead to conditions such as cancer.

In response to these various forms of damage, cells developed a whole range of repair mechanisms. A classification of them is given in Box 8.6. Different mechanisms correct for different types of lesions. In single-strand breaks, base excision repair, nucleotide excision repair, and direct reversal of the DNA damage are employed; in the case of double-strand breaks, homologous recombination and nonhomologous end joining are used. In addition, it is necessary to correct mismatches from replication errors. The importance of efficient DNA repair systems (see Chapter 4) is documented by the approximately 130 human genes participating in DNA repair, and by the severe diseases that affect persons with deficient repair systems. Clinically, patients with

Box 8.5 Agents that cause DNA damage.

1. External
  1.1. Ionizing radiation: gamma and X-rays can cause single-strand or double-strand breaks in the sugar–phosphate structure.
  1.2. Ultraviolet radiation: major source is from sunlight; causes cross-linking between adjacent pyrimidines leading to pyrimidine dimers.
  1.3. Environmental chemicals: hydrocarbons, aflatoxins, substances used in cancer chemotherapy; alkylating agents can transfer a methyl or other alkyl group onto DNA bases inducing cross-linking between bases within a strand or between different strands.
2. Internal
  2.1. Depurination: due to spontaneous hydrolysis of the base–sugar link.
  2.2. Deamination: producing uracil or hypoxanthine.
  2.3. Reactive oxygen: superoxide anions and related molecules are generated as a by-product of oxidative metabolism in mitochondria.
  2.4. Nonenzymatic methylation: the methylated adenine or guanine bases distort the double helix and interfere with vital DNA–protein interactions.
  2.5. Other types of damage during normal DNA metabolism: errors in replication or recombination can induce strand breaks in the DNA.

Source: Strachan and Read 2011 [3].

defective repair machinery may suffer symptoms that include hypersensitivity to sunlight, neurological and/or skeletal problems, anemia, and especially a high incidence of various cancers [3].

When the DNA damage affects developmental pathways, morphological abnormalities that are due to disruptions of the general processes of differentiation, pattern formation, and morphogenesis occur. The conditions appearing at birth are globally referred to as congenital malformations, and the agents conditioning them are labeled as *teratogenic*. Chemicals that can induce such effects during pregnancy include alcohol, certain antibiotics, thalidomide, retinoic acid, and cocaine. In addition to exogenous chemicals, mutations influencing developmental pathways can also lead to these drastic effects. These mutations have been found in genes

Box 8.6 DNA repair mechanisms in humans.

1. Single-strand breaks
  1.1. Base excision repair: glycosylase enzymes remove abnormal bases by breaking the sugar–base bond. Afterward, an endonuclease and a phosphodiesterase cut the sugar–phosphate structure at the position of the missing base and remove the sugar–phosphate residue. The gap is filled by resynthesis with a DNA polymerase, and the remaining nick is sealed by DNA ligase III.
  1.2. Nucleotide excision repair: removes thymine dimers and large chemical adducts. The sugar–phosphate backbone is cleaved at the site of the damage, and exonucleases remove a large stretch of the surrounding DNA. As in base excision repair, the gap is filled by resynthesis and sealed by DNA ligase.
  1.3. Direct reversal of the DNA damage: three main genes have been implicated in this mechanism. The best characterized encodes a methyltransferase that remove methyl groups from guanines that have been incorrectly methylated.
2. Double-strand breaks
  2.1. Homologous recombination: a single strand from the homologous chromosome invades the damaged DNA and acts as a template for accurate repair.
  2.2. Nonhomologous end joining: large multiprotein complexes are assembled at broken ends of DNA molecules and DNA ligases rejoin the broken ends regardless of their sequence.
3. Correcting mismatches from replication errors: the mechanisms involve at least five proteins.

Source: Strachan and Read 2011 [3].

related to transcription factors, signaling pathways, and components of the cell or extracellular matrix [3]. Induced and spontaneous mutations have been considered in Chapter 4. They are mostly subjected to the influence of negative selection.

What is better, more or less gene product?

The answer to this question is "it depends." In terms of pathology, mutations can produce either (a) loss of function, when the product formed has reduced or no function; or (b) gain of function, when the product is abnormal and can have increased or new function. Most loss-of-function mutations are inherited in a recessive way. The classical examples are the inborn errors of metabolism. However, loss-of-function conditions may be dominant (a) when there is haploinsufficiency (i.e., 50% of the product's normal level is not sufficient for normal function); an example of haploinsufficiency is the Waardenburg syndrome type 1, caused by mutations in the *PAX3* gene; or (b) when there is loss of function due to a dominant-negative effect; that is, the product of the mutant allele not only is nonfunctional, but also interferes with the function of the remaining normal allele, resulting in less than 50% residual function. Proteins that build multimeric structures, such as fibrillar collagens, probably due to their structures and the subtle interactions that are needed for their functioning, are particularly vulnerable to these effects.

Gain-of-function mutations, as the name implies, are those in which a mutation causes a gene to acquire new or enhanced activity. In gain-of-function mutations, the presence of a normal allele does not prevent the mutant allele from behaving abnormally. As a result, gain-of-function mutations are almost always dominant. Generally, gain-of-function mutations affect the way in which a gene or its product reacts to regulatory signals. The gene may be expressed at the wrong time, in the wrong tissue, at the wrong level, or in response to the wrong signal. Box 8.7 provides some examples. Six different types of malfunction are listed, resulting in a wide array of defects, such as sex reversal, a lethal bleeding disease, paramyotonia, or Huntington disease. Two of the mutations listed are only observed somatically, the constitutional forms (occurring in the whole body) probably being lethal.

In some cases, loss-of-function and gain-of-function mutations may occur in the same gene. Five examples are given in Box 8.8. The loss-of-function mutation in *PAX3*, mentioned above, leads to Waardenburg syndrome type 1, but a gain-of-function mutation in the same gene leads to alveolar rhabdomyosarcoma, a very different condition. In the other four examples, the two types of mutations also lead to different pathological outcomes. For instance, in Xq12 a loss-of-function mutation leads to the testicular feminization syndrome, while a gain-of-function mutation leads to spinobulbar muscular dystrophy.

Both types of mutations are generally subjected to negative selection, since they lead to diseases or sterility.

Box 8.7 Characterization of some gain-of-function diseases.

Malfunction	Gene	Disease	Additional information
Overexpression	*NR0B1*	Male-to-female sex reversal	Result of gene duplications
New substrate	*PI* (Pittsburgh allele)	Lethal bleeding disease	Due to conformational changes in the alpha-1 antitrypsin gene
Inappropriate ion channel	*SCN4A*	Paramyotonia congenital of von Eulenburg	Delayed closing
Protein aggregation	*HD*	Huntington disease	Proteins with expanded polyglutamine runs form toxic aggregates
Receptor permanently on	*GNAS1*	McCune–Albright disease	Only somatic mutations observed; constitutional form probably lethal
Chimeric gene	*BCR-ABL*	Chronic myeloid leukemia	Somatic mutations only

Source: Strachan and Read 2011 [3].

Box 8.8 Examples of genes in which loss-of-function and gain-of-function mutations cause different diseases.

Gene	Chromosome location	Effect	Diseases	Symbol
PAX3	2q35	–	Waardenburg syndrome type 1	WS1
		+	Alveolar rhabdomyosarcoma	RMS2
RET	10q11.2	–	Hirschsprung disease	HSCR
		+	Multiple endocrine neoplasia type IIA	MEN2A
PMP22	17p11.2	–	Charcot–Marie–Tooth neuropathy type 1A	CMT1A
		+	Tomaculous neuropathy	HNPP
GNAS1	20q13.2	–	Albright hereditary osteodystrophy	PHP1A
		+	McCune–Albright syndrome	MAS
AR	Xq12	–	Testicular feminization syndrome	TFM
		+	Spinobulbar muscular dystrophy	SBMA

Source: Strachan and Read 2011 [3].

Genetic manipulation of animals to study health and disease

Animal models are of vital importance for the investigation of how genes function in cells, and in testing new drugs and therapies. Most experiments are performed using transgenic animals. Transgenes are usually introduced into a fertilized oocyte and in this way are integrated in the genome of the developing animal. Gene targeting can also be performed using cultivated pluripotent stem cells, which are subsequently injected in the animal's germline.

Many methods have been devised to study cellular functions in animal systems, and Box 8.9 lists five of them. A whole range of genetic manipulations is now possible, involving gene inactivation, gene insertion, chromosome engineering, and intervention at the RNA level. In addition, exposure of the animals to radiation or chemical substances is used in experimental mutagenesis.

A wide range of animals are being used in the experiments listed in Box 8.9. They can be (a) unicellular organisms such as *Escherichia coli* and *Saccharomyces cerevisiae* (about 30% of human genes known to be involved in disease have functional homologs in yeast) [3], (b) invertebrates such as *Caenorhabditis elegans* or *Drosophila melanogaster*, (c) nonmammal vertebrates (chickens, frogs (*Xenopus*), and zebrafish), (d) nonprimate mammals (sheep, pigs, rats, dogs, cats, and especially mice), and (e) nonhuman primates (chimpanzees) [3].

Vertebrate gene knockouts have largely been performed in the mouse, and the International Mouse Knockout Consortium (http://www.knockoutmouse.org) seeks to mutate all known protein-coding genes

Box 8.9 Selected methods employed to study mutant effects in human diseases using animal models.

Method	Description
1. Gene knockout	Inactivation or modification of one gene to investigate loss-of-function mutations.
2. Gene knock-in	The target gene is inactivated and transgenic sequences are expressed under the regulation of its promoter.
3. Chromosome engineering	Use of microbial recombinases to produce chromosomal deletions, inversions, and translocations.
4. Gene knockdown	Intervention at the RNA level in specific RNA transcripts.
5. Insertional mutagenesis	Specific transgenes are inserted into the genome causing gene inactivation in a random or semi-random way.

Source: Strachan and Read 2011 [3].

in the mouse and to make knockouts widely available to the scientific community.

Reproductive fitness and health

Reproductive fitness can be defined as the ability of individuals to transmit their genes to subsequent generations. This parameter is basically related to measures of fertility and mortality, key factors in any study of the evolutionary consequences of health and disease. In humans, the influence of biological (genetic) factors on these variables is easily demonstrated in relation to pathological traits, since most hereditary diseases have clear mortality and/or fertility effects. More difficult to evaluate are differences in the normal range, since parent–child correlations in family sizes are most influenced by the family environment (socioeconomic level, education, contraceptive methods) or other characteristics (marital age, ability to acquire a mate).

The Hutterites, a religious minority that originated in South Tyrol in the 16th century, and that in the 1870s migrated in significant numbers to the United States and Canada, can be considered an exception in terms of the possibility of reaching mortality and fertility genetic estimates. They practice a communal agrarian lifestyle, ensuring a similar environment for their members, and have equal access to resources such as wealth, education, and medical care. Besides, they desire large families, avoiding contraception. Kosova et al. [10] studied 450 Hutterite couples, members of a single 13-generation pedigree. Their results are summarized in Table 8.1. Three types of characteristics were considered: (a) total number of births in completed families; (b) number of births per year of marriage; and (c) age at which the wife had her last child. These measures were corrected for age, cohort effects, and length of the reproductive period when needed. There was wide variability among couples in relation to these characteristics. While the average number of births was 7.1, this number varied between 1 and 17. Similar diversity was found in relation to number of births per year of marriage (0.2–1.0), and age at which the wife had her last child (22–47). The most general result found was a significant degree of heritability for the three characteristics in males, but not in females. Different genetic factors related to reproduction in the two sexes may explain these differences. Other subtle variations in the types of autosomal and X-linked variances were observed. Be as it may, it seems that reproductive traits are amenable to genetic mapping studies, giving new clues about the factors influencing natural fertility in our species.

Table 8.1 Sample characteristics and heritabilities of reproductive fitness traits partitioned by sex among the Hutterites.

Sample characteristics	Sample size	Average	Range
1. Total number of births in completed families (CFS)[a]	353	7.1	1–17
2. Number of births per year of marriage (BR)	459	0.5	0.2–1.0
3. Age at which the wife had her last child (ALR)	353	35.1	22–47

Heritabilities[b]	h^2_A	h^2_X	H^2	Probability value
Females				
1. CFS	0.0	0.2	0.3	0.29
2. BR	0.0	0.3	0.3	0.21
3. ALR	0.2	0.1	0.2	0.14
Males				
1. CFS	0.1	0.1	0.7	0.001
2. BR	0.1	0.1	0.5	0.02
3. ALR	0.3	0.0	0.3	0.001

Source: Reference 10.

[a] Families in which the wife was >45 years of age and was not widowed before then, or the couple had not had child in >6 years.

[b] Heritability is a measure of the degree to which genetic, as opposed to environmental, factors influence a given trait. It can be classified as follows: h^2_A, narrow heritability caused by autosomal additive effects; h^2_X, narrow heritability caused by X-linked additive effects; and H^2, broad heritability.

Consanguinity

Despite many social restrictions (requirement of dispensation in the Roman Catholic Church, or the prohibition of such marriages in 31 of 50 states in the United States), currently couples related as second cousins or closer ($F \geq 0.0156$) and their progeny account for about 10.4% of the world population [11]. The concerns over consanguineous marriage due to the fact that eventual problems related to autosomal recessive disorders (for instance, an average excess mortality at first-cousin level of 3.5% [11]) may be outweighed by social benefits (more stable marital relationships, greater compatibility with in-laws, lower domestic violence, lower divorce rates, and landholdings maintenance). Therefore, cousin marriages are still quite frequent especially in African and Asian countries.

Pedigree-based estimates of consanguinity do not provide information on such marriages that occurred in distant generations, and therefore underestimate cumulative inbreeding effects. They are being replaced by high-density genome scans that estimate individual autozygosity from uninterrupted runs of homozygosity. This new method has already furnished precious information on complex diseases such as schizophrenia, bipolar disorder, or autism [11].

Violence

Violence can be viewed as a distinctive pathology. Aggression, physical or mental attacks on other persons, mediates competition for food, mating partners, or dominance hierarchies. Violence can be individual or collective, and we will focus first on individual violence. From a genetic perspective, except for a few isolated examples, aggression is a quantitative trait due to multiple segregating genes that are environmentally sensitive. General evaluations point to heritability estimates of around 50% [12]. Abnormal expression of aggressive behavior is a common consequence of traumatic brain injury (especially in the frontal lobe), neuropsychiatric diseases, alcohol and drug abuse, and neurodegenerative disorders. Box 8.10 provides a list of 11 pathological impulse conditions that can lead to violence. They can be divided into two large categories, of disordered impulses in general or of impulse control. In the first category, we can mention depression, bipolar disorder, phobias, addictions, obsessive–compulsive disorder, Tourette syndrome, schizophrenia, and delusion conditions. In the second category, we can mention antisocial personality, intermittent explosive mood, borderline personality disorder, childhood conduct affections, and attention deficit hyperactivity disorder. As the listing indicates, they are a complex mix of conditions that are influenced by both genetic and nongenetic factors.

Human anger was analyzed in detail in evolutionary terms by Sell et al. [14]. They proposed that anger is produced by a neurocognitive program engineered by natural selection to use bargaining tactics to resolve conflicts of interest in favor of the angry individual. Through two experimental studies they verified that men with enhanced abilities to inflict costs (stronger) or women who may confer benefits (attractive) have considerable bargaining power, and are more prone to anger. They concluded that the internal logic of the anger program reflects the ancestral payoffs characteristic of small-scale social world rather than rational assessments of modern payoffs in large populations.

Childhood maltreatment is a universal risk factor for antisocial behavior. Maltreatment increases the risk of later criminality by about 50%, but most maltreated children do not become delinquent or adult criminals. Why? The answer may be in the levels of the neurotransmitter-metabolizing enzyme monoamine oxidase A (MAOA). It was found, using subjects of the Dunedin Multidisciplinary Health and Development Study (which had no population stratification confounds), that maltreated children with MAOA high-level expression were less likely to develop antisocial problems than those with low levels, a good example of gene–environment interaction [15]. Other factors, of course, can also play a role.

Is it possible to ascertain the ethical behavior of a person by just looking at his/her face? A series of six studies published between 1999 and 2012 suggested that simple facial traits could be used to predict aggressive, unethical, or other types of behaviors. One trait specifically proposed is that men with wider faces relative to facial height would be more likely to develop unethical behavior mediated by a psychological sense of power. Intrigued by these results, Gómez-Valdés et al. [16] considered 4960 individuals from 94 modern human populations, ethnographical records, samples of male prisoners of the Mexico City Federal Penitentiary condemned by crimes of variable level of interpersonal aggression, and the relationship between the trait and

Box 8.10 Pathological impulse conditions that can lead to violence.

1. Diseases with disordered impulse

1.1. Depression and anxiety disorders

Depression is characterized by decreased mood, appetite, interest in sex, and energy. Bipolar disorder alternates cycles of depression and mania, with depression predominating as the person ages. Anxiety diseases include panic disorder and phobia. Suicide is a major outcome risk.

1.2. Addictions

They can be related to both impulsivity and disordered mood. Drug and nicotine dependence, as well as alcoholism, are examples.

1.3. Obsessive–compulsive disorder

Compulsion for cleaning, checking, counting, and many others.

1.4. Tourette syndrome

Disordered motor movements, such as facial and vocal tics, many times involving coprolalia.

1.5. Schizophrenia

Characterized by hallucinations and delusions.

1.6. Delusional disorder

Persons believe that there is a conspiracy directed against them.

2. Disorders of impulse control

2.1. Antisocial personality disorder

Symptoms are lying, stealing, and torturing animals, and those affected may become serial killers.

2.2. Intermittent explosive disorder

Severe impulsive behavior, which may lead to senseless murders.

2.3. Borderline personality disorder

Poorly regulated emotionality, impulsive behavior.

2.4. Childhood conduct disorder

Children lie, cheat, and steal, commit vandalism, and engage in early sexual behavior and drug use.

2.5. Attention deficit hyperactivity disorder

Characterized by inattention, for instance, forgetfulness, distractibility, and loosing things; hyperactivity, such as inability to stay seated, restlessness, and excessive talking; and impulsiveness, for example, not waiting one's turn.

Source: Goldman 2012 [13].

reproductive success. Overall, the results suggested that facial attributes are poor predictors of aggressive behavior.

One of the most perverse forms of violence is torture used by counterintelligence agents against prisoners. With the opening of the United States Central Intelligence Agency (CIA) archives to interested persons, horrifying details of these processes have been made available. CIA's 1963 Kubark Counterintelligence Interrogation Manual methodically indicated effective forms of "mind control," "brainwashing," and other procedures used to "break" prisoners or to induce confessions. These procedures would go to the extreme of interrogating prisoners about the amount of distress that they had under different types of physical torture or psychological manipulations, and Box 8.11 lists the results of a survey involving 12 physical and 7 psychological forms of torture. The list speaks for itself; no further comments are necessary.

Sexual conflict is also widespread, especially sexual coercion by males. It may include direct coercion (forced copulation, harassment, and intimidation) or indirect coercion (coercive mate guarding). In cultures of the Mediterranean and Middle East, men will beat even their own kinswomen for sexual misdemeanors to demonstrate their commitment to chastity. Spousal violence, related to fear of infidelity (the husband considers his wife as his property), is not rare, and rape could be due to a demonstration of power and dominance and/or due to the fact that the rapist could not attract sexual partners in a voluntary way. Obviously, all these types of pathological behaviors have reflections in our species' gene pool and have been considered in evolutionary analyses [19].

Box 8.11 Results of a study comparing distress ratings of 300 Yugoslavian torture victims.

Type of procedure	Distress rating (0–4)
1. Physical torture	
1.1. Palestinian hanging (hanging by the wrists tied at the back)	3.8
1.2. Suffocation/asphyxiation	3.8
1.3. Electric shock	3.7
1.4. Falanga (beating the soles of the feet)	3.6
1.5. Burning of parts of body	3.6
1.6. Forced extraction of teeth	3.6
1.7. Stretching of the body	3.5
1.8. Beating	3.5
1.9. Hanging by hands or feet	3.5
1.10. Needles under toenails or fingernails	3.4
1.11. Beating over the ears with cupped hands	3.4
1.12. Pulling/dragging/lifting by hair	3.2
2. Psychological manipulations	
2.1. Sham executions	3.7
2.2. Witnessing torture of close ones	3.6
2.3. Threats of rape	3.6
2.4. Threats against family	3.4
2.5. Witnessing torture of others	3.4
2.6. Threats of death	3.3
2.7. Fluctuation of interrogator's attitude	2.8

Sources: Price 2007, Part 1 [17]; Part 2 [18].

Jealousy, a deeply negative emotion that arises when an important relationship is threatened by a rival (Figure 8.1), probably evolved outside the mating context, as a response to competition between siblings. However, morbid jealousy develops when persons display a conviction, most often delusional, that their mates are cheating on them. It occurs in both sexes, and may lead to physical injury or homicide [20].

Moving to collective violence, there is a dialectical relationship between intrapopulation altruism (patriotism) and interpopulation aggression (war). Bowles [21,22] and Choi and Bowles [23] considered these factors in a quantitative way. Their conclusions are that genetic differences between prehistoric groups would be large enough so that lethal intergroup competition could be a significant factor influencing these relationships. Important in this regard are characteristics such as sharing food beyond the immediate family, monogamy, and other forms of reproductive leveling. Both altruism to group members and parochialism, hostility toward individuals not of one's own ethnic, racial, or other type of affiliation, should be considered in these analyses.

The evolution of lethal intergroup violence was evaluated by Kelly [24]. According to him, after a period of Paleolithic warlessness due to low population density, an appreciation of benefits of being on good terms with neighbors, and a respect for their defensive capabilities, a period of segmental forms of organization engendered the origin of war. The unit involved in combat would be adult male raiding parties, with a target at the sleeping quarters at the core of the enemy group's territory. Attackers would benefit from weaponry, surprise, and numerical superiority. The earliest archeological evidence for attacks on settlements is a Nubian cemetery dated at 12,000–14,000 years before present (YBP). War originated independently in other parts of the world at dates extending to 4000 YBP [24].

An important aspect of all this evolution is the acquired capacity, as old as 2.0–2.5 million years ago, of killing or injuring members of their own species at a distance. Australopithecines were the first animals in the history of the Earth to acquire this ability [25]. Bingham [25] is categorical: "Coercive violence exploiting the uniquely human capacity to kill remotely is apparently essential to all human social cooperation above the level of tiny kinship groups. According to the theory this will remain so, inevitably and forever."

This view is basically unethical. War is always immoral. Simpson [26] asserts that not only the official who ordains the soldier to kill the enemy, but also the soldier himself would be responsible for the death of another human being, an intrinsically evil act.

When conflict groups are in contact, psychological factors that may perpetuate the conflict are important. Bruneau et al. [27] considered this problem examining behavioral and neural responses of samples of Arab, Israeli, and (as a control) South American subjects. Arabs and Israelis reported feeling significantly less compassion for each other's pain and suffering, which did not happen in relation to the other, distant outgroup. However, brain regions that respond to others' tragedies (identified through magnetic resonance imaging) showed an

Figure 8.1 An old painting that can be used to illustrate jealousy between women. Morbid jealousy, especially among men, may lead to serious physical injury or death inflicted in their partners. (*Source*: http://www.smithsonianmag.com/history/marie-antoinette-134629573/?no-ist.) (See the Color Plates section.)

ingroup bias relative to the distant outgroup only. Further studies relating behavior judgments to neural characteristics are clearly necessary.

Cancer

Cancer is a condition in which cells divide without control. Six basic characteristics can lead to this situation, and they are listed in Box 8.12. The cells should be insensitive to anti-growth signals and should have the ability to replicate indefinitely, as well as to trigger angiogenesis, vascularize, and establish secondary tumors.

Tumors, masses of non-inflammatory cells formed by abnormal proliferation, can be classified according to their tissues of origin: (a) carcinomas, derived from epithelial cells; (b) sarcomas, derived from bone or connective tissue; and (c) leukemias and lymphomas, derived from blood cell precursors. Generally, the beginning of the process is tissue hyperplasia or benign (noninvasive) tumors, which can later develop into more progressive (invasive) stages.

It was verified that on average six or seven successive mutations are needed to convert a normal epithelial cell into an invasive carcinoma [3]. The chance that a single cell will be the recipient of seven independent mutations is vanishingly small. However, two mechanisms may be responsible for an event like this within a single cell: (a) some mutations would enhance cell proliferation, which may lead to an expanded target cell population favorable to the next mutation; and (b) certain mutations affect the stability of the whole genome, increasing the mutation rate.

Two types of genes are of key importance in the whole process of cancer development: *oncogenes* and *tumor suppressor genes*. We will first consider oncogenes, genes that promote cell proliferation. Located at specific regions of our genome, gain-of-function mutations may lead to abnormal growth. Their nonmutant versions are called proto-oncogenes, and a selected list of them is given in Box 8.13. The altered functions are of diverse types, conditioning different types of cancer. The activation process leading to malignancy may involve (a) amplification, (b) point mutation, (c) chromosomal rearrangement, and (d) translocation to a region of transcriptionally active chromatin [3].

The second major class of genes that are mutated in tumors consists of the tumor-suppressing genes. Loss of heterozygosity in nearby regions, point mutations, and methylation changes may occur within these genes, and a list of 10 rare familial cancer diseases due to these kinds of genes is presented in Box 8.14. They affect different organs and tissues, always involving, however, solid tumors.

Box 8.12 The six basic characteristics of cancer cells.

1. Independence of external growth signals.
2. Insensitivity to external anti-growth signals.
3. Ability to avoid apoptosis.
4. Ability to replicate indefinitely.
5. Ability of a mass of such cells to trigger angiogenesis and vascularize.
6. Ability to invade tissues and establish secondary tumors.

Source: Strachan and Read 2011 [3].

Recent sequencing technologies are furnishing unprecedented details of the changes present in tumor cells. For instance, a study found 63,000 changes in a tumor cell not present in the patient's normal genome [3]. It is important, however, to distinguish *driver mutations* (directly involved in the development of the tumor) from *passenger mutations* (incidental consequences of the many cell divisions and genomic instability of cancer cells).

Degenerative diseases

Aging can be characterized as a progressive loss of adult physiological functionality over time, which is reflected in a reduction of the rate of survival and fecundity. Must all organisms age? The overwhelming majority of metazoans show definitive aging, and even unicellular organisms, potentially immortal, show this phenomenon through asymmetrical cell division. In bacteria, one of the two products of a cell division, the daughter, has younger regions than the other, the mother, and those that inherit the younger regions live longer. It therefore appears that all organisms must inevitably age and die.

Why do we age? First, it is clear that aging does not have a function, and second, the answer to the question is antagonistic pleiotropy. Any mutation that sufficiently improves reproductive performance early in life, even if it increases the risk of death or infirmity later in life, will be positively selected. Thus, increased reproduction reduces life span, the result being the disposability of the soma to the benefit of the germline [2].

Humans are the longest-lived primate, and this fact has biological consequences. Table 8.2 shows that death from senescence is 10× more frequent in traditional humans than in feral chimpanzees. The degenerative diseases include, for instance, ischemic heart disease, neurodegeneration (neuronal loss, neuritic plaques, fibrillar degeneration), and cancer.

Humans may have more cancer than other species for at least three reasons: (a) we are surviving longer than

Box 8.13 Selected list of cellular proto-oncogenes in the human genome.

Function	Cellular proto-oncogene	Map location
1. Secreted growth factors		
1.1. Platelet-derived growth factor B subunit	*PDGFB*	22q13.1
2. Cell surface receptor		
2.1. Epidermal growth factor receptor	*EGFR*	7p11.2
2.2. Macrophage colony-stimulating factor receptor	*CSF1R*	5q32
3. Signal transduction component		
3.1. Receptor tyrosine kinase	*HRAS*	11p15.5
3.2. Protein tyrosine kinase	*ABL1*	9q34.1
4. DNA-binding proteins		
4.1. AP-1 transcription factor	*JUN*	1p32.1
4.2. DNA-binding transcription factor	*MYC*	8q24.21
4.3. DNA-binding transcription factor	*FOS*	14q24.3
5. Cell cycle regulators		
5.1. Cyclin D1	CCND1	11q13
5.2. Cyclin D2	CCND2	12p13
5.3. Cyclin D3	CCND3	6p21

Source: Strachan and Read 2011 [3].

Box 8.14 Rare familial cancers due to tumor suppressor gene mutations.

Disease	Gene	Map location
1. Familial adenomatous polyposis coli	*APC*	5q21
2. Lynch syndrome I	*MSH2*	2p21
3. Breast–ovarian cancer	*BRCA1*	17q21
4. Li–Fraumeni syndrome	*TP53*	17p13
5. Gorlin basal cell nevus syndrome	*PTCH1*	9q22.3
6. Ataxia telangiectasia	*ATM*	11q22.3
7. Retinoblastoma	*RB1*	13q14
8. Neurofibromatosis 1 (von Recklinghausen disease)	*NF1*	17q11.2
9. Familial melanoma	*CDKN2A*	9p21
10. Von Hippel–Lindau syndrome	*VHL*	3p25.3

Source: Strachan and Read 2011 [3].

we did in the past and now have a long postreproductive life span; (b) we have not yet adapted to new risk factors, such as tobacco smoking, alcohol, high-calorie, high-fat diets, sedentary habits, contraceptives, and pollutants; and (c) we have an unusual sexuality, characterized by continuous cycling, continuous receptivity, continuous activity, and now contraception, all of which increase the number of cell divisions, and therefore of somatic mutations [2].

Aging is more than the accumulation of diseases over time. The healthiest 50-year-old person in the world cannot sprint as fast as he/she could at age 20. This is due to declines in multiple physiological systems, influenced by many genes. There are also what has been called the "public" mechanisms of aging, shared by many organisms. The most well-known environmental factor is dietary restriction, which leads to increased longevity. In relation to single genes, two can be singled out: (a) apolipoprotein E, a blood cholesterol transporter shown to strongly influence brain health and neurodegeneration; and (b) growth hormone/insulin/insulin-like growth factor 1, mutations in which lead to up to 70% increased life span and retarded aging in mice. Although direct extension of this last result to humans is controversial, the subject merits further studies [29,30].

Table 8.2 Causes of death (%) in feral chimpanzees and traditional humans.

Causes of death	Feral chimpanzees	Traditional humans[a]
Infections	67	73
Violence/ accidents	32	17
Senescence	1	10

Source: Reference 28.

[a] Hunter-gatherers and forager-farmers with limited access to modern medicine.

The influence of dietary restriction, mentioned above, is only one aspect of the influence of diet on aging. One important factor in human evolution was the increased consumption of animal tissues. Advantages of eating a diet rich in animal protein include higher density caloric content and concentrated micronutrients; however, increased trace metals and fat ingestion can also lead to pathogenesis. Also, uncooked meat from scavenged old carcasses, consumed by early humans, should have led to increased exposure to infectious pathogens, and cooking accelerates nonenzymatic glycoxidation to form advanced glycation end products, which are diabetogenic and proatherosclerotic. Therefore, trade-offs should have to be developed to counteract these undesirable aspects with the advantages of eating meat [28].

The activation of the immune system, important during the reproductive life to eliminate infections, could also lead to undesirable effects in old age, since chronic inflammation produces mutagenic protons and reactive nitrogen species [2]. All these considerations point to the complexity of factors inherent to human evolution. We are special due to the development of culture, and unique forms of research are needed to unravel our past, understand the present, and plan for the future.

Ecogenetics, pharmacogenetics, and pharmacogenomics

Good health (and its opposite, disease) results from an interaction between a person's genotype and his/her environment. The field that deals with these relationships is called ecological genetics, or ecogenetics, and involves a wide array of conditions and situations. Thus, environmental pollution can lead to DNA damage and disease, as presented in a previous section of this chapter, and more generally lifestyles, as characterized by food energy consumption (nutrigenetics) and energy expenditure (derived from physical activity), should be considered. Particularly important is drug metabolism, since we are being exposed to an increasing number of them either due to medical treatment (10,000 medicines are currently in use or have been employed in the past) [4] or due to sheer recreation (e.g., nicotine, marijuana, and cocaine).

We begin by considering evolutionary energetics. Our survival and reproduction is a reflection of our total energy budget. Maintenance energy expenditure involves resting (basal) costs, as well as those related to daily activities of work, play, and recreation, while production energy expenditure is that associated with growth from infancy into adulthood, and offspring production afterward.

Table 8.3 presents selected information about three indices related to evolutionary energetics (TDEE: total daily energy expenditure; BMR: basal metabolic rate; and PAL: physical activity level) for some subsistence-level populations located in Africa and the Americas (Figure 8.2) that were compared with those of a sample of people living in an industrial society (the United States). The data are separated by sex due to male/female different physiques and lifestyles. Generally, males are heavier and expend more energy than females, and agriculturalists show a higher level of physical activity than hunter-gatherers or pastoralists. On the other hand, persons living in an industrialized environment are heavier and expend less energy in physical activities.

The result is a higher prevalence of obesity and chronic metabolic disorders (adult type 2 diabetes,

Table 8.3 Energy expenditure and physical activity in selected subsistence-level and industrial populations.

Population	Sex	Average weight (kg)	TDEE[a] (kcal/day)	BMR[b] (kcal/day)	PAL[c]
Hunter-gatherers					
!Kung (Botswana)	M	46.0	2319	1383	1.68
	F	41.0	1712	1099	1.56
Inuit (Canada)	M	65.0	3010	1673	1.80
	F	55.0	2350	1305	1.80
Ache (Paraguay)	M	59.6	3327	1531	2.17
	F	51.8	2626	1394	1.88
Pastoralists					
Evenki (Russia)	M	58.4	2681	1558	1.75
Agriculturalists					
Gambia	M	61.2	3848	1604	2.40
	F	50.3	2500	1236	2.03
Aymara (Bolivia)	M	54.6	2713	1355	2.00
	F	50.5	2376	1166	2.03
Quechua (Ecuador)	M	61.3	3810	1601	2.38
	F	55.7	2460	1252	1.96
Average of 10 subsistence level populations	M	58.0	3015	1525	1.98
	F	51.6	2294	1257	1.82
Average of U.S. industrial populations	M	70.1	2873	1659	1.73
	F	58.6	2234	1300	1.72

Source: Reference 32.
[a] TDEE: total daily energy expenditure.
[b] BMR: basal metabolic rate.
[c] PAL: physical activity level.

Figure 8.2 Member of an Ache community of Paraguay prepared for an event. The Ache are among the most well-studied groups that up to quite recently had a hunter-gatherer way of subsistence. (Photograph reproduced with kind permission of Kim Hill and A. Magdalena Hurtado.) (See the Color Plates section.)

hypertension, and dyslipidemia, that is, elevated serum triglycerides with reduced HDL cholesterol) in industrialized societies. In the United States, the prevalence of obesity was at alarming numbers in 2003–2004 (31% in men and 33% in women), and these frequencies are probably higher presently [2,31]. As for type 2 diabetes, Table 8.4 gives its regional prevalence in 2000, and the estimate made in 2001 of what these values would be in 2010. As shown, 151 million people were affected by the disease in 2000, and the estimates were that this number would rise to 221 million in 2010, a 46% increase. More recent data will probably disclose even higher numbers, since the trend is accelerating. The tendency for increase is especially high in South America, Africa, and Asia.

Table 8.4 Regional prevalences of type 2 diabetes in 2000 and the estimates made for 2010 (in millions of cases).

Region	2000	2010 (estimates)	% Increase
North America	14.2	17.5	23
Europe	26.5	32.0	24
Australia	1.0	1.3	33
South America	15.6	22.5	44
Africa	9.4	14.1	50
Asia	84.5	132.3	57
Total, world	151.0	221.0	46

Source: Reference 32.

The main factor responsible for these public health problems is, as indicated previously, an imbalance between energy intake and energy expenditure. The type of diet consumed by people in so-called modern societies is generally not appropriate not only in amount, but also in quality. Americans derive 33% of their dietary energy from fat, 15% from protein, and the remaining 52% from carbohydrates [31]. The carbohydrates generally come from simple sugars or refined grain products, which condition a high glycemic load, which may be responsible for insulin resistance. In addition, the fats are mostly unsaturated, leading to increased chronic disease risks.

Diets, of course, are strongly influenced by taste perception. Food preferences are determined by multiple types of sensory input, such as taste, texture, and smell. Perception of these properties depends on the stimulation of receptor proteins located on specialized cells that are located in the tongue, mucosal surfaces of the mouth, nose, and pharynx. Five modalities of taste can be detected: sweet, salt, sour, bitter, and umami (savory or the taste of monosodium glutamate), and all of them are regulated by genes, pointing to the importance of genetic factors in these preferences [31]. The bitter taste receptor genes (*TAS2R*s) were investigated in a whole range of mammals. Gene duplications were particularly evident in the ancestral branches of the anthropoids, suggesting adaptive evolution. The human genome contains five segmental duplications, and our species is polymorphic for a deletion that involves *TASR2R45* and *TAS2R68P* [33].

Turning now to questions of pharmacogenomics, the core for any attempt at personalized medicine should be personalized prescription. Patients have different susceptibilities to drugs, and prescriptions considering just age

may lead either to absence of the desired effect or to adverse reactions. It is estimated that these reactions are responsible for not less than 100,000 deaths per year in the United States, and for 1 in 15 of all hospital admissions in the United Kingdom [3].

Some of these adverse reactions are due to environmental causes, such as disease, personal lifestyle, or interaction with other drugs. However, many differences are due to genetics, and *pharmacogenetics* is the study of the roles of specific genes in these effects, whereas *pharmacogenomics* uses genome-wide methods for the same purpose.

Studies on the absorption, activation, metabolism, and excretion of drugs are labeled *pharmacokinetics*. *Pharmacodynamics*, on the other hand, focus on the actual target response. Pharmacokinetics considers what the body does to a drug, while pharmacodynamics examines what the drug does to the body. Genetic factors are important for the two processes.

Four stages can be envisaged when considering how genetic variability can influence a patient's response to a drug: (a) absorption, that is, transport of a drug to the bloodstream; (b) activation, since some drugs are given in an inactive form that must be converted into the active conformation; (c) target response, the local concentration of the drug is important; and (d) catabolism and excretion, how the drug is broken down and disposed.

Two phases can also be visualized in the processes described above: (a) phase 1, characterized by oxidation, hydroxylation, and hydrolysis, to produce a polar compound that is water soluble; and (b) phase 2, when the molecule is conjugated with acetyl, glucuronosyl, or glutathionyl groups in order to be excreted. The P450 cytochromes (coded by about 60 genes) are the main complexes responsible for the phase 1 metabolism of drugs, while the conjugation reactions of phase 2 are mainly performed by acetylators (NAT1 and NAT2 enzymes), a glucuronosyltransferase (UGT1A1), glutathione *S*-transferases (GSTM1 and GSTT1), and a thiopurine methyltransferase (TPMT). Other systems have already been identified as responsible for variations in the drug targets (receptors, enzymes, and signal transduction substances) [3].

Detection of genetic diseases

Genetic diseases are much more common than was previously thought. Monofactorial, Mendelian disorders may be individually rare, but when all of them are considered the numbers become significant. In addition, a large number of multifactorial diseases, in which there is an interaction between genetic and environmental conditioning factors, are among the most frequent of human ailments. It is of prime importance, therefore, to adequately detect the frequencies of these traits at the population level. Box 8.15 presents some selected aspects related to this detection. The first part of Box 8.15 relates to the purposes of such studies. Three reasons can be listed: (a) to provide information to people at risk; (b) to treat affected individuals; and (c) to answer scientific questions.

When should detection testing be done? The answer is, throughout the life cycle. The sooner the detection, the better for the prevention of certain hereditary conditions. When the disease is of late onset, however, diagnoses have to be postponed. However, they are important for relatives of those affected, and proper measures can lead to avoidance of the disease among them.

Finally, intervention can take many forms, and its modalities are detailed in the next sections.

Genetic counseling

One of the most traditional forms of interaction between carriers or persons with a family history of genetic diseases and geneticists is genetic counseling. The latter is a communication process that at the beginning was performed basically by professional geneticists, but presently involves multidisciplinary teams located at clinical institutions and, in the United States, persons who specialized in the area, *genetic counselors*.

Who seeks genetic counseling? Box 8.16 furnishes a list of 10 reasons for seeking such advice. As indicated, being a carrier or having a family history of a monogenic disease was traditionally the most common type of reason. However, with the advances in genetics and molecular biology and the ways of dealing with such afflictions, many other situations have arisen. These situations involve children with multiple congenital malformations, advanced maternal age, recurrent abortions, exposure to noxious agents, and other subjects. A recent development involved the creation of online services called direct-to-consumer genetic testing (DCGT) that provide genome information direct to any interested person. They are given by commercial companies that

Box 8.15 Major aspects related to the detection of genetic diseases.

1. Purposes
- **1.1.** To provide information to people at risk. Sometimes, these risks have a high degree of certainty, as in Mendelian conditions. In other instances, however, risk assessment depends on empirical information previously obtained in affected families.
- **1.2.** To treat affected individuals. For many conditions, detection can lead to valuable, even life-saving treatment (examples: phenylketonuria, with the establishment of a low-phenylalanine diet; bipolar disorder, with the use of lithium and other mood-stabilizing drugs).
- **1.3.** To answer scientific questions. Genotype–phenotype relationships can furnish knowledge of both theoretical and practical importance.

2. Timing of detection
- **2.1.** Premarital
- **2.2.** Preconceptual
- **2.3.** Preimplantation
- **2.4.** Prenatal
- **2.5.** Neonatal
- **2.6.** Childhood and adulthood

3. Intervention
- **3.1.** Genetic counseling
- **3.2.** Management and treatment
  - **3.2.1.** Avoidance of noxious agents
  - **3.2.2.** Dietary restriction
  - **3.2.3.** Replacement of deficient substances
  - **3.2.4.** Cofactor supplementation
  - **3.2.5.** Drug administration
  - **3.2.6.** Surgical excision or correction
  - **3.2.7.** Organ transplantation
  - **3.2.8.** Stem cell therapy
  - **3.2.9.** Gene therapy

Source: Rosenberg and Rosenberg 2012 [4].

Box 8.16 Reasons to seek genetic counseling.

1. Being a carrier or having a family history with a Mendelian (monofactorial) disease.
2. Child with multiple congenital malformations or with cognitive deficits.
3. Advanced maternal age.
4. Recurrent abortions.
5. Family history of cancer.
6. Exposure to mutagenic or teratogenic agents.
7. Inbreeding (marriage to a relative).
8. Planning for testing related to late-onset diseases.
9. Follow-up to abnormal prenatal or neonatal test result.
10. Interpretation of direct-to-consumer results.

Source: Rosenberg and Rosenberg 2012 [4].

emphasize that they are assessing risk or susceptibility to a given condition, not diagnoses. However, this unregulated activity may inflict more harm than help to these consumers and recently the companies doing DCGT have come under scrutiny by the FDA [34].

Why DCGT services may be harmful? The answer is that genetic counseling is a complex, delicate process, involving many facets, detailed in Box 8.17. As indicated there, the process includes the understanding of the diagnosis, the need to be fully aware of the role of heredity in the condition, and the establishment of the options most consistent with the person's ethical and religious views. Adjustment to the situation is generally not easy, especially when there is no cure or remedy, and this is the reason why most genetic counseling teams include psychological support.

Box 8.17 Aims of genetic counseling.

Genetic counseling is a communication process whose broad aim is to provide information to individuals and families with or at risk for genetic diseases. Its aims are to help those seeking counseling to
1. understand the facts related to diagnosis, likely outcome, and medical management;
2. appreciate the role of heredity in the condition, especially recurrence risks to family members;
3. furnish the options available in relation to the recurrence risks;
4. choose a course of action consistent with their family principles, as well as their ethical and religious views; and
5. make an ideal adjustment to the disease.

Source: Rosenberg and Rosenberg 2012 [4].

A basic question is whether genetic counseling should be directive or nondirective. Should the provider merely present the facts and options, or be more assertive? Originally, genetic counselors were told to be non-directive in the United States; however, over the past decade, most have changed their policy to be more directive. In many cases, the answer will be different according to the consultant level of knowledge, or the patients' expectation or need for specialized advice.

Wertz and Fletcher [35] conducted a cross-cultural study of 1096 geneticists living in 19 nations located in Europe, Asia, the Americas, and Australia on issues and attitudes related to genetic counseling. They were asked what type of action they would perform in relation to 14 specific cases, which covered five types of ethical problems: (a) confidentiality versus duties to third parties (3 cases); (b) full disclosure of sensitive information to consultants (2 cases); (c) full disclosure of laboratory test results (3 cases); (d) indications for prenatal diagnosis (3 cases); and (e) directive/nondirective counseling (3 cases).

Overall, there was less degree of international consensus than was previously anticipated. Cultural differences in ethics still seemed substantial, at least at the time the survey was conducted (25 years ago). Some of the differences in relation to the total were the following: (a) most respondents were willing to discuss donor egg and surrogate mothering in a nondirective context; (b) most would do prenatal diagnosis for maternal anxiety only; (c) in some nations, significant minorities would be willing to perform prenatal diagnosis for sex selection unrelated to X-linked diseases; and (d) about half of the respondents by nation would approve mandatory genetic screening in the workplace. These findings emphasize again the assertion that genetic counseling involves complex and difficult questions that cannot be easily dismissed.

Treatment

The time is over when a genetic condition was viewed as not amenable to treatment or cure. Today, there are many possibilities for the management or treatment of genetic conditions, as indicated in Box 8.15. First, avoidance of a noxious agent is the most straightforward. For instance, xeroderma pigmentosum patients may prevent skin cancers by simply avoiding sunlight. Second, marked dietary restriction of a substance is a classical approach to treat many inborn errors of metabolism (example: diets low in phenylalanine for phenylketonurics). Third, replacement of deficient or low-level molecules can also be performed, the classical example being factor VIII transfusion in hemophiliacs. Fourth, cofactor supplementation (for instance, vitamins) functions in cases of impaired enzymes. Fifth, other forms of management, such as drug administration, surgical operations, and organ transplantation, are effective for many conditions. Sixth, sophisticated approaches, such as stem cell therapy or gene therapy, can be tried.

Stem cells are unspecialized cells capable of self-renewal and of giving rise to specialized cells of different tissues. They are found (a) in embryos, and are capable of differentiating into virtually all organs or tissues (are *pluripotent*); or (b) in adult tissues, in every organ or tissue, but can differentiate only into the specialized cell types of the organ or tissue from which they were extracted (are *multipotent*). Adult cells can also be induced to be pluripotent stem cells (iPSCs) by appropriately supplementing the culture medium on which they are cultivated with the appropriate chemical signals and/or stimuli.

The clinical application of these types of cells for what has been called *regenerative medicine* has great promise. For instance, it could replace nerve cells in spinal cord

injury or Parkinson disease, or cardiac myocytes in persons who had suffered myocardial infarcts. However, the techniques needed to implement such measures are not trivial, and in addition there are ethical concerns about using human embryos (even those that were stored in fertility clinics and were not used, and that therefore have to be discarded) for research purposes.

Another type of intervention is gene therapy, a process by which a normal gene is introduced into a patient's cells to treat or cure his/her disease. Theoretically, this insertion could be made in either somatic or germline cells, but the latter procedure is strongly opposed to avoid eventual harm to future generations. As for somatic cell gene therapy, two types of treatment are possible: (a) *ex vivo*: the normal gene is introduced into cells removed from the patient, treated, and returned to him/her via an appropriate vector; or (b) *in vivo*: the normal gene is introduced directly (via a vector) into the patient's tissue requiring treatment.

The steps required for the development of a successful gene therapy are detailed in Box 8.18. The procedure seems straightforward, but there are many problems that plague the field since the first trials, which occurred some 25 years ago. Be as it may, two unquestioned successes of *ex vivo* therapy are now known, related to adenosine deaminase deficiency and X-linked adrenal leukodystrophy, as well as one of *in vivo* therapy (direct insertion of the normal gene into the retina of patients with Leber's congenital blindness that led to an impressive restoration of vision) [4].

Box 8.18 Steps necessary for the development of a successful gene therapy.

1. Ascertain risk–benefits. Gene therapy is a complex procedure that should be tried only if no other option is available.
2. Isolate and clone the normal gene. At present, this is a relatively easy task.
3. Define the regulatory regions necessary for gene expression. Since different genes are regulated in diverse ways, this step is necessary so that the gene product will be processed in adequate amounts.
4. Insert the gene (with or without an adjacent regulatory region) in a suitable vector (usually a virus turned nonpathogenic).
5. Transduce the patient's cells with the vector. Stability of the transduction is essential for a successful therapy.
6. Monitor the patient for undesirable effects.

Source: Rosenberg and Rosenberg 2012 [4].

Conclusions

Medicine, and the concepts of health and disease, should be always considered in an evolutionary framework. This chapter indicates some of the approaches that should be followed, in terms of both research and patient care.

The technological progress of molecular and bioinformatic methods defies adjectives. The costs of sequencing DNA have dropped continuously over the years, and the so-called next-generation sequencing platforms furnish information about millions or billions of short DNA fragments. Even more ambitious targets of analysis than simple sequencing are being developed; an example is the possibility of forming non-natural DNA with a fifth synthetic base that could specifically bind to a target molecule, such as a disease-related protein. This technique is presently at the experimental phase, but its potential for the development of new treatment drugs is enormous [36].

On the other hand, developments in genomic medicine for the 21st century and beyond face the challenge of changing the "diagnose and treat" to the "predict and prevent" paradigms. The time may arrive when by donating a drop of blood or saliva we could obtain our whole genome, with indications of our pattern of genetic disease susceptibility for the whole life. The next step, which is much more difficult to achieve, is to have specific means of preventing or curing these future ailments. However, scientific developments alone are not enough; all these questions must be tempered by issues of cost effectiveness and universal access to the tools for a healthy life. If we create shiny new technologies that remain economically out of reach for the majority of the population, we will loose the opportunity of contributing for a happier future for all [1].

Review questions and exercises

1 A whole movement, disseminated on the Internet, asserts that we should abandon the modern diets and strictly adhere to food such as those that were ingested by our prehistoric antecessors. Do you agree?

2 Was agriculture malefic or benefic to human evolution? Elaborate your answer.

3 How should pollution and environmental degradation be halted?

4 Can the frequencies of acts of violence be diminished and eventually completely eliminated? Or are we genetically programmed for such acts and can do nothing to prevent them?

5 A client recently diagnosed as having Huntington disease, which may lead to serious neurological disorder in old age, refuses to permit disclosure of the diagnosis and relevant genetic information to siblings who may be at risk (the condition is inherited as an autosomal dominant trait). How would you behave in this case?

6 A female undergoes diagnosis for infertility, and the tests show that she is chromosomally male (XY), carrying an androgen insensitivity syndrome. Would you reveal this finding to her?

7 A couple already had four children of the female sex, and desperately wants a boy. The woman is pregnant again and they request prenatal diagnosis. They say that if the embryo is female they will abort it. What would you do?

References

1 MacArthur DG, Lek M. 2012. The uncertain road towards genomic medicine. *Trends Genet* **28**:303–306.
2 Stearns SC. 2012. Evolutionary medicine: its scope, interest and potential. *Proc R Soc B* **279**:4305–4321.
3 Strachan T, Read A. 2011. *Human Molecular Genetics*, 4th edition. Garland Science, New York.
4 Rosenberg LE, Rosenberg DD. 2012. *Human Genes and Genomes: Science, Health, Society*. Elsevier, Amsterdam.
5 Tabor HK et al. 2014. Pathogenic variants for Mendelian and complex traits in exomes of 6,517 European and African Americans: implications for the return of incidental results. *Am J Hum Genet* **95**:183–193.
6 Stearns SC et al. 2010. Evolutionary perspectives in health and medicine. *Proc Natl Acad Sci USA* **107** (Suppl 1): 1691–1695.
7 Haig D. 2008. Intimate relations: evolutionary conflicts of pregnancy and childhood. In: Stearns SC, Koella JC (eds), *Evolution in Health and Disease*, 2nd edition. Oxford University Press, Oxford, pp 65–76.
8 Karlsson EK et al. 2014. Natural selection and infectious disease in human populations. *Nat Rev Genet* **15**:379–393.
9 Harper KN, Armelagos GJ. 2013. Genomics, the origins of agriculture, and our changing microbe-scape: time to revisit some old tales and tell some new ones. *Am J Phys Anthropol* **57**:135–152.
10 Kosova G. et al. 2010. Heritability of reproductive fitness traits in a human population. *Proc Natl Acad Sci USA* **107** (Suppl 1): 1772–1778.
11 Bittles AH, Black ML. 2010. Consanguinity, human evolution, and complex diseases. *Proc Natl Acad Sci USA* **107** (Suppl 1): 1779–1786.
12 Anholt RRH, Mackay TFC. 2012. Genetics of aggression. *Ann Rev Genet* **46**:145–164.
13 Goldman D. 2012. *Our Genes, Our Choices: How Genotype and Gene Interactions Affect Behavior*. Academic Press, Elsevier, Amsterdam.
14 Sell A et al. 2009. Formidability and the logic of human anger. *Proc Natl Acad Sci USA* **106**:15073–15078.
15 Caspi A et al. 2002. Role of genotype in the cycle of violence in maltreated children. *Science* **297**:851–854.
16 Gómez-Valdés J et al. 2013. Lack of support for the association between facial shape and aggression: a reappraisal based on a worldwide population genetics perspective. *PLoS One* **8**: e52317.
17 Price DH. 2007. Buying a piece of anthropology. Part 1. Human ecology and unwitting anthropological research for the CIA. *Anthropol Today* **23** (3): 8–13.
18 Price DH. 2007. Buying a piece of anthropology. Part 2. The CIA and our tortured past. *Anthropol Today* **23** (5): 17–22.
19 Borgerhoff Mulder M, Rauch KL. 2009. Sexual conflict in humans: variations and solutions. *Evol Anthropol* **18**:201–214.
20 Harris CR. 2004. The evolution of jealousy. *Am Sci* **92**:62–71.
21 Bowles S. 2006. Group competition, reproductive leveling, and the evolution of human altruism. *Science* **314**:1569–1572.
22 Bowles S. 2009. Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors? *Science* **324**:1293–1298.
23 Choi J-K, Bowles S. 2007. The coevolution of parochial altruism and war. *Science* **318**:636–640.
24 Kelly RC. 2005. The evolution of lethal intergroup violence. *Proc Natl Acad Sci USA* **102**:15294–15298.
25 Bingham PM. 2000. Human evolution and human history: a complete theory. *Evol Anthropol* **9**:248–257.
26 Simpson GG. 1949. *The Meaning of Evolution*. Yale University Press, New Haven, CT.
27 Bruneau EG et al. 2012. Social cognition in members of conflict groups: behavioral and neural responses in Arabs, Israelis and South Americans to each other's misfortunes. *Philos Trans R Soc B* **367**:717–730.

28 Finch CE. 2010. Evolution of the human lifespan and diseases of aging: roles of infection, inflammation, and nutrition. *Proc Natl Acad Sci USA* **107** (Suppl 1): 1718–1724.

29 Ackermann M, Pletcher SD. 2008. Evolutionary biology as a foundation for studying aging and aging-related disease. In: Stearns SC, Koella JC (eds), *Evolution in Health and Disease*, 2nd edition. Oxford University Press, Oxford, pp 241–252.

30 Anstad SN, Finch CE. 2008. The evolutionary context of human aging and degenerative disease. In: Stearns SC, Koella JC (eds), *Evolution in Health and Disease*, 2nd edition. Oxford University Press, Oxford, pp 301–311.

31 Leonard WR. 2008. Lifestyle, diet, and disease: comparative perspectives on the determinants of chronic health risks. In: Stearns SC, Koella JC (eds), *Evolution in Health and Disease*, 2nd edition. Oxford University Press, Oxford, pp 265–276.

32 Bamshad M, Motulsky AG. 2008. Health consequences of ecogenetic variation. In: Stearns SC, Koella JC (eds), *Evolution in Health and Disease*, 2nd edition. Oxford University Press, Oxford, pp 43–50.

33 Hayakawa T et al. 2014. Frequent expansions of the bitter taste receptor gene repertoire during evolution of mammals in the Euarchontoglires clade. *Mol Biol Evol* **31**:2018–2031.

34 Hof R. 2014. Seven months after FDA slapdown, 23andMe returns with now health report submission. Forbes Magazine, June 20, 2014 (http://www.forbes.com).

35 Wertz DC, Fletcher JC. 1989. *Ethics and Human Genetics: A Cross-Cultural Perspective*. Springer, Berlin.

36 Hirao I. 2014. Creating new medicines using non-natural DNA. *RIKEN Res* **9** (3): 18–21.

CHAPTER 9

Recent human evolution: an integrative approach

For the first time in human evolution, the individual life is long enough, and the cultural transformation swift enough, that the individual mind is now a constituent player in the global transformation of human culture.

—*William Irwin Thompson* [1]

SUMMARY

The study of human evolution is enriched from the findings of a number of scientific disciplines that are traditionally distinct and otherwise seemingly unconnected. In particular, the study of recent human evolution benefits from physical and social anthropology, linguistics, musicology, archaeology, statistics, computational biology, biometrics, embryology, population genetics, history, and molecular biology, among others. Knowledge from each of these fields individually nourishes the body of information that gradually and sometimes dramatically modifies our view of how modern humans originated, evolved, and populated our planet. Yet, it is when information from several of these areas of research is combined and comparatively analyzed that synergism often occurs. The potential corroboration of data from different disciplines obtained independently and from unique sources allows scientists to refute or support a theory. To illustrate the integrated nature of the study of human evolution, in this chapter, we provide a number of case studies in which the different fields have contributed to the advancement and resolution of scientific problems, from the migration out of Africa that resulted in the peopling of the world to specific dispersals. The Austronesian expansion represents a good example of how linguistic, music, art, physical anthropology, molecular biology, population genetics, and fauna/flora studies are being employed to assess this impressive trek. Linguists have been able to follow this dispersal while their artistic expressions are traceable in the traditional Lapita pottery. Molecular biology and the use of various genetic marker systems, such as mitochondrial DNA (mtDNA) and Y chromosome-specific loci, coupled with powerful population genetics and statistical algorithms, have provided corroborative data, yet conflicting at times, on this diaspora and the interactions with Melanesians, the original inhabitants of parts of Oceania.

Recent human evolution

The study of contemporary human populations involves the application of a number of disciplines including anthropology, linguistics, and molecular biology. Boxes 9.1 and 9.2 outline the most pertinent characteristics of anthropological, linguistic, and genetic markers in the study of recent human population genetics. The applications of these fields together usually have positive and synergistic impacts on investigations. At times, information from one area of work may clue investigators to search for support from other approaches. Although at times results from different lines of work contradict, congruent data are a powerful mechanism to confirm conclusions. Disagreement could also be beneficial in signaling potential faulty scientific methods and/or rationale.

Linguistics is the study of the evolution of languages, extinct and extant. It analyzes symbols, alphabets, grammar, syntax, and spelling and is particularly useful in determining the spread of culture, yet it fails to

Genomes, Evolution, and Culture: Past, Present, and Future of Humankind, First Edition.
Rene J. Herrera, Ralph Garcia-Bertrand, and Francisco M. Salzano.

© 2016 John Wiley & Sons, Ltd. Published 2016 by John Wiley & Sons, Ltd.

Box 9.1 Characteristics of archaeological and linguistic markers.

Linguistic data	Archaeological data
Assessment of the evolution of languages Include analyses of symbols, alphabets, grammar, syntax, and spelling	Assessment of human remains Include the analysis of bones, artifacts, tools, musical instruments, food, and radioactive decay dating
Useful for determining spread of culture	Useful for assessing morphological changes in time, as well as determining the location and times of human settlements and the culture responsible for those sites
Unfortunately, fail at differentiating spread of culture (acculturation) from demic expansions of people (gene flow)	Unfortunately, susceptible to gaps in the fossil record and deterioration with time, as well as contamination from other geological strata and human intervention
Limited to analysis of cultures for which language information is available	Limited by sparse distribution of artifacts, which are dependent upon the environmental conditions that destroy or preserve them
Limited to events that occurred subsequent to the development of language and writing	Limited (typically) by low sample sizes, minimizing the utility of statistical analyses

Box 9.2 Characteristics of molecular markers.

Advantages	Limitations
Molecular markers largely reflect similarity via *ancestry* as opposed to *state*	Limited availability of DNA from ancient samples
Large variety of neutral markers	Degradation of ancient DNA, which is susceptible to contamination from ancient and contemporary material
Sample sizes from modern populations can be large allowing statistical analyses	Age limitation (~400,000 YA) of samples that can be analyzed
More amenable to quantitation (e.g., allelic frequencies, genetic distances)	Does not allow extrapolation of samples from the distant past
Provide direct evidence for or against gene flow	Difficulties in calibration of molecular clocks (number of mutations per unit time) due to different rates of change (not mutation) for different genes
Speed of data collection and analysis	

discriminate between acculturation events and demic expansion of people or gene flow. The field of linguistics is not useful when dealing with populations for which language information is not available or for time periods prior to the development of language and writing. In combination with anthropological and molecular biology data, linguistics can help elucidate whether art, culture, and technological innovations were driven by information transfer and/or gene flow.

Archaeology is the study of human remains and the physical culture that they left behind. Archaeology usually involves the analysis of bones, artifacts, tools, musical instruments, food, and radioactive decay dating. These types of data are particularly informative for assessing morphological changes as a function of time, especially if the fossil record is not sporadic. In addition, archaeological information is paramount to determine the location, ages, and periods of human occupation and the culture responsible for the sites. Furthermore, when human remains and artifacts are studied in the context of climatic conditions, it can provide insights into the environmental and geographical forces that directed

evolutionary change, helped shape the genetic constitution of populations, and drove dispersals. The synergistic effects resulting from the combined analysis of data from these disciplines are well illustrated in the study of the evolution of North African hominins east of the East African Ridge and their early migration into Arabia. All of these disciplines, in combination with ancient DNA data, are capable of ascertaining the genetic constitution of extinct populations, provide for comparisons with contemporary humans, and allow for the exploration of changes in the gene pools of populations in specific geographical regions. Unfortunately, archaeology is subject to gaps in the fossil record and deterioration resulting from elapsed time and human intervention, contamination of the artifacts, and geological distortion. Environmental conditions such as heat and humidity are factors that can expedite deterioration of artifacts and remains. In terms of statistical analysis of the data, when dealing with fossils, limited number of individuals prevents the application of robust statistical analyses.

Since its inception in human population studies in the late 1970s, molecular biology has played an increasingly important role in assessing phylogenetic relationships among contemporary and ancient human groups. The strength of molecular markers (e.g., DNA, RNA, and proteins) resides in the number of samples available for analysis, especially when dealing with contemporary samples. The large number of individuals usually available for study allows for highly stringent statistical and phylogenetic analyses (see Chapter 3). In addition, a large number of selectively neutral markers have been identified and are now available in commercial kits allowing increasing number of investigators to rapidly assess genetic differences among human groups. Furthermore, molecular markers are more amenable to objective quantitation of the data in the form of allelic frequencies or genetic distances, for example. These types of analyses are designed to test for gene flow. One of the most important attributes of molecular markers is their capacity to discriminate between ancestral and derived states. This ability to identify the ancestral condition (allele or nucleotide) provides for a polarity in the direction of mutations, very valuable in phylogenetic studies. This quality of molecular markers is clearly evident in indels such as *Alu* insertions where the insertional state is almost invariably the derived state. With the increasing refinements of DNA sequencing techniques of human fossils extending the analysis to samples close to half a million years old, molecular markers in combination with anthropology and archeology are playing an increasing synergistic role in ascertaining phylogenetic relationships of ancient humans. Yet, perhaps the most important criticism to molecular biology markers is that due to the limited numbers of testable samples and age limitations of informative fossilized material, most of the molecular biology studies are performed on contemporary specimens and the question always remains, can past evolutionary events be extrapolated correctly from the present genetic makeup of populations?

Out of Africa

The Levant versus the Horn of Africa

Current genetic, anthropological, archeological, and geological evidence suggests that modern humans originated in Africa. Genetically, the degree of diversity in sub-Saharan Africa is much greater than that in any other part of the world, congruent with an African genesis of modern humans. In addition, mtDNA estimates indicate that the most common female ancestor lived in Africa about 200,000 to 150,000 years ago (YA) [2]. The fossil record also shows that the region that is today Ethiopia was the cradle of *Homo sapiens sapiens*. Geological and climatic evidence, discussed in Chapter 6, also corroborates an African origin. Furthermore, anthropological evidence specifically demonstrates that early humans migrated out of Africa into Eurasia on a number of occasions during the Upper Pleistocene (126,000 to 11,700 YA) [3]. Yet, considerable debate still exists regarding the times and routes utilized by these early migrants. Two strategically located passageways have been delineated in connection with dispersals out of Africa: one in Northeast Africa, the Levant, into what is today the Sinai Peninsula; and the other, the southern corridor from the Horn of Africa to what is today Yemen (Figure 9.1). It is likely that several dispersal events took place since the genesis of modern humans about 200,000 YA. For example, it has been postulated that at least four major migrational events have occurred, three in the Upper Pleistocene and a fourth one more recently [4]. In the context of this discussion, it is important to recognize that modern humans not only migrated out of Africa but also returned back to Africa, perhaps using the same corridors used to emigrate.

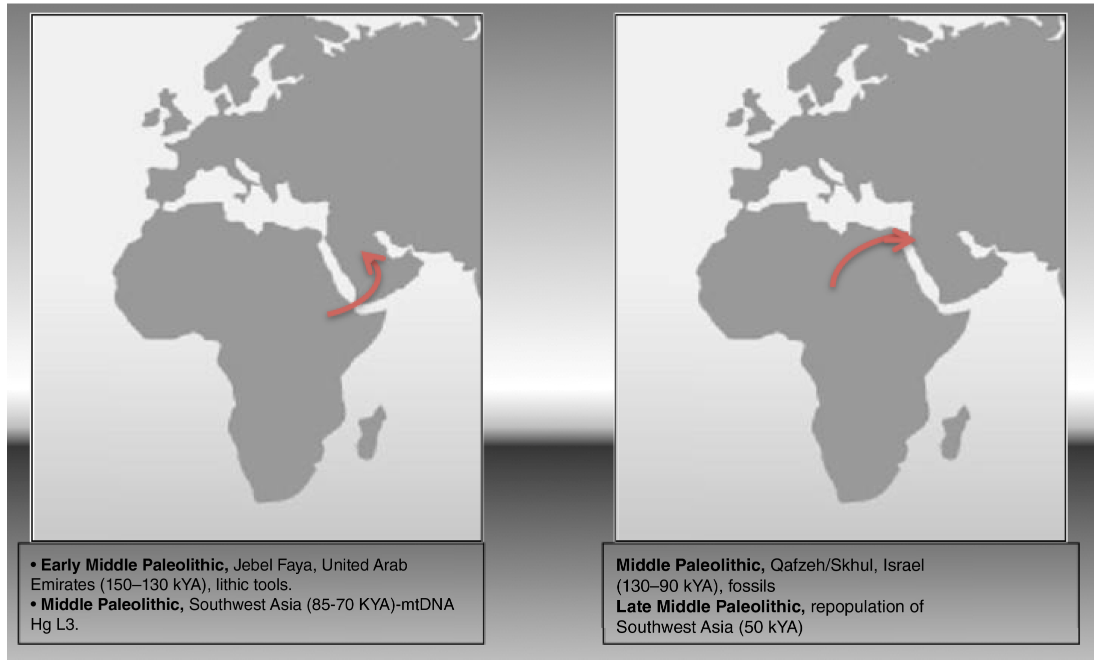


Figure 9.1 The Levant (northern route) versus the Strait of Bab-el-Mandeb (southern route). The two well-established migrational corridors are indicated. The dates of the crossings out of Africa as well as the location and type of evidence are indicated.

Via the Horn of Africa

The southern path across the Strait of Bab-el-Mandeb at the southern end of the Red Sea is thought to be the gateway for the oldest known dispersal out of Africa (150,000 to 130,000 YA). This period of time corresponds closely with low sea levels resulting from a glacial period (180,000 to 125,000 YA). It is even possible that a land bridge actually existed during the glacial maxima. It has been estimated that the sea could have been as much as 130 m lower compared with the current level [3]. A continuous land bridge would allow migrants moving from Northeast Africa to Southwest Asia to cross without the need of rafts. This drop in sea level, during glacial epochs, happens because water is sequestered in the enlarged ice caps. Exposed areas, previously under the sea, provided easy passage for early modern humans to move into previously submerged coastal regions and occupy previously unexplored areas.

In connection with the Red Sea corridor is the Jebel Faya site radiometrically dated to about 127,000 to 95,000 YA [3]. In this Southeast Arabia site, lithic tools manufactured using *façonnage* technology exhibit striking similarities to Northeast African counterparts suggesting a connection between the two. In addition, the observed parallelisms between the Northeast African and Jebel Faya stone tool traditions dating to similar time periods may be indicative of a long-term occupation of early modern humans. Unfortunately, the Jebel Faya site does not provide human remains, only stone tools. Also, it is not clear how long this early settlement in the Arabian Peninsula lasted and whether these migrants ventured further east into Asia. In this regard, the earliest evidence of modern humans in Asia is considerably more recent from about 100,000 YA in Southeastern China, in the region of Zhirendong [5].

Across the Nile corridor

Early modern human remains at the sites of Qafzeh and Skhul in present-day Israel signal an occupation that likely penetrated Arabia via the Levant (Figure 9.1). Unlike the Jebel Faya site in which only lithic material has been discovered, the Qafzeh and Skhul cave digs are made up of fossils and artifacts dating to 130,000 to 90,000 YA [6]. Significantly, these early human remains were discovered in close proximity to Neanderthal fossils. Yet, it is not clear to what extent these two hominin

species were interacting at these locations. Specifically, it is not evident, for example, whether the disappearance of these early modern humans from the Levant was the result of direct competition with Neanderthals, assimilation into a common gene pool, and/or retreat back to Africa. At these Levant sites, the contemporaneous Neanderthals and early modern humans shared the Levallois technique for making tools as well as cave habitats (see Chapter 6). In fact, it is not possible to determine whether individual blades and triangular flakes were made by Neanderthals or early modern humans. This uncertainty prevents meaningful comparisons between modern human artifacts in the Levant and Northeast Africa. What is clear about this cohabitation scenario is that no modern human presence is seen in the Levant after the period of the Qafzeh and Skhul occupation. Modern humans are not seen again in the region until 50,000 YA as a result of the resettlement of Southwest Asia [7] (Figure 9.1). Also, as with the Jebel Faya settlement, it is not clear whether the Levantine human population was able to migrate further east into Asia.

The Nile passageway was possibly used by modern humans from Northeast Africa into the Levant about 50,000 years ago [8]. This reoccupation of the Levant and Southwest Asia by modern humans subsequently led to the settlement of Southeast Europe and eventually Western Europe about 40,000 to 35,000 YA by way of the Danube River [7].

Two routes have been proposed for the migration from Northeast Africa to the Levant corridor. One scenario envisions a coastal dispersal northward along the Red Sea, while a second one theorizes a path across central Sahara utilizing a river corridor that cut through the Libyan Desert and led to the Mediterranean coast. Considering that this Libyan wet passageway existed about 130,000 to 117,000 YA [9], it is possible the Qafzeh and Skhul settlements are the result of human movement along its basin. Lithic deposits along the Red Sea coast such as Abdur dating to about 125,000 YA and Asfet [10], both in present-day Eritrea, render support to the coastal Red Sea putative route to the Nile corridor.

Genetic data

Based on genetic data, the southern route is generally associated with the dispersal of modern humans from the Arabian Peninsula further eastward into South Asia across what is today the Strait of Hormuz into the Indian subcontinent toward Southeast Asia, Melanesia, and Australia [8]. Contemporary mtDNA data indicate that the L2 and L3 maternal lineages experienced an expansion during a period of 85,000 to 55,000 YA (see Chapter 5 for explanation of haplotypes and how they change over time). According to these uniparental mtDNA lineages (haplotypes), the best age estimates suggest that a limited group of females carrying haplotype L3 migrated out of Africa approximately 70,000 YA via the Strait of Bab-el-Mandeb (Figure 9.1) [11].

The paternal counterpart of mtDNA inheritance is the Y chromosome. Also uniparentally transmitted but along the male line of descent, Y chromosome-specific markers, as part of haplotypes, have been employed to follow the male-driven out of Africa migrations. Based on Y chromosome-specific DNA haplogroups, it is known that the paternal lineages of all modern humans had their origins in Africa [12]. The most recent common ancestor (MRCA) of all modern human Y chromosomes has been recently dated between 120,000 and 156,000 YA, yet a wide range of ages have been estimated, ranging from 142,000 to 338,000 YA [13]. The most ancient modern human Y haplogroups are A and B. These two Y chromosome types are exclusively African in origin (Figure 9.2). Specifically, haplogroup A is mostly found among Khoisan populations, while haplogroup B is specific to pygmies and other groups from the Congo. It is likely that haplogroups A and B never left Africa during the out of Africa episodes. It has been postulated that the migrants that dispersed out of Africa and populated the rest of the world carried an ancient point mutation on the Y chromosome known as M168. From this M168 Y chromosome, all non-African haplogroups originated outside of Africa. It is estimated that M168 individuals dispersed out of Africa about 70,000 to 50,000 years ago and differentiated into haplogroups C to T in Southwest Asia [14]. These date estimations suggest that the early modern human migrants' maternal and paternal lineages left Africa at contemporaneous times and possibly as part of the same migration. This parallelism in departure date for both uniparental set of markers favors the Horn of Africa as the putative route. Figure 9.3 illustrates the worldwide distribution of major Y chromosomes.

More recent migrations across the north and south corridors

Since the early dispersals out of Africa (Figure 9.1), genetic signals from several additional migrational episodes are apparent [15,16]. The signatures of these migrations are

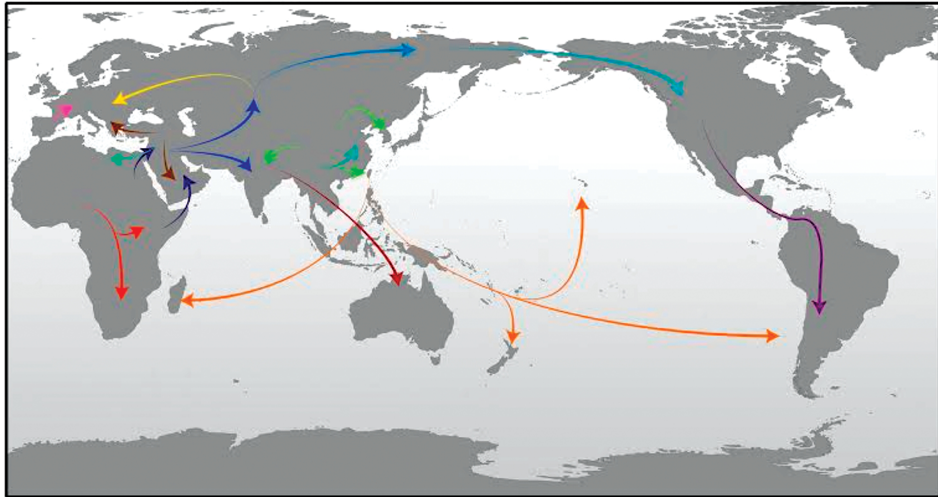


Figure 9.2 Major modern human migrations. Approximate dates and diagnostic genetic markers are indicated for each individual expansion. (See the Color Plates section.)

evident when specific mtDNA and Y chromosome-specific markers are examined within Africa and in Eurasia. Table 9.1 summarizes the time periods, putative corridor used by migrants, and informative markers.

All of the gene flow episodes, except for the one marked by the M1 mtDNA haplotype, date from the Upper Paleolithic (40,000 to 10,000 YA) to the Neolithic (10,000 YA to the present). In other words, except for the M1 mtDNA haplotype, these genetic systems do not detect earlier dispersals including the older, fast track coastal migration along Southwest Asia and the Indian subcontinent that led to the settlement of Melanesia and Australia about 68,000 to 50,000 YA [16,17]. It is likely that the older signals have been lost due to genetic dropout since uniparental DNA markers, as part of single haplotypes, are notoriously susceptible to deletion from the gene pool. Alternatively, the absence of markers from this (these) earlier migration (s) out of Africa has been diluted out by overwhelming numbers of individuals carrying various mutations from subsequent more recent migrations.

Another trend seen in this listing of genetic signals is that the majority of them involved the Levant crossing,

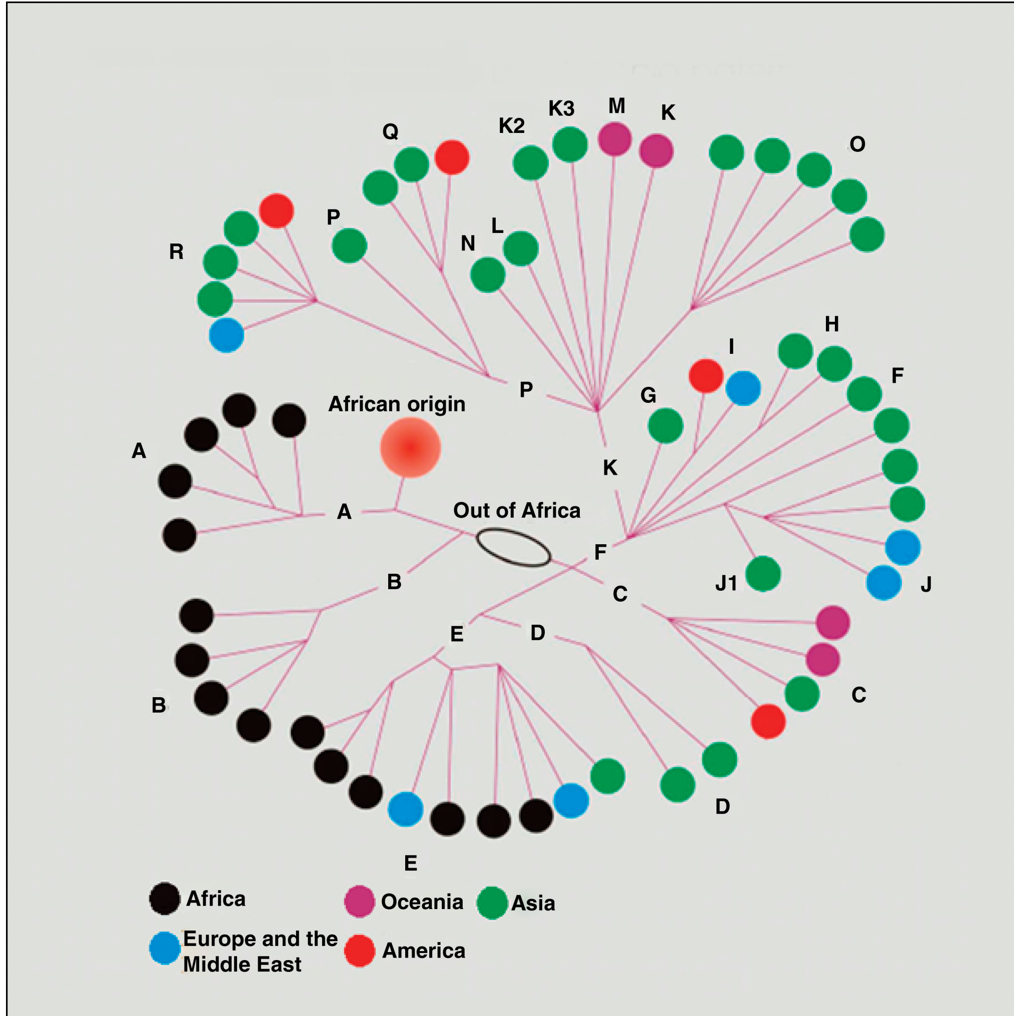


Figure 9.3 Worldwide distribution of Y chromosome haplogroups. (See the Color Plates section.)

most of them going back to Africa from Eurasia (Table 9.1). It is possible that at the time most of these migrations occurred, during the Upper Paleolithic and Neolithic epochs, drier, arid conditions affected specifically the Horn of Africa and Southwestern Arabia regions, preventing passage across the Strait of Bab-el-Mandeb. It is also conceivable that interglacial periods could have compromised dispersals through the southern route due to treacherous conditions resulting from deeper waters and currents in the strait.

Worthwhile noting is that all the signals going out and into Africa through the Strait of Bab-el-Mandeb derived from mtDNA markers (haplogroups M1, K, pre-HV1, and T1). Furthermore, it is also notable that a higher frequency of sub-Saharan mtDNA (haplogroups L1, L2, L3, and U) compared with Y chromosome (E1b1b1a-M78 and E1b1b1c-M123) lineages is found in the Arabian populations (Figures 9.4 and 9.5).

Back to Africa

A review of the information provided by these uniparental DNA markers reveals that the majority of the migrations detected are not out of Africa but back to Africa (Table 9.1). These back to Africa dispersals are

Table 9.1 Summary of time periods and migrational paths of modern humans out of and into Africa.

mtDNA		
Haplotype/haplogroup	**Levant (time span)**	**Horn of Africa (time span)**
H	E to A (UP-LGM)	NE
J	E to A (UP-LGM)	NE
J1b	E to A (UP-LGM to N)	WE
K	E to A (UP-LGM to N)	E to A (UP-LGM to N)
M1	NE	A to E or E to A (MP to UP)
N1b	E to A (UP-LGM to N)	NE
Pre-HV1	E to A (UP-LGM to N)	E to A (UP-LGM)
T1	E to A (UP-LGM to M)	E to A, WE (?)

NRY		
YCC group/haplogroup	**Levant (time span)**	**Horn of Africa (time span)**
E1b1b1a-M78	A to E (N)	NE
E1b1b1c-M123	A to E (N)	NE
E1b1b1b-M81	NE	NE
G-201	E to A (N)	NE
J2-M172	E to A (N)	NE
T-M70	E to A (LGM)	NE for LGM
R1*-M173	E to A (LGM)	NE
R1b-M17	E to A (N)	NE
R1b-M269	E to A (N)	NE

The time periods of possible migration are provided in parentheses. A "?" indicates that there is not enough evidence to estimate time interval. Abbreviations are as follows: E to A, Eurasia to Africa; A to E, Africa to Eurasia; UP, Upper Paleolithic; MP, Middle Paleolithic; N, Neolithic; M, Mesolithic; LGM, Last Glacial Maximum; NE, no evidence for use during the Upper Paleolithic to Neolithic; WE, weak evidence for use during the Upper Paleolithic to Neolithic.

clearly observed when the distribution of Y chromosome-specific (Figure 9.5) and mtDNA (Figure 9.4) haplogroups of Eurasian origin is illustrated in regions inside Africa and in the Southwest Asian side of the south and north corridors.

In terms of Y chromosome-specific markers, there are a number of markers and haplogroups that had their genesis in Asia, subsequent to the departure of early modern humans from Northeast Africa about 70,000 years ago [15, 16], and then dispersed into Africa. For example, although the DE-M1 suprahaplogroup (YAP insertion) has been postulated to have its origin in East Africa approximately 65,000 years ago [18], more recent reports indicate Asian origin and an introduction into Africa later [19]. The same can be said about haplogroup E, a descendant of DE. Haplogroup E has been postulated to originate in East Africa [20] and Asia [21] by two independent research groups. If indeed haplogroup E, the most abundant lineage in Africa, had its genesis in Asia, it represents the most successful back migration of modern humans into Africa. Other Y chromosome-specific haplogroups that have penetrated Africa are G-M201, J2-M172, T-M70, R1*-M173, R1b-M17, and R1b-M269. G-M201 is thought to have originated in the Northwestern Arabia about 30,000 to 9,500 YA and since then crossed into North and Northeast Africa [22]. Similarly, J2-M172 present in West Asia around 22,000 to 15,000 YA moved into North and Northeast Africa during the Neolithic [18]. T-M70, with a homeland also in West Asia and dating back to 30,000 to 19,000 YA [23], seems to have traveled to Africa using both northern and southern pathways; during the Last Glacial Maxima

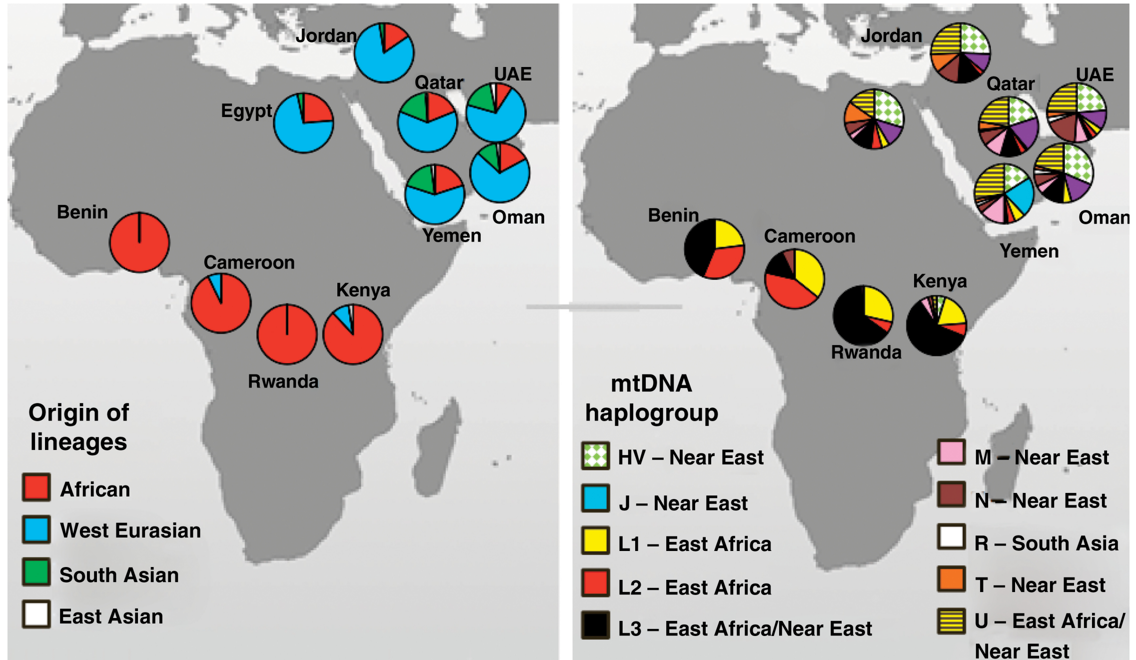


Figure 9.4 Origins and distribution of mitochondrial DNA haplogroups within Africa and on the southwest Asian side of the Strait of Bab-el-Mandeb and the Nile corridors. (Adapted from Rowold et al. 2007 [15]. Reproduced with permission of Macmillan Publishers Limited.) (See the Color Plates section.)

(LGM) through Nile corridor into North Africa [16]. Haplogroup T-M70 exhibits particularly high frequencies in the Horn of Africa [16]. The R1-M173, R1b-M17, and R1b-M269 have origins in South Asia. The most basal R1 mutation (R1-M173) dates back to about 12,500 to 25,700 YA [24] moving into Africa some time later. The high frequencies of R1-M173 in Cameroon suggest a sizeable migration of individuals carrying this mutation into sub-Saharan Africa [25].

The mtDNA makers that signal migrations back to Africa include H, J, J1b, K, M1, N1b, pre-HV1, and T1. All but one (M1) of the dispersals signaled by these mutations are relatively recent. M1 originated around 60,000 YA in Africa or Asia [26]. If M1 originated in South Asia, it has been theorized that it moved into Africa about 40,000 YA. T1, on the other hand, is thought to have entered Africa during a period of time from the Upper Paleolithic to the Mesolithic (20,000 to 10,000 YA). The rest of the mtDNA back to Africa signals date to a period of time between the LGM (26,000 to 20,000 YA) and the Neolithic (10,000 YA to the present). The majority of these migrations utilized the Nile corridor (Table 9.1). It is possible that many of the out of Africa migrations occurred early, soon after the initial dispersal that populated Southwest Asia, South Asia, Melanesia, and Australia, and their signals have been weakened below detection or erased by time. In other words, the signals that we detect are from recent crossings involving humans that expanded into Africa with new more advanced technology, for example, the agriculturist migrants that moved from the Near East to North Africa in the Neolithic. In terms of the preference for the Nile corridor over the Horn of Africa for dispersals exiting and entering Africa during the LGM to the Neolithic, the aridity maximum around 22,000 to 13,000 YA may have limited or closed the Strait of Bab-el-Mandeb to traffic only leaving the north crossing for migrations [27].

Beyond Arabia

In the next sections, we will provide a relatively detailed picture of the population movements in Asia, Europe, America, and Oceania, as case studies illustrating how

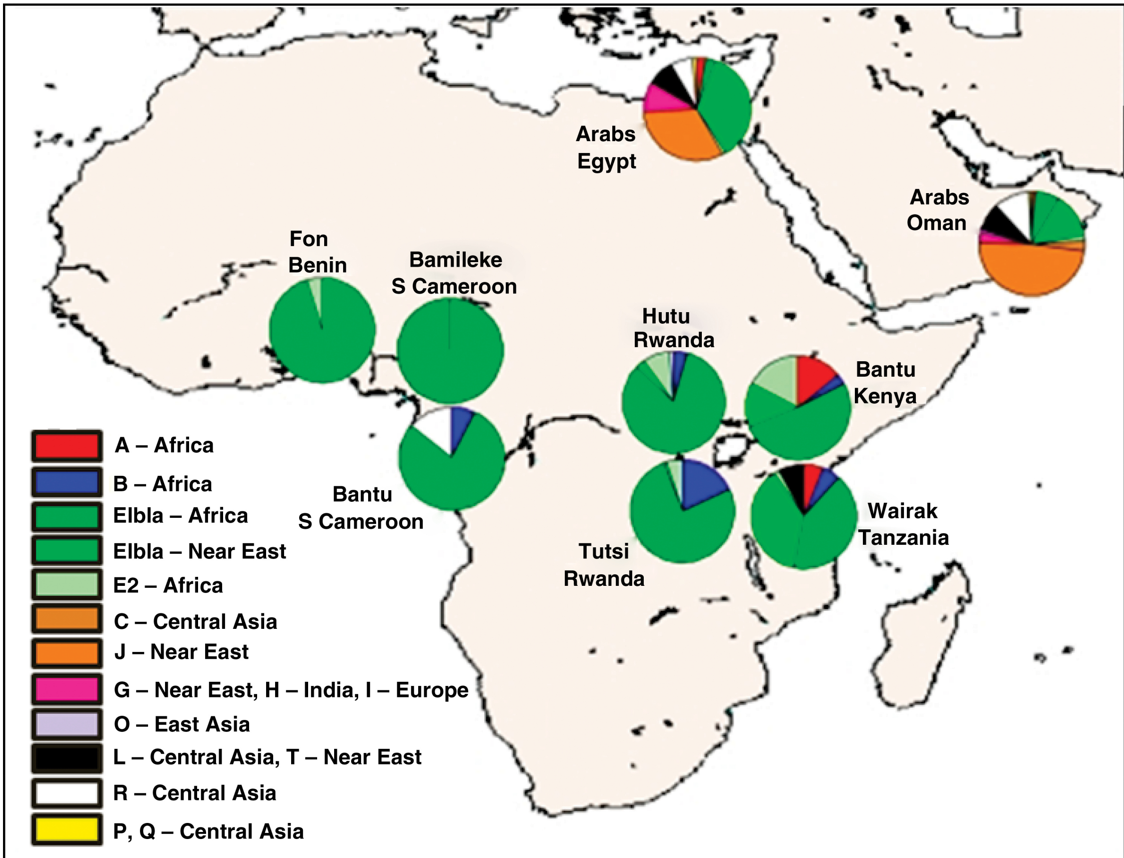


Figure 9.5 Origins and distribution of Y chromosome haplogroups within Africa and on the southwest Asian side of the Strait of Bab-el-Mandeb and the Nile corridors. (Adapted from Luis et al. 2004 [16]. Reproduced with permission of Elsevier.) (See the Color Plates section.)

several lines of evidence can furnish information about our evolution.

The colonization of Australia

It is generally recognized that once early modern humans penetrated the Arabian Peninsula via the Horn of Africa about 70,000 YA, it set in motion an eastward migration (see Figure 9.2) [8]. It is thought that this dispersal was coastal along the Indian subcontinent as it continued to Southeast Asia [28]. Considering the time these humans exited Africa and the original colonization of Australia (approximately 60,000 YA), a number of investigators think that this west to east transcontinental migration, which culminated in the settlement of Asia, Melanesia, and Australia, took place very rapidly. The trek lasted approximately 10,000 years. Considering the coastal distance between Southwestern Arabia and Australia, the speed of this migration would have been about 1 km per year. In addition to this fast-track migration that colonized Southeast Asia, a number of subsequent major dispersals took place. These dispersals with corresponding dates and diagnostic genetic markers are illustrated in Figure 9.2.

Southeast Asia was the immediate destination of this coastal migration [29]. Subsequently, it is thought that modern humans dispersed north into what is today Southern China and then Northern China. Another branch of this migration ventured south into the Malaysian Archipelago in route to Melanesia and Australia (Figure 9.2). At the time of arrival of the Australian aborigines' ancestors, the sea level in and near Australia was approximately 150 m lower resulting from a period of glaciation. The decrease in sea levels in the region allowed New Guinea and Tasmania to join in a

meta-continent with Australia. This ancient mass of land is known as Sahul. Also, this meta-continent extended further into what is today the Timor Sea and closer to continental Asia. Yet, in spite of the lower sea level, Australia was separated from the primordial Malaysian Archipelago by about 90 km of water [30]. It is theorized that these migrants moved by island hoping between Sulawesi and New Guinea and/or via Timor into Northwest Australia. Thus, these early travelers had to cross water, deliberately or accidentally. It is not clear how humans with limited technology could perform such a crossing requiring rafts.

Archeological and anthropological data from Australia indicate that at least two forms of aborigines existed in the continent, the gracile and the robust [31]. Proponents of the candelabra theory, which advocates multiregional origins for modern humans, suggested that the robust form is indicative of continuity with earlier *Homo erectus* populations in Southeast Asia. Both gracile and robust remains have been recovered from fossils from the Pleistocene and early Holocene. The oldest human remains in Australia were found in Southeastern Australia in a dry lake known as Mungo and they date back to at least 60,000 YA. The Mungo samples belong to the gracile type suggesting that this type arrived prior to the robust [31]. These two anatomical types have been linked to different continental penetrations. Yet, the anatomical parameters of the robust type are outside the indices of contemporary aboriginals. Considering the distance of the dry Mungo Lake, about 2700 km from the putative zone of initial incursion into northern Australia, it is likely that the first arrival by modern humans dates prior to the Mungo settlement by several thousand years.

Mitochondrial and Y chromosome-specific DNA studies indicate that Australia was settled 40,000 to 70,000 YA from a single genetically heterogeneous wave or several independent penetrations [32]. Based on assessments from these two marker systems, contemporary Australian aborigines are genetically quite heterogeneous and for the most part are similar to residents of Asia. The observed genetic similarities between Australian and New Guinean aborigines suggest a common source of migration(s) and/or admixture events subsequent to colonization.

A number of uniparental markers signal the original settlement of Australia. In terms of mtDNA, haplogroup N14, a descendant of the African L3, was brought into Australia during the original migration. It is thought that N14 originated in Asia around 71,000 years ago [33]. Another haplogroup, M, possibly of South Asian origin from around 60,000 YA, also marks the entrance into Australia [33]. Both haplogroups M and N are associated with the initial out of Africa crossing and the dispersals that populated the world. The Y chromosome-specific mutations linked with the arrival to Australia include C4-M347 and S-P60. Just like the mtDNA haplotypes, C4-M347 and S-P60 are tied to the original circum-Indian dispersion. About 99% of the Australian tribal population possesses either C4-M347 or S-P60 Y haplotype. C4-M347 is exclusively Australian and its parent major haplotype C originated in Asia about 53,000 years ago [34]. C4-M347 is the most abundant Y haplogroup with frequencies around 60% of the population [34]. The other major native Y haplogroup, S-P60, is found at levels as high as 40% [35]. The abundance of these two predominant indigenous haplogroups varies within the Australian continent. For example, in South Australia the frequency of S-P60 is approximately 60% while that of C4-M347 is around 40% [36]. It is not clear whether these regional differences resulted from genetic drift or multiple migrations and clinal expansions (e.g., migration of C4-M347 individuals as they moved southward into S-P60 territory). S-P60's parent haplogroup, S-P405, has its origin in Southeast Asia or New Guinea subsequent to the initial arrival of modern humans approximately 41,000 to 28,000 YA. It is not clear whether S-P60 is of Australian origin or had its genesis in Southeast Asia or New Guinea like its parent haplogroup S-P405.

The mtDNA and Y haplogroups that define the aborigines are ancient and basal in the phylogenetic tree. Ancient mtDNA analysis also indicates that Australia possesses the deepest lineage of anatomically modern humans in the world [31]. The deep haplogroups in Australian aborigines suggest relative isolation subsequent to the initial settlement of the continent. Thus, the uniparental genetic data support the view that the technological developments later seen in Australia, including the backed-blade lithic industry, are *in situ* developments and not introduced by more recent migrations. More recent genome-wide investigations also support the coastal southern route with comparable arrival dates in Australia [37]. Yet, these investigations have detected more recent gene flow from India about 4230 years ago, which coincides with the introduction of certain technological advances including new tool manufacturing styles, food processing methods, and

the introduction of the dingo (*Canis lupus dingo*). These data demonstrate that Australia did not remain in complete isolation since the initial settlement about 60,000 YA.

Modern humans in Europe and America

The settlement of Europe and America is relatively recent. In the case of Europe, contemporary humans populated the region around 45,000 YA during the Upper Paleolithic where they encountered Neanderthals and shared the continent until the Neanderthals became extinct about 27,000 YA [38]. These first modern humans were hunters and gatherers and lived in small groups in a large expansion of land. The beginning of the Neolithic, about 10,000 YA, is marked by a number of cultural and technological developments known as the agricultural revolution in the Levant. Although there were a number of transitions from hunter-gatherer to agrarian societies in the world, the one that resulted in the repopulation of Europe has motivated considerable research. Some of the attention generated revolves around the issue of whether the agricultural advances in Europe were brought about by acculturation or demic diffusion of Neolithic humans from central Anatolia, present-day Turkey. In other words, to what degree the change to an agrarian society was due to just communication of technology as opposed to the migration of individuals. It turns out that the answer to the question is region specific with certain areas reflecting various proportions of Paleolithic and Neolithic DNA [39].

The colonization of America was also relatively recent. The estimates range from 40,000 to 15,000 YA [40]. Although a number or important issues involving the movement of modern humans into the New World are intensely debated, recent multidisciplinary developments in molecular biology, archaeology, linguistics, and paleoclimatology have provided fresh information on this diaspora. Central to the settlement of America are the climatic episodic openings and closings of the migratory corridors from Asia to Beringia and into the Americas [41].

The dispersal from Asia to America can be partitioned into three phases. The first was the northern, trans-Siberian migration, as early as 40,000 YA [41], of Paleolithic people from the Altai region of Central Asia. These Paleo-migrants eventually populated Beringia (the now submerged body of land between Siberia and Alaska) [42]. In Beringia, a migratorial stop, likely induced by harsh climatic conditions during the last glacial period, may have resulted in the encapsulation of these Paleo-natives in between the two continents for as long as 15,000 years, and thus allowing for genetic variability and subpopulation differentiation resulting from genetic isolation [40]. The third stage was the demographic dispersal into what is now Alaska and beyond, all the way south to Tierra del Fuego. This southbound dissemination of people was most likely facilitated by the onset of global warming, approximately 12,500 YA after the LGW.

Although the Asian origin of Native Americans is well established, a number of issues still remain unanswered. One point of contention involves the time and the speed of the trip from Asia to the southernmost territories in South America. It has been calculated that, at a very rapid rate of 200 km per century, proto-Amerinds could have trekked the approximately 16,000 km of difficult terrain in about 7000 years [43]. Yet, the presence of archaeological remains in Monte Verde, Chile, dated at 14,600 YA, requires this exodus from Beringia southward into austral South America to have been initiated at least 21,600 YA, considerably before the end of the last glacial period (12,500 YA). A number of investigators contest that a rapid movement of people could have taken place south to the cone of South America via a Pacific coastal route in about 2000 years [44].

The number of migrations that penetrated America is also a subject of controversy. Critical to this argument is whether or not random stochastic fluctuations generated by genetic drift, bottleneck episodes, and founder effects once humans arrived in the New World could account for the geographical patterns of genetic diversity observed in the extant Native Americans, or whether it is necessary to suggest separate migrations. It has been suggested that the marked cranial differences among Paleo-natives at various sites in Washington, Nevada, and Brazil, ranging in time from 5600 to approximately 10,000 YA, and contemporary Native Americans cannot be explained by random genetic drift alone subsequent to the incursion into America and propose that there were two major dispersals from different ancestral populations in Asia [45]. Other investigators contest that the observed craniofacial diversity can be explained with a single migrant population with high levels of diversity and the action of genetic drift and multiple bottleneck events.

In addition to morphological data, genetic studies have been directed at assessing the time(s) of inception of specific genes into America as well as the likelihood of a single migration contributing the genetic diversity currently

observed. The early studies during the second part of the 20th century using classical genetic markers such as blood groups and HLA polymorphism established the Asian origins of Native Americans. Subsequent mtDNA investigations at the beginning of the 1990s [46] provided evidence for several crossings confirming Greenberg's tripartite theory based on Amerind, Na-Dene, and Aleut/Eskimo linguistic groups [47]. Yet, later mtDNA work suggested that the genetic partitioning and time estimates were best explained by a single major dispersal [48]. About the same time, Y-SNP and Y-STR experiments were not conclusive as to the number of dispersals into America [49]. Later Y chromosome studies suggest multiple migrations [50]. More recent genome-wide SNP results are congruent with Greenberg's tripartite theory, but with a single wave contributing most of the migrants [51].

The Asian agricultural revolution and the Austronesian expansion

Two major human dispersals populated the Pacific Ocean

The settlement of the Pacific Ocean represents the climax of the global expansion that started in Northeast Africa about 70,000 to 80,000 years ago. The peopling of the Pacific Ocean took place in two main dispersal events. The first was an outshoot of the initial coastal migration out of Africa that settled Australia. This early migration from Southeast Asia into peninsular and insular Indonesia, New Guinea, and the Solomon Islands started about 40,000 years ago during the beginning of the Late Paleolithic. It culminated with the settlement of the Bismarck Archipelago by around 33,000 YA and the Solomon Islands approximately 29,000 YA. Thus, this first dispersion only encompassed and reached Near Oceania. At present, most of the descendants of this initial Pacific dispersal speak Papua and are referred to as Melanesians. On the other hand, the second dispersal known as the Austronesian expansion is relatively more recent, starting about 3000 YA from Southeast Asia. This second migration introduced the Austronesian language to Oceania.

An agriculture-driven migration

The agricultural revolution is the name given to a number of cultural transformations that allowed humans to change from a hunting and gathering subsistence to one of agriculture and domestication [52]. This profound difference in lifestyle let on to the establishment of stable settlements with large number of residents and sufficient food for storage for consumption during lean times. This development and dependence on agriculture and domestication involved a number of other environmental adaptations including deforestation, irrigation, and the allocation of land for specific crop cultivation. In addition, it led to various other dramatic innovations such as division of labor, new tool technologies (e.g., the harvester's sickle), trading, architecture, centralized political systems, and nonverbal systems of communication, such as writing. The last development provided for far-reaching consequences in the form of the present-day information revolution. This shift in mode of subsistence also provided for more leisure time and safe existence leading to free time to think and to be creative.

The development of agriculture took place independently in different parts of the world and at different times. Although the term agricultural revolution is usually associated with the Near East, the Fertile Crescent, and specifically Anatolia (present-day peninsular Turkey), marking the start of the Neolithic (12,000 to 10,000 YA) period, other regions such as the Mexican Plateau (5000 to 4000 YA), the Andean region of South America (5000 to 4000 YA), New Guinea (9000 to 6000 YA), Mid-North America (4000 to 3000 YA), Northwest sub-Saharan Africa (5000 to 4000 YA), and the Yangtze and Huang River basins in Eastern Asia (9000 YA) have experienced parallel developments [53].

In what is today China, several species of plants were domesticated at different times and in a number of regions. Archaeobotanical studies suggest that rice was domesticated independently in what is today Southern China as well as in the middle and the lower Yangtze River basins at about 5000 YA [54]. Other crops such as millets started in Eastern Inner Mongolia around 8000 YA, whereas soybeans were first cultivated in Northern China approximately 5000 YA.

It has been theorized that the impetus for the settlement of what is today the island of Taiwan, formerly Formosa, by mainland Southeast migrants was provided by the Asian agricultural revolution as farmers with the newly developed crops expanded their domains within China and beyond to Island Southeast Asia [55]. Genetic and linguistic evidence suggests that these farmers were Daic-speaking groups from Southeast Asia [56]. The agrarian settlement of Taiwan happened about 6000 to

5000 YA from coastal Mainland Southeast Asia. A considerable portion of the island of Taiwan is occupied by a rugged central cordillera with small farming villages in the coastal alluvial plains. It is thought that these original agriculturalists of Taiwan migrated to other Southeast Asian islands such as the Philippines approximately 4000 YA possibly in search for additional lands for cultivation [57]. This dispersal into Island Southeast Asia and ensuing island hoping into the Pacific Ocean was likely driven by population growth and the need for more uncontested land. In this process, it is likely that these agrarian migrants reached to the Batan Islands for more land. The distance between Taiwan and the Philippine's Batan Islands to the south is just 190 km. The Batan Islands were populated by Taiwanese aborigines about 3000 YA, carrying their agricultural heritage, and since then these two groups of populations have been culturally and economically linked. It has been postulated that this migration by Taiwanese tribal farmers initiated a dispersal throughout Island Southeast Asia that culminated in one of the human's major diasporas, the peopling of Oceania. The Austronesian expansion represents the most geographically extensive dispersal undertaken by humans. Thus, what started as an agrarian acquisition of land in the Mainland Southeast Asian Neolithic revolution resulted in the peopling of two-thirds the circumference of the world, from the island of Madagascar in East Africa to the west and Easter Island, off the coast of Chile, to the east.

Linguistic evidence

Today, the island of Taiwan is populated by about half a million (2% of the total population) aborigines that belong to nine major tribes: the Ami, Atayal, Paiwan, Bunun, Puyuma, Rukai, Tsou, Saisiyat, and Yami. All of the tribes speak different Austronesian languages. The rest of the population is represented by different groups of Han Chinese, mainly the Min and Hakka subgroups. Taiwanese aborigines are unique in that 9 out of the 10 currently spoken Austronesian linguistic subgroups are found in Taiwan. The rest of the Austronesian-speaking world speaks only one subgroup, the Malayo-Polynesian or Extra-Formosan branch. This high linguistic diversity is regarded as evidence that the Austronesian language family has its roots in the island. It is noteworthy to highlight that the natives of Orchid Island, 60 km off Southeast Taiwan in the direction of Oceania, also speak the Extra-Formosan Austronesian branch [58]. The Austronesian language family encompasses a wide geographical range bound by Madagascar to the west, Easter Island to the east, New Zealand to the south, and the Hawaiian chain to the north [58].

Cultural parallelisms

In addition to the Austronesian language that links the majority of Pacific populations to the Taiwanese aborigines, cultural data also suggest connections between the two groups. For example, the use of tattoos as ritual body decorations is an ancient practice in common among Taiwanese aborigines, Austronesians from the Philippines, Indonesians, and Borneo as well as Polynesians [59]. Across the Austronesian dominion, the tattoos usually represent geometric patterns and figures of plants and animals as well as anthropomorphic expressions. These same representations are seen in the pottery throughout the region. Specifically, the Lapita pottery is characterized by geometric dentate-stamped themes, which suggests a connection between the Taiwanese aborigines and the people of Oceania [59]. The finding of Taiwanese-mined nephrite within the Austronesian domain could also be interpreted as signs of ancestral kinship or even trade [59]. A societal system based on agriculture and trade, government by patrilineal chiefdoms, and practicing pantheistic religions are shared as well. Of course, some of these attributes may be derived from Mainland Southeast Asian populations such as the Daic. Other cultural commonalities that point to Taiwan as the source of the expansion include the Austronesian-wide use of outrigger sailing canoes with shared designs, slit drums, and hula-type dancing.

Genetic data

Some of the genetic data suggest that the agriculturalist Taiwanese aborigines are the contemporary descendents of the migrants that initially set off to sea approximately 4000 YA and in about 3000 years populated the Pacific and part of the Indian Ocean (Madagascar and the East African coast) [60, 61]. Although in recent years a number of reports have suggested that the Daic domain within South China, and not Taiwan, was the source of the Austronesian expansion [56], the island still lies at the root of these interrogatives. The Daic, also agriculturists, are considered to be the original inhabitants of China's southeast coast and their origin can be traced to 20,000 YA [56]. Today, the Daic is the second largest ethnic group in China, after the Han, and the population has strong presence in Thailand, Laos, Vietnam, Myanmar, and India [56].

Several explanations have been proposed to delineate the mode by which Austronesians reached their destinations. Yet, three hypotheses have received greater attention by geneticists, linguists, and archeologists: the "express train," "entangled bank," and "slow boat" theories. The "express train" model, also referred to as "out of Taiwan" model, suggests that Austronesians originated in Taiwan and traveled rapidly through Micronesia and Melanesia with minimal genetic admixture with Melanesians, the preexisting population of the area before settling throughout Oceania [61]. The "entangled bank" hypothesis, on the other hand, posits that a dispersal from Southeast Mainland Asia into Oceania during the Mid-Holocene (7000 to 5000 YA), along with continued and extensive gene flow with Melanesian natives throughout the trek, is responsible for the current genetic characteristics of Austronesian groups [62]. The "slow boat" model, which combines some of the main points of both aforementioned ideas, theorizes that though Austronesian migrants are of Asian descent (most likely Taiwanese), they traveled slowly through Island Southeast Asia assimilating some of the preexisting genetic substrata and eventually reaching Near and Far Oceania [60].

Since the 1990s, genetic investigations have been performed to address the genesis of the Austronesian people [63]. Genetic data have the potential to complement other lines of evidence such as linguistics and archeology to shed light on this subject. Central to the use of genetic information to address the Austronesian expansion is to what degree the language and archeological parallelisms observed result from acculturation or gene flow.

The presence of the mtDNA haplogroup B4a1a1 (the so-called Polynesian motif) and its ancestral lineage B4a1a among Taiwanese tribes as well as in Polynesian groups provided, for the first time, direct genetic evidence for a connection between these two groups of populations [64]. In addition, phylogenetic analyses based on mtDNA haplogroups from the Taiwanese tribes and Pacific Ocean populations (excluding Australia and Inland New Guinea), in general, echo linguistic relationships [53]. Furthermore, the genetic phylogeny generated using mtDNA data parallels a geographical progression starting in Taiwan, where it is most diverse, traveling south toward the Philippines and Indonesia [63].

The following mtDNA studies began to reflect a more complex story in which the genetic heritage of Pacific Austronesians was a combination of both Asia, possibly of Taiwanese aboriginal descent, and Melanesians [65]. Melanesians are probably the original inhabitants of Near Oceania possibly dating back to the original coastal migration that populated Australia. The Melanesian domain includes islands from Eastern Indonesia eastward to the Solomon Islands. Melanesians speak different forms of Papua, although some islands are linguistically Austronesian. Some of these islands within the Melanesian range that speak Austronesian exhibit a blend of Melanesian and Austronesian genetic elements [65]. For example, in these admixed island populations, high frequencies of Asian mtDNA and Melanesian NRY (nonrecombining Y chromosome) haplogroups were detected. These results are of interest since Polynesian societies are matrilocal. It seems that in these interphase islands, although there was complete language replacement, the gene pools remained admixed. These results are not surprising considering the dynamics and speed of language acquisition versus complete genome substitution. The orthodoxy at the time started thinking that Austronesians, prior to reaching the Middle and Far Oceania, had a somewhat extended layover in Melanesian territory, with a stay long enough to allow admixture between the two groups. Thus, it was beginning to look as if there was never an "express train" after all but more like a "slow boat." More recent studies have also uncovered previously undetected concordance linking the mtDNA of Pacific Islands populations with Melanesian tribes from the Bismarck Archipelago in Indonesia [66]. The time estimate for the separation of these two groups of populations is approximately 9000 YA. This predates the Austronesian expansion by as much as 3000 years. Considering the earlier TMCA (time to most recent common ancestor) derived from this study, it reinforces the likelihood that an older mtDNA strata preexisted in Oceania prior to the arrival of the Austronesians.

Although early Y chromosomal studies indicated genetic connections between Taiwanese tribes and Polynesians [67], ensuing studies were controversial. The source for the differences of opinion started when close genetic ties between Melanesia and Polynesia were reported in the literature [68]. In fact, at that time, some of the leading authorities in the field were not seeing clear Y chromosome evidence for affinity between Polynesia and tribal Taiwan [68]. A revival of the contention between Taiwanese aborigines and the Austronesian expansion occurred when it was discovered that a

subgroup of haplogroup O3, specifically O3a2, is widely distributed throughout Island Southeast Asians, Indonesians, and Polynesia [69]. Of interest in connection with Mainland Southeast Asia and specifically the Daic groups as a source of the Austronesians was the finding that the O3a2 haplogroup is absent from those populations. More recently, a specific genetic relationship involving the O3a2c*-P164 subhaplogroup was detected between the Ami (one of the main Taiwanese aboriginal tribes) and Polynesian populations [70]. O3a2c*-P164 is found at very low levels and in only some Mainland East Asian populations. The Daic populations examined lacked O3a2c*-P164. These findings established a direct genetic link between a specific Taiwanese tribe and Polynesian groups previously undetected due to the minimal resolution of O3-derived Y chromosomes afforded by previous studies.

Genome-wide studies, on the other hand, have detected minimal Melanesian DNA in Micronesian and Polynesian populations [71]. The results of these genomic scans, overall, provide a more general panorama of the genome of the Pacific populations since they assess recombining DNA at many sites as opposed to uniparental lineages. It could be argued that mtDNA and Y chromosome-specific haplogroups lack the resolution afforded by thousands of recombining autosomal loci and that autosomal markers are less subject to lineage dropouts. The genome-wide results indicate that Melanesian groups are genetically very different from each other, partitioning along islands and island locations (e.g., coastal versus inland). In general, Melanesian island interiors are more homogeneous than coastal populations, the latter exhibiting various degrees of Austronesian (Asian) admixture (<20%). No Papua-speaking populations from the interior of islands exhibit Austronesian (Asian) admixture. Non-Melanesian groups (e.g., Polynesians) tend to be genetically more homogeneous than Austronesian-speaking Melanesians. On the other hand, Polynesians segregate with Micronesians, Taiwanese tribes, and insular East Asians, but not with Melanesians. The genomic data suggest that the passage of Austronesians along the Melanesian domain was not slow since the degree of genomic admixture was low between these two groups. In other words, according to the genomic scan data, the boat was not idle.

A recently published genome-wide study also indicates that Austronesians possess more affinity to tribal Taiwanese than to Melanesians or Mainland Southeast Asian populations [72]. These data not only confirm previous genome scans but also suggest that Oceanic Austronesians' immediate ancestors are not mainland populations such as the Daic but insular aborigines from Formosa. Yet, this investigation also provided evidence for a mainland genetic component in western Island Southeast Asia in addition the Austronesian signals. These results suggest that insular Southeast Asia may have received gene flow from Austronesians, likely from Taiwan, as well as from Southeast Asia mainlanders.

The motivation

It is not clear what prompted Austronesians to sail into the open sea not knowing their destination. It is possible that some of the colonization events were accidental, as fishing parties may have drifted out to sea and carried away with the prevailing currents to distant islands. Along those lines, a recent pertinent case involving Chilean fishermen who ended up in the shores of Hawaii, months later, comes to mind. And although the distances among islands tend to be large, at times huge, it is possible that during the timescale of hundred of years, such unintended migrations may have occurred periodically. Yet, it is highly unlikely that most of the settlement events resulted from serendipity.

Many of the islands in the Pacific Ocean are separated by thousands of kilometers of just open water. Were Austronesians running out of land to cultivate? This is a real possibility considering that most of the islands in Oceania are volcanic and small and possess very jagged mountainous profiles with limited suitable flat land for crops. Was the migration motivated by internal conflicts, possibly feuds among families or clans? Maybe sailing to the unknown were desperate actions resulting from ecological collapse, or maybe was just the desire to explore the unknown. Considering the duration, distances, and dangers associated with the voyages, it seems that independent of the nature of the motivations, Austronesians were highly driven. Also central to this issue, did Austronesians think of the ocean as a barrier or a highway for commerce and cultural exchange? What is known is that in just about 4000 years or less, Austronesians moved from Southeast Asia to Easter Island off of Chile, South America. These travelers traversed the entire Pacific Ocean.

The timeline of this odyssey goes as follows: the Tonga and Samoan Archipelagos, in Central Oceania, were settled about 2800 years ago, where the migrants

experienced a hiatus of approximately 500–1000 years [73]. This pause may have been required to allow for the development of the adequate nautical technology to sail the distance of over 2500 km to the Society Islands in French Polynesia. The ensuing migrations were faster. From Central Polynesia, Austronesians reached the Eastern Pacific including the Cook Islands, the Hawaiian Archipelago, French Polynesia, New Zealand, and then Easter Island about 1200 years ago [74]. These last legs of the Austronesian expansion spring boarded from the island of Ra'iatea in the Society Group in French Polynesia. Easter Island or Rapa Nui, possibly the last land to be colonized, was settled by just 30–100 individuals in one or two landings [73].

Evidence from plants and animals

Biological markers of human dispersal

Austronesians traveled in double vaulted canoes fitted with one or more huts for their protection and preservation of their goods. The double vaulted design is known to provide the needed stability for sailing the high sea. These amazing vessels were capable of carrying 50–100 individuals. Their cargos were not just humans. They took with them supplies for consumption during the voyage and plants and animals that would help them survive in and colonize the new land. These live cargos can in fact be used as an extension and reflection of human activity. In other words, they are markers signaling human presence. Among the flora and fauna that Austronesians transported, there is evidence for the introduction of the dog, pig, chicken, rat, and dozens of plants throughout the Pacific Ocean [75]. In theory, fossils, remains, and the contemporary descendants of these stocks could be informative in efforts to trace the origin of the Austronesian people. In addition, these plants and animals may be revealing in answering questions regarding number of colonization events, failed human settlement attempts, and postcolonization contacts. In other words, by comparisons with potential sources, the species and varieties in the islands (dead or alive) may reveal their origins and in turn the origins of the travelers.

It is known that a number of species of plants and animals were transported from islands in the Fiji, Samoa, and Tonga Archipelagos in Central Polynesia to the islands further east, north, and south in Oceania [75]. It seems that these islands became centers of communication, commerce, and cultural exchange. To the east, the island of Ra'iatea in Eastern Polynesia received this flow of goods and information from Central Polynesia and in turn became a launching platform for the exploration and distribution to the islands in the north (the Hawaiian Archipelago), south (New Zealand), and the Far East Polynesia (the Marquesas Archipelago and Easter Island). The Marae or center of ceremony in Ra'iatea still can be seen as an impressive complex that was used for departure in trans-island voyages [76]. In other words, ensuing the initial discovery, the sea became less of a barrier and more of a highway for humans and their goods.

Dogs

Only a few species have been investigated as markers in an effort to trace the origins of Austronesians. These species include the dog, pig, chicken, rat, and the sweet potato. The dog found in Polynesia has been studied using ancient and contemporary mtDNA [77]. The results of these inquiries indicate that the domestic dog originated in South China about 10,000 to 16,000 YA. The timeline for the diffusion of the domestic dog indicates that it was present in the Gulf of Thailand around 4000 YA and it was first detected in the Moluccan Archipelago in Eastern Indonesia about 3300 YA, approximately the same time periods when pigs were found in the corresponding regions. Dogs were part of the Yüan-shan culture homestead in northern Taiwan as early as 4500 YA [78] and they are seen in Polynesia by 2000 YA [79]. This chronological–geographical progression in the appearance of the dog in Oceania generally mirrors the Neolithic movement of humans eastward into the Pacific Ocean. It is likely that just as with the dingoes in Australia, dogs in Polynesia were primarily used as a source of food and possibly companionship.

Mitochondrial DNA results from Central, North, and South Polynesia (the Cook Islands, Hawaii, and New Zealand, respectively) indicate the presence of only two dog haplotypes, Arc1 and Arc2. These two mtDNA haplotypes are found in South China, Mainland Southeast Asia, and Indonesia but not in Taiwan or the Philippines [79]. Similar haplotypes have been detected in the Australian dingoes. These data suggest that Austronesians from Taiwan did not transport the dog to Polynesia. Alternatively, the Formosan breed ancestral to haplotypes Arc1 and Arc2 became extinct after the

dispersal from Taiwan commenced. Thus, it is possible that the dog was introduced into Polynesia via Mainland Southeast Asia and Indonesia and it was picked up from Melanesia as the Austronesians traveled through their territory. The story provided by dog mtDNA data posits a complex scenario for the peopling of Polynesia with inputs from different regions, not just Taiwan. In other words, the results suggest that the Polynesian culture probably had a complex origin, with components from Taiwan as well as Indonesia and Melanesia.

Pigs

Pigs were independently domesticated in Europe and Asia from the wild boar (*Sus scrofa*) about 9000 YA [80] and introduced by humans east of the Wallace line [80]. In Oceania, two varieties of the Asian pigs exist, the Pacific and East Asian, with distinctive mtDNA haplotypes. The East Asian form is found within Micronesia in islands such as the Mariana and Palau as well as in Taiwan. The Pacific stock, on the other hand, exhibits a distribution that parallels a route from Mainland Southeast Asia into peninsular and insular Indonesia, Papua, New Guinea, the Solomon Islands, and Polynesia. In some islands such as Kosrae and Hawaii in western and northern Polynesia, respectively, the two strains coexist. These data advocate for two separate dispersions of the domesticated pig during the Neolithic. One, involving the Pacific strain, is thought to be exclusively linked to the Lapita culture and the Austronesian expansion, while the other (the East Asian) is associated with a dispersal from Taiwan into the Philippines and eastward into Micronesia [81]. The presence of both haplotypes in Kosrae most likely is the result of secondary imports into the island while the side-by-side habitation in Hawaii may represent the postcolonial introduction of the Pacific type.

These two distinct varieties of pigs are clearly characterized by distinct phylogenetic mtDNA clades [81]. The total lack of the Pacific clade in Mainland China, Taiwan, and the Philippines indicates that any migration of people from Taiwan to Island Southeast Asia and Melanesia (the out of Taiwan scenario) did not include pigs. In other words, the pigs taken in the voyages by Austronesians must have come from some other place.

Chickens

The domestication of the chicken is thought to have taken place in Southeast Asia from the wild red jungle fowl during the Asia agricultural revolution about 8000 YA [82]. This contention is supported by linguistic reconstruction of the proto-Austronesian word for chicken, *manuk*. The proto-Austronesian language is thought to have its genesis in Southeast Asia. It is thought that initially chickens were not primarily used for food but for fighting. From Southeast Asia, the domesticated race was taken to India where a number of varieties were developed and exported worldwide. All the mtDNA haplotypes currently found in the Americas, Europe, Middle East, and Africa are variations from the Indian subcontinent haplotypes [83].

Recent molecular and phylogenetic analyses have traced the Polynesian chicken to Island Southeast Asia [84]. Specifically, these whole mitochondrial genome (WMG) studies pinpoint the Philippines as the birthplace of the Austronesian chicken lineages [84]. Although two prominent major haplogroups were uncovered in Oceania, D and E, only lineage D, found only in the Pacific, is the authentic founding mtDNA chicken lineage associated with the Austronesian expansion. Haplogroup E seems to be a recent postcolonial introduction. Domesticated chickens were most likely transported from Micronesia to New Guinea via the Indonesian Archipelago about 3850 YA and from there to the Solomon Islands, Vanuatu, and then further east into the Far Pacific [84]. Interesting, the chicken and the rat were the only animals that reached or managed to survive the harsh conditions of Easter Island at the fringes of the expansion [75].

Although these studies specifically link the Polynesian chicken to the Philippines, the investigators surprisingly failed to sample the island of Taiwan. Thus, the question of whether the samples from the Philippines got there by way of Mainland Southeast Asia or by way of Taiwan remains unanswered. Alternatively, the Filipino haplotypes may have been the result of *in situ* domestication. Depending on which of these scenarios represents the authentic route of the chicken out of Southeast Asia, it would corroborate or not the out of Taiwan model.

Rats

The Pacific rat (*Rattus exulans*) is a ubiquitous omnivore distributed throughout Polynesia. The available data from mtDNA studies suggest that this animal originated in Southeast Asia and was then transported by Austronesians during their dispersal [85]. Its range includes Taiwan, the Philippine Archipelago, Mainland Southeast Asia, peninsular and insular Indonesia, and most of the

Pacific Islands. The Pacific rat is smaller than the more popular European brown and black rats and is considered a separate species. The fact that this rat cannot swim and perishes when in water, does not interbreed with the European brown and black rats, and was not present in Near Oceania prior to the Neolithic makes it a reliable proxy for tracing the Austronesian expansion including reconnaissance trips, number of colonization events, unsuccessful human settlement attempts, and contacts. In addition, the high levels of genetic variation present in the Pacific rat provide for high-resolution genetic studies [85].

It is likely that the rat's dissemination was at times accidental, although in some instances, it may have been planned. Just like the common rat is notorious for sneaking into more recent and contemporary cargo, *R. exulans* likely did the same during the Austronesian migrations. Nevertheless, its relationship with humans after landing should have been commensal and disruptive to the autochthonous fauna and flora. As a food source, the rat, though small in size, could help support humans in lean times. Its hardiness in stringent environments and high reproductive capacity allowed its survival in the harsh conditions of Easter Island [53].

Mitochondrial DNA studies indicate the presence of three major Pacific haplotypes, I, II, and III, for the Pacific rat with very distinctive and mutually exclusive distributions [86]. These data may suggest that each of these lineages may derive from unique sources and/or migrations, although further work is necessary to solidify this conclusion. Haplotype I is observed exclusively in Island Southeast Asia. Haplogroup II has representatives in Southeast Asia and Near Oceania, while lineage III is found starting in the Bismarck Archipelago and beyond into the rest of Polynesia. Fine-resolution analyses of the individual Polynesian haplotypes as well as phylogenetic and simulation tests derived from these studies were used to argue for a distribution center in Central Polynesia including Samoa, Tonga, and the Cook and Society Islands. These islands seem to have experienced multiple contacts with the rat. In addition, although isolation was the general pattern in East Polynesia (e.g., the Marquesas and Easter Island), evidence for multiple introductions was detected, especially involving Hawaii from Central Polynesia (Samoa and Tonga) and New Zealand from the Cook and Society Islands [86].

Although the phylogenetic and simulation data based on rat mtDNA suggested a route of dispersal starting in the Philippine Archipelago, the island of Taiwan was not included as part of the study. Therefore, the question of whether the Filipino haplotypes came from Taiwan remains unanswered. In spite of this sampling gap, the rat mtDNA distribution is generally consistent with a putative migration route of Austronesians from Island Southeast Asia into the Indonesian Archipelago, the Melanesian domain, and Polynesia. For example, the time of the arrival of the rat in New Zealand around 1280 of the present era is consistent with the dating from human material [87]. An important conclusion from the rat research is that the Polynesians did not live in complete isolation ensuing the initial settlements of individual islands. In other words, to different degrees, depending on the groups of islands considered, there was communication within the Polynesian domain.

When animal studies are taken together, a consistent theme for the origin of the Austronesian expansion is lacking. In addition, none of the animal species provide specific signals connecting the Austronesian expansion with Taiwan. The chicken and rat data leave the door open for a Taiwanese tribal origin since both trace the animals to the Philippine Archipelago but failed to sample Taiwan. In the case of the dog and pig, the Polynesian counterparts are not the ones found in Taiwan. The mtDNA of these animals connects them to Mainland Asia. Yet, it is possible that the Polynesian variety became extinct subsequent to the voyages out of Taiwan, but these scenarios are not the most parsimonious. Therefore, although the animal data point to a South Asian origin of the species, details of the migrations seem to be different. In light of these results, it is possible that Austronesians picked up some of their stocks while in Indonesia or perhaps Melanesia. Furthermore, in evaluating these results, it is important to consider all of the animal studies with caution since, thus far, all of them are based on mtDNA sequences only and therefore are susceptible to haplotype dropouts.

Contacts between South America and Polynesia

Polynesian chickens in Chile and South American sweet potatoes in Polynesia

A number of observations of the fauna and flora distribution point to contacts between Polynesia and South America. One is the existence of the sweet potato

(*Ipomoea batatas*) throughout Polynesia [88]. This plant is not Asian in origin but American. In fact, the sweet potato was domesticated about 10,000 YA at the beginning of the Andean cultural revolution. Radiocarbon dating indicates that the plant got to the Cook Islands at least 1000 years ago and to Mangaia Island in Central Polynesia about the same time [89]. Based on these precolonization dates prior to the incursions of Europeans into the Pacific, it has been theorized that the plant was transported from South America to Polynesia approximately 1300 years ago [88]. It is also possible that Native Americans transported the plant but it is likely that the Austronesians with their superior maritime skills were the venue. Also, it is unlikely that prevailing currents deposited floating seeds across the Pacific Ocean since this plant reproduces by vine cuttings.

Radiocarbon and genetic evidence of chicken remains from El Arenal, a site on the west coast of Chile, also suggests contact between Polynesians and South America [90]. Yet, controversy exists regarding dates of bones, contamination with contemporary DNA, and the significance of the presence of the mtDNA haplogroup E [84]. Central to the allegations of faulty results is the issue of contamination of the chicken ancient DNA from El Arenal with contemporary chicken DNA in commercial stocks of shrimp DNase (an enzyme employed to destroy extraneous contaminating DNA). In addition, the use of mtDNA haplogroup E by the investigators as a marker for Polynesian incursions into South America has been argued since this lineage is ubiquitously distributed in both the Pacific and South America [84]. It turns out that only the Polynesian-specific mtDNA haplotype D is not found in South America. Yet, the investigators proposing the link between Polynesia and South America claim that mtDNA haplotype E was exclusively introduced from Southeast Asia into Oceania about 4000 YA [90]. Also, there were no chickens in South America before the settlement of Europeans. Therefore, the question still remains how the chickens got to Chile prior to the arrival of Europeans.

Native American DNA in Easter Island

In addition to contacts between Southeast Asia and Oceania, whole-genome scans of human DNA have uncovered gene flow from South America to Polynesia [73]. Specifically, this gene flow has been observed from the cone of South America toward Easter Island or Rapa Nui. It is likely that Rapa Nui represents one of the last colonization events of the expansion, contemporaneous with the incursion into New Zealand. It is known that Easter Island was settled by Polynesians about 800 YA and it was rediscovered by Europeans in 1772. Prior to the landing of the first Europeans, the precarious enclave of Rapa Nui experienced a rapid decline in population, likely the outcome of ecocide resulting from widespread deforestation [53]. This ecological collapse essentially entrapped the Polynesians in their small island, an ironic and somber fate after their ancestors traveled more than 15,000 km from Southeast Asia to Rapa Nui in less than 3000 years. Subsequent to the ecological downfall, in the 1860s, the Austronesian population of Rapa Nui suffered a profound decline from about 4000 individuals to 100 as a result of the Peruvian slave trade and the ensuing European-introduced plagues. These population bottleneck episodes in tandem likely brought about a sharp decline in genetic diversity. Therefore, the number of 50–100 original settlers of the island based on current genetic diversity may represent an underestimation.

Results based on the HLA complex markers have provided suggestive evidence for gene flow from Native American populations to Polynesians from Easter Island [91]. More recently, by comparing the relative proportions of nuclear DNA from contemporary Europeans, South Americans, and Polynesians present in individuals of Polynesian descent from Easter Island (Rapa Nui), an average of 6 and 16% of South American and European components, respectively, were detected in an Austronesian genetic background [73]. These admixture levels were based on phylogenetic and statistical results derived from structure analyses (see Chapter 3 for description) and ancestry tract length distributions. The latter method takes advantage of the indirect relationship that exists between the length of linkage disequilibrium DNA tracks (continuous unrecombined pieces of DNA; see Chapter 5) and the time foreign chromosomes were introduced by introgression.

The time of the Native American admixture with the Rapa Nui, based on DNA track sizes, is between 1280 and 1425 of the present era [73]. These dates are of importance since they indicate not only a time interval ensuing the settlement by the Austronesians and prior to the rediscovery of Rapa Nui by the Europeans, but also a period of time before the cessation of long-distance travel by Polynesians at about 1450 [92]. In other words, at the time of the putative South American contact,

Austronesians were still on the expansion mode. By the time Europeans landed on Easter Island, in 1772, the ship building technology and canoes were not compatible with long trips. This corrosion of the Polynesian naval tradition among the Rapa Nui may have been linked to their overall decline as a society resulting from their ecological demise.

Although one-way contacts from South America represent a possible scenario, it is more likely that gene flow occurred via round trips by Polynesians since no architectural elements from South America are seen in Polynesia and Polynesians possessed a superior maritime technology compared with Native Americans. A point to consider in this discussion is the likelihood that Polynesians had contact with the Americas. In other words, is it likely that Polynesians missed America, a whole continent, at the apex of their age of expansion, when they were able to find Easter Island, a speck of land in the vast expanse of the Pacific Ocean? Although comprehensive and detailed investigations of Native American populations are not available, no indications of significant gene flow from Polynesians can be seen in the present Native American genome [51]. Thus, it is possible that if such contacts occurred, they may have been sporadic. It is also likely that any Polynesian genetic flow may have been diluted out after its introduction into the vast expanse of Native American DNA.

Review questions and exercises

1 How a multidisciplinary approach benefits studies on recent human evolution?

2 Enumerate and elaborate on the type of data provided by linguistics, anthropology, archeology, climatology, and genetics that may benefit investigations of recent human evolution.

3 What linguistic parameters are investigated when contrasting populations?

4 What are the advantages and disadvantages of linguistic, anthropological, and genetic data in the study of recent human evolution?

5 How the East African Ridge and Range impacted recent human evolution?

6 Is it scientifically reliable to extrapolate past evolutionary events from the distribution of genetic markers from contemporary populations?

7 Describe the migrations and dates of human dispersals across the Levant and the Horn of Africa.

8 What type of archeological evidence was recovered from the Jebel Faya, Qafzeh, and Skhul sites in the Arabian Peninsula? Include dates of each.

9 Speculate on the reasons for the fact that most intercontinental dispersals between Africa and Eurasia detected with DNA markers are towards Africa and relatively recent in time.

10 To what extent Neanderthals and early modern humans interacted in Arabia?

11 Is it possible for the maternal and paternal human lineages (mtDNA and Y chromosome) to have left Africa at different times as part of different migrations?

12 It is theorized that the initial out of Africa migration that led to the settlement of Australia was coastal and rapid. Why?

13 Fossils of Australian aborigines indicate two forms, gracile and robust. Is this proof of the multiregional theory for the origin of modern humans?

14 Argue for and against the statement "Australian aborigines represent a highly isolated group of people."

15 What may have motivated Mainland Southeast Asians to migrate to Formosa?

16 Trace the hypothetical route taken by Austronesians out of Taiwan to Easter Island.

17 Comment on the significance of the language of natives of Orchid Island, off the Southeast coast of Taiwan, being Extra-Formosan.

18 How could Madagascar natives be Austronesian speakers (Extra-Formosan) while other Indian Ocean populations are not Austronesian speakers?

19 Define Lapita culture.

20 List the similarities between Taiwanese and Polynesian populations.

21 Contrast the mtDNA, Y chromosome, and whole-genome evidence in relation to the out of Taiwan theory.

22 Argue for and against the “express train,” “entangled bank,” and “slow boat” theories.

23 List the advantages and disadvantages of uniparental, autosomal, and whole-genome genetic data.

24 Of all the potential reasons for Austronesians to populate Oceania and part of the Indian Ocean, which is the most compelling? Why?

25 How useful are the plants and animals carried by the Austronesians to the study of the expansion?

26 What are the advantages and limitations of the presently available data from the plants and animals transported by Austronesians?

27 Was the sea a barrier or a highway to the Austronesians?

28 Based on the available data, how likely a Polynesian–American contact would have been?

29 What reasons are usually given for the decline of the Rapa Nui population?

30 Explain how DNA track length and linkage disequilibrium are employed to generate TMCA values and population ancestry.

31 Considering the bottleneck episodes experienced by the Rapa Nui people, how realistic is the estimation of a founding population number of 50–100?

References

1 Thompson WI. 2010. *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Harvard University Press, Cambridge, MA.

2 Cann R. 1988. DNA and human origins. *Ann Rev Anthropol* **17**:127–143.

3 Beyin A. 2011. Upper Pleistocene human dispersals out of Africa: a review of the current state of the debate. *Int J Evol Biol*, doi: 10.4061/2011/615094.

4 Armitage SJ et al. 2011. The southern route out of Africa: evidence for an early expansion of modern humans into Arabia. *Science* **331**:453–456.

5 Liu W et al. 2010. Human remains from Zhirendong, South China, and modern human emergence in East Asia. *Proc Natl Acad Sci USA* **107**:19201–19206.

6 Mercier N et al. 1993. Thermoluminescence date for the Mousterian burial site of Es-Skhul, Mt. Carmel. *J Archaeol Sci* **20**:169–174.

7 Lahr MM, Foley R. 1994. Multiple dispersals and modern human origins. *Evol Anthropol* **3**:48–60.

8 Mellars P. 2006. Archeology and the dispersal of modern humans in Europe: deconstructing the Aurignacian. *Evol Anthropol* **15**:167–182.

9 Osborne AH et al. 2008. A humid corridor across the Sahara for the migration of early modern humans out of Africa 120,000 years ago. *Proc Natl Acad Sci USA* **105**:16444–16447.

10 Walter RC et al. 2000. Early human occupation of the Red Sea coast of Eritrea during the last interglacial. *Nature* **405**:65–69.

11 Macaulay V et al. 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* **308**:1034–1036.

12 Wang C-C, Li H. 2013. Inferring human history in East Asia from Y chromosomes. *Invest Genet* **4**:11–21.

13 Poznik GD et al. 2013. Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science* **341**:562–565.

14 Underhill PA et al. 2000. Y chromosome sequence variation and the history of human populations. *Nat Genet* **26**:358–361.

15 Rowold DJ et al. 2007. Mitochondrial DNA geneflow indicates preferred usage of the Levant Corridor over the Horn of Africa passageway. *J Hum Genet* **52**:436–447.

16 Luis JR et al. 2004. The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *Am J Hum Genet* **74**:532–544.

17 Bowler JM et al. 1970. Pleistocene human remains from Australia: a living site and human cremation from Lake Mungo, Western New South Wales. *World Archaeol* **2**:39–60.

18 Cruciani F et al. 2004. Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. *Am J Hum Genet* **74**:1014–1022.

19 Chandrasekar A et al. 2007. YAP insertion signature in South Asia. *Ann Hum Biol* **34**:582–586.

20 Semino O et al. 2004. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am J Hum Genet* **74**:1023–1034.

21 Chiaroni J et al. 2009. Y chromosome diversity, human expansion, drift, and cultural evolution. *Proc Natl Acad Sci USA* **106**:20174–20179.

22 Cinnioğlu C et al. 2004. Excavating Y-chromosome haplotype strata in Anatolia. *Proc Natl Acad Sci USA* **114**:127–148.

23 Mendez F et al. 2011. Increased resolution of Y chromosome haplogroup T defines relationships among populations of the Near East, Europe, and Africa. *Hum Biol* **83**:39–53.

24 Sharma S et al. 2009. The Indian origin of paternal haplogroup R1a1* substantiates the autochthonous origin of Brahmins and the caste system. *J Hum Genet* **54**:47–55.

25 Cruciani F et al. 2002. A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am J Hum Genet* **70**:1197–1214.

26 Gonzalez AM et al. 2007. Mitochondrial lineage M1 traces an early human backflow to Africa. *BMC Genomics* **8**:223.

27 Hooghiemstra H et al. 1992. Vegetational and climatic changes at the northern fringe of the Sahara 250,000–5000 years BP: evidence from 4 marine pollen records located between Portugal and the Canary Islands. *Rev Palaeobot Palynol* **74**:1–53.

28 Macaulay V et al. 1980. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Sci Mag* **308**:1034–1036.

29 Chu JY et al. 1998. Hypothetical ancestral migration routes to the Far East. *Proc Natl Acad Sci USA* **95**:11763–11768.

30 Gerritsen R. 2011. *Beyond the Frontier: Explorations in Ethnohistory*. Batavia Online Publishing, Canberra, Australia.

31 Adcock GJ et al. 2001. Mitochondrial DNA sequences in ancient Australians: implications for modern human origins. *Proc Natl Acad Sci USA* **98**:537–542.

32 Gyllensten U et al. 2003. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res* **13**:1600–1606.

33 Pierson MJ et al. 2006. Deciphering past human population movements in Oceania: probably optimal trees of 127 mtDNA genomes. *Mol Biol Evol* **23**:1966–1975.

34 Zhong H et al. 2010. Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia. *J Hum Genet* **55**:428–435.

35 Scheinfeldt L et al. 2006. Unexpected NRY chromosome variation in Northern Island Melanesia. *Mol Biol Evol* **23**:1628–1641.

36 Hudjashov G et al. 2007. Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *Proc Natl Acad Sci USA* **104**:8726–8730.

37 Pugach I et al. 2013. Genome-wide data substantiate Holocene gene flow from India to Australia. *Proc Natl Acad Sci USA* **110**:1803–1808.

38 Currat M, Excoffier L. 2004. Modern humans did not admix with Neanderthals during their range expansion into Europe. *PLoS Biol* **2**:e421.

39 Barbujani G et al. 1998. Evidence for Paleolithic and Neolithic gene flow in Europe. *Am J Hum Genet* **62**:488–492.

40 Regueiro M et al. 2013. On the origins, rapid expansion and genetic diversity of Native Americans from hunting-gatherers to agriculturalists. *Am J Phys Anthropol* **150**:333–348.

41 Malyarchuk B et al. 2011. Ancient links between Siberians and Native Americans revealed by subtyping the Y chromosome haplogroup Q1a. *J Hum Genet* **56**:583–588.

42 Dulik MC et al. 2012. Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and indigenous Altaians. *Am J Hum Genet* **90**:229–246.

43 Gibbons A. 1998. Mother tongues trace steps of earliest Americans. *Science* **279**:1306–1307.

44 Bodner M et al. 2012. Rapid coastal spread of first Americans: novel insights from South America's southern cone mitochondrial genomes. *Genome Res* **22**:811–820.

45 Perez SI et al. 2009. Discrepancy between cranial and DNA data of early Americans: implications for American peopling. *PLoS One* **4**:e57.

46 Torroni A et al. 1993. mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am J Hum Genet* **53**:591–608.

47 Greenberg JH et al. 1986. The settlement of the Americas: a comparison of the linguistic, dental and genetic evidence. *Curr Anthropol* **27**:477–497.

48 Bonatto S, Salzano M. 1997. A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proc Natl Acad Sci USA* **94**:1866–1871.

49 Karafet TM. 1999. Ancestral Asian source(s) of New World Y chromosome founder haplotypes. *Am J Hum Genet* **64**:817–831.

50 Schurr TG, Sherry ST. 2004. Mitochondrial DNA and Y-chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. *Am J Hum Biol* **16**:420–439.

51 Reich D et al. 2012. Reconstructing Native American population history. *Nature* **488**:370–374.

52 Bocquet-Appel J-P. 2011. When the world's population took off: the springboard of the Neolithic demographic transition, *Science* **333**:560–561.

53 Diamond J, Bellwood P. 2003. Farmers and their languages: the first expansions. *Science* **300**:597–603.

54 Fuller DQ. 2007. Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the Old World. *Ann Bot* **100**:903–924.

55 Jiao T. 2007. *The Neolithic of Southeast China: Cultural Transformation and Regional Interaction on the Coast*. Cambria Press, Amherst, New York.

56 Li H. 2008. Paternal genetic affinity between western Austronesians and Daic populations. *BMC Evol Biol* **8**:146–157.

57 Ruhlen M. 1994. *The Origin of Language: Tracing the Origin of the Mother Tongue*. John Wiley & Sons, Inc., New York.

58 Gray RD, Jordan FM. 2000. Language trees support the express-train sequence of Austronesian expansion. *Nature* **405**:1052–1055.

59 Kirch PV. 1998. Lapita and its aftermath: the Austronesian settlement of Oceania. In: Goodenough WH (ed.), *Prehistoric Settlement of the Pacific*. American Philosophical Society, Philadelphia, PA.

60 Kayser M et al. 2008. Genome-wide analysis indicates more Asian than Melanesian ancestry of Polynesians. *Am J Hum Genet* **82**:194–198.

61 Diamond JM. 1988. Express train to Polynesia. *Nature* **336**:307–308.

62 Oppenheimer SJ, Richards M. 2001. Polynesian origins: slow boat to Melanesia? *Nature* **410**:166–167.

63 Melton T et al. 1995. Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *Am J Hum Genet* **57**:403–414.

64 Trejaut JA et al. 2005. Traces of archaic mitochondrial lineages in Austronesian-speaking Formosan populations. *PLoS Biol* **3**:e247.

65 Kayser M et al. 2008. The impact of the Austronesian expansion: evidence from mtDNA and Y chromosome diversity in the Admiralty Islands of Melanesia. *Mol Biol Evol* **25**:1362–1374.

66 Soares P et al. 2011. Ancient voyaging and Polynesian origins. *Am J Hum Genet* **88**:239–247.

67 Su B et al. 2000. Polynesian origins: insights from the Y chromosome. *Proc Natl Acad Sci USA* **97**:8225–8228.

68 Kayser M et al. 2000. Melanesian origin of Polynesian Y chromosomes. *Curr Biol* **10**:1237–1246.

69 Karafet T et al. 2010. Major east–west division underlies Y chromosome stratification across Indonesia. *Mol Biol Evol* **27**:1833–1844.

70 Mirabal S et al. 2012. Increased Y-chromosome resolution of haplogroup O suggests genetic ties between the Ami aborigines of Taiwan and the Polynesian Islands of Samoa and Tonga. *Gene* **492**:339–348.

71 Friedlaender J et al. 2008. The genetic structure of Pacific Islanders. *PLoS Genet* **4**:e19.

72 Lipson M. 2014. Reconstructing Austronesian population history in Island Southeast Asia. *Nat Commun* **5**:1–7.

73 Moreno-Mayar V et al. 2014. Genome-wide ancestry patterns in Rapanui suggest pre-European admixture with Native Americans. *Curr Biol* **24**:1–8.

74 Rieth TM, Hunt TL. 2008. A radiocarbon chronology for Samoan prehistory. *J Archaeol Sci* **35**:1901–1927.

75 Ponting C. 2007. *A Green History of the World: The Environment and the Collapse of Great Civilizations*. Penguin Press, New York.

76 Kirch PV. 1980. *Evolution of the Polynesia Chiefdoms*, Cambridge University Press, Cambridge, MA.

77 Oskarsson MCR et al. 2011. Mitochondrial DNA data indicate an introduction through Mainland Southeast Asia for Australian dingoes and Polynesian domestic dogs. *Proc R Soc B* **279**:967–974.

78 Bellwood P. 1997. *Prehistory of the Indo-Malaysian Archipelago*. University of Hawaii Press, Honolulu, HI.

79 Friedlaender JS, Matisoo-Smith E. 2007. Animal translocations, genetic variation and the human settlement of the Pacific. In: Friedlaender JS (ed.), *Genes, Language and Culture History in the Southwest Pacific*. Oxford University Press, Oxford, UK.

80 Giuffra E. 2000. The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* **154**:1785–1791.

81 Greger L. 2007. Phylogeny and ancient DNA of *Sus* provides insights into Neolithic expansion in Island Southeast Asia and Oceania. *Proc Natl Acad Sci USA* **104**:4834–4839.

82 West B, Zhou BX. 1988. Did chickens go north? New evidence for domestication. *J Archaeol Sci* **14**:515–533.

83 Zeder MA et al. 2006. Documenting domestication: the intersection of genetics and archaeology. *Trends Genet* **22**:139–155.

84 Thomson VA et al. 2014. Using ancient DNA to study the origins and dispersal of ancestral Polynesian chickens across the Pacific. *Proc Natl Acad Sci USA* **111**:4826–4831.

85 Matisoo-Smith E et al. 1998. Patterns of prehistoric human mobility in Polynesia indicated by mtDNA from the Pacific rat. *Proc Natl Acad Sci USA* **95**:15145–15150.

86 Matisoo-Smith E, Robins JH. 2004. Origins and dispersals of Pacific peoples: evidence from mtDNA phylogenies of the Pacific rat. *Proc Natl Acad Sci USA* **101**:9167–9172.

87 Wilmshurst JM et al. 2008. Dating the late prehistoric dispersal of Polynesians to New Zealand using the commensal Pacific rat. *Proc Natl Acad Sci USA* **105**:7676–7680.

88 Van Tilburg JA. 1994. *Easter Island: Archaeology, Ecology and Culture*. Smithsonian Institution Press, Washington, DC.

89 Hather J, Kirch PV. 1991. Prehistoric sweet potato (*Ipomoea batatas*) from Mangaia Island, Central Polynesia. *Antiquity* **65**:887–893.

90 Storey AA, Matisoo-Smith EA. 2014. No evidence against Polynesian dispersal of chickens to pre-Columbian South America. *Proc Natl Acad Sci USA* **111**:E3583.

91 Thorsby E et al. 2009. Further evidence of an Amerindian contribution to the Polynesian gene pool on Easter Island. *Tissue Antigens* **73**:582–585.

92 Jones TL et al. 2011. *Polynesians in America: Pre-Columbian Contact with the New World*. Altamira Press, Lanham, MD.

CHAPTER 10

Bioethics: consequences and implications of genetic technology on human evolution

If ethics is not the engine of success, in the train of growth, it sure is a guard, with a flag, which may be green, or at times red.

—*Priyavrat Thareja* [1]

Summary

Two major forms of human evolution, social evolution and biological evolution, have shaped biology, and formed our current social structure. Social evolution sometimes referred to as cultural evolution involves demographics, technology, economic forces, and religious and ethical beliefs. While the principal forces driving biological evolution are genetic variation, drift, and natural selection, social evolution is driven by environment, material conditions, technology, and spirituality, which maintain overall fitness within society [2]. Part of our social evolution is redefining what we believe to be right and wrong and setting standards for our behavior between individuals and populations. The evolving concepts of justice and morality have included our treatment of individuals at various levels including the biological.

The study of bioethics began in earnest shortly after World War II as a result of concerns over medical experimentation. Since its inception, the field has struggled to keep up with genetic technology. Mandatory DNA testing of crime suspects and military personnel and DNA dragnets have led to concerns over privacy. More than a dozen nations are using genetic information to screen their populations in "biobank" projects raising concerns about how this information will be used and who has access to it.

For centuries humans have been selecting plants and animals for specific genetic characteristics directing the development of current agricultural practices. The advances in agriculture and technology led to the current social structure. Just as agriculture and the formation of communities led to the increased pace of human evolution, technology and modern culture may increase the current rate of human evolution.

Today in developing countries almost everyone reaches reproductive age. Current and evolving medical technologies will continue to ensure a longer life span and the ability to reproduce. Gene therapy, vaccines, proteomics, directed evolution of proteins, evolutionary medicine, personalized medicine, genomics, and metagenomics are already a focus of medical research and will continue to develop to eventually become part of common medical practice. Natural selection is relatively slow in comparison with cultural, technical, and environmental changes imposed by humans. Diseases will continue to exert a strong selection pressure, but many of these pressures may be solved by modern medicine (specific drugs and genetic modification) before natural selection acts. Many view these technological advances as a natural process by which we can better society through euphenics, but others see this as a cryptic approach to eugenics.

In this chapter, we will consider the debate over the evolution of morality and whether it is a result of learned behavior or a product of natural selection, or a combination of both. The chapter will briefly describe some of the Western philosophers who shaped our moral and ethical beliefs, and then focus on how social evolution and biological evolution have both contributed to our current social structure and our value and belief systems, especially bioethical concerns that have arisen over the last century due to modern technology and our increased understanding of what makes us human. We will examine the technology and ethical issues behind assisted reproduction, genetic enhancement, gene therapy, stem cell therapy, genetic privacy, genetic screening, and DNA profiling, and point out how these may be affecting human evolution.

Genomes, Evolution, and Culture: Past, Present, and Future of Humankind, First Edition.
Rene J. Herrera, Ralph Garcia-Bertrand, and Francisco M. Salzano.

© 2016 John Wiley & Sons, Ltd. Published 2016 by John Wiley & Sons, Ltd.

Social and biological evolution

From a historical perspective, there are several major social events that led to dramatic shifts that contributed to biological evolution. Changes from hunting and gathering to agriculture, technological advances, changes in climate, the development of trading and economic systems, the development of religion and ethics, land acquisition, and the increase in population size all contributed to social structure and in many cases modified human biological evolution. Population size alone contributed to a need for more resources, more cooperation, and even a change in the epidemiology of diseases, all factors under selection. There are a number of behaviors that influence or have influenced allele frequencies. Nonrandom mating has had a direct influence on changing allele frequencies, for example, within social hierarchies the ruling class has had children with females of the underclass, rape during wartime, and populations that practice consanguinity in small isolated communities. Migration has also led to allelic variation through people's ability or inability to adapt to parasites, diseases, new foods, climate changes, and mixed resources. Populations that have been isolated by choice or by force have also resulted in allelic variation through genetic drift.

The information and behavior transmitted by learning is probably the most extraordinary process developed through human evolution. The development of symbolism, religion, and superstition and the realization of mortality and self-awareness have distinguished us from our primate ancestors. The development of culture allowed for social organization, child-rearing practices, mate selection, social selection, art, violence, and the eventual development of language and speech.

Although some aspects of social and biological evolution may seem far removed from each other, they acted concurrently to shape our current state and continue to act today to influence social life and our biology. Mating behavior and gender roles [3], parental roles, and access to food and technology all have direct effects on biological fitness [4]. Less apparent are the effects of violence, politics, economics, race, ethnicity, ethics, and religious beliefs. However, current technological advances, economic status, politics, ethnicity, ethics, and religion may be playing more of a role in biological evolution than ever before.

Growing concern in the United States and Europe over technology and security, tension over the immigration of people of specific race and ethnicity, and a growing dichotomy in economic status have led to beliefs that there are lives that may not have similar value. All of these forces have had an effect on human evolution.

Social science research in the 1950s and 1960s associated behavioral problems with economics, environment, and social status. However, as genetic technology advanced and scientists such as Seymour Benzer and T.H. Morgan began to map behavioral genes in *Drosophila*, and other scientists began to associate human genes with diseases and mental illness, the public jumped to the conclusion that genetics controlled human behavior and our fate, leading to the idea of determinism. This in turn led to social and political change from a 1950's and 1960's philosophy of "there are people problems" to a current philosophy that "there are problem people." As genetic technology continued to progress, even some prominent scientists began to promote determinism. In 1989, James Watson stated [5] "We used to think our fate was in the stars. Now we know, in large measure, our fate is in our genes." In the 1980s, determinism became an argument against welfare, and prison reform, and blaming groups of individuals helped to absolve society from any responsibility. Ironically, during the past decade as the thousands of human genome sequences revealed our similarities, many minority groups have begun to see the information as a further way to discriminate by questioning civil and reproductive rights. Unfortunately, sociological studies over the last 18 years have confirmed this fear by demonstrating that genetic information can generate divisive beliefs and attitudes toward race [6].

Overview of ethics and philosophical influences on Western ethics

A discussion of current practices and concerns in bioethics, and how they affect human evolution, cannot begin without an overview of the terminology and a brief history of Western philosophers who have contributed ideas that have shaped our current perspectives on ethics, morality, and justice.

The term ethics typically refers to the branch of philosophy associated with moral behavior (ability to distinguish right from wrong, good from bad, and virtue from vice). Ethics is typically described as a set of principles guiding behavior and distinguishing right from wrong,

good from evil, and justice from injustice. Ethical principles vary from culture to culture and even among individuals from similar cultures. Ethical principles are based on context, past experience, and beliefs. From an academic point of view, ethics is a part of philosophy and is divided into three areas of study based on the application of ethical principles: (1) meta-ethics is the abstract theoretical study of ethics dealing with the definitions of right and wrong, and the foundations of ethics; (2) normative ethics looks at the principles to distinguish between good and bad actions, and looks at the course of action determined within a society based on virtue, duty, and consequences; and (3) applied ethics deals with controversial topics, asking what is the right thing to do given a specific situation [7].

The methodology used in making major ethical decisions is complex and typically involves academics, journalists, media, and politicians. A full discussion of methodology is beyond the focus of this chapter; however, it is good to know the basic approaches used in making ethical decisions.

The fundamental approach looks at ethical theory and then applies it to the case details to arrive at a conclusion. The case-based approach looks first at the case details and then through intuition comes up with a conclusion. A problem with the fundamental approach is that theories do not always fit case details. Problems with the case-based approach are that bias may be involved since not all cases are similar, and it is difficult to generalize intuition.

Western ethical practices and the contemporary study of ethics are associated with philosophers of the past who influenced and shaped the way we think about ethics, morality, and justice. These philosophers all expressed different ideas and their contributions can be seen today in modern business, politics, bioethics, and the justice system [8]. The list of philosophers is in chronological order to facilitate an understanding of how our principles have changed over time. In each case, following the description of their contributions, there are questions about their beliefs and how their principles relate to normative and applied ethics. By addressing these questions, it is easy to see that no one philosophy can be utilized when making ethical decisions.

Socrates

An ancient philosopher whose teaching methods centered on inductive reasoning (critical thinking) and who is credited with the Socratic methods used by scientists and educators is often referred to as the Father of Ethics. This title is controversial since Socrates left no written record of his philosophy, and instead his teachings are relayed in the writings of his students after his death. Plato is credited with the dissemination of Socratic theory, as it relates to ethics, through his Socratic dialogues. In the Socratic dialogues, Plato describes Socrates as a person living a simplistic way of life in pursuit of virtue, and that Socrates taught that the pursuit of virtue outweighs all other pursuits. According to Plato, Socrates described virtue as the source of happiness and denied that virtue could be taught and instead described it as a divine bequest. Socrates believed that no one desires to do wrong or knowingly desires evil.

If Socrates was right, do people have free will when it comes to doing wrong or evil or is it an inherent property that they cannot control? Is everyone given equal virtue?

Aristotle

Aristotle, who was a student of Plato, had very different beliefs from his predecessors about justice and virtue. Aristotle believed that justice was about giving people their due, what they deserve. When considering matters of distribution, Aristotle argued that one must consider the goal, the telos, or the purpose of what is being distributed. For Aristotle, justice is inherently unequal and tied to merit, virtue, and honor—fitting a person's virtues with an appropriate role. Citizens who contribute most to the purpose of the community are the ones who should be most rewarded. This is not utilitarian because those most worthy could be the minority, and those rewarded serve the greater community. But how do we know the purpose of a community or a practice? How does Aristotle address the issue of individual rights and the freedom to choose? If our place in society is determined by where we best fit, doesn't that eliminate personal choice? One of the most glaring objections to Aristotle's views on freedom is his defense of slavery as a fitting social role for certain human beings.

John Locke

The philosopher John Locke (Figure 10.1) would be described today as a moderate libertarian. He believed that individuals have fundamental rights that no government can take away. The rights to life, liberty, and property are a fundamental part of nature, and were given before government and laws were created.

According to Locke, our natural rights can be neither given up nor taken away. Locke believed that when we choose to live in a society we give our consent to obey the laws passed by a majority. Therefore, laws are legitimate and compatible with individual rights, as long as they apply to everyone. Locke also believed that people were born with a "blank slate" and that behavior was learned not innate.

If we all have rights to life, liberty, and property, how do we equitably distribute liberty and property?

Immanuel Kant

Immanuel Kant was an academic, challenging, influential philosopher who had a distinct philosophy known today as Kantianism. Kant rejected utilitarianism. He argued that each of us has certain fundamental duties and rights that take precedence over maximizing utility. Kant rejected the notion that morality is about calculating consequences and final outcomes. He believed that when we do something simply because it is right, only then do our actions have moral worth. For example, a person who passes up the chance to cheat only because his/her reputation might suffer or he/she might go to jail would not be acting morally. According to Kant, the person's action has no moral worth, because he/she did the right thing for the wrong reason.

Immanuel Kant says that what confers moral worth is our capacity to rise above self-interest and inclination and to act out of duty. Kant's test for determining a moral action is to identify the principle expressed in our action and then ask whether that principle could ever become a universal law that everyone could accept. Kant's categorical imperative [9] states that you should "act only on the maxim (principle) whereby you can at the same time will it to be universal law." Immanuel Kant's stringent theory of morality allows for no exceptions. Kant believed that telling a lie, even a white lie, is a violation of one's own dignity.

Can there ever be principles that everyone can act on universally? Are there specific situations when a white lie is appropriate after considering the long-term consequences? If cheating on an exam does not hurt anyone, what is the cost to society?

Jeremy Bentham

Jeremy Bentham proposed that morality should be guided by utilitarian principles, the doctrine that the right thing to do is whatever produces "the greatest good for the greatest number." Bentham argued that human happiness is the achievement of pleasure and the avoidance of pain and that happiness in the community is the sum of individual interests. Choices should be made on the basis of the amount of pleasure or pain brought about by the choice. Companies and governments often use utilitarian logic under the name of "cost–benefit analysis."

John Stuart Mill

John Stuart Mill (Figure 10.2) was a utilitarian philosopher who defended utilitarianism against the objections raised by critics. Mill concluded that many people misunderstood utilitarianism by assuming that utility was in opposition to pleasure. He defined utility as the principle for the greatest happiness, and argued that seeking "the greatest good for the greatest number" is compatible with protecting individual rights, because protecting individual rights will, in general, maximize utility. He felt that utilitarianism can make room for a distinction between higher and lower pleasures because the higher pleasure is always the pleasure preferred by a well-informed majority, and people best qualified to distinguish higher and lower pleasures are those who have experienced them.

Although this is a popular Western philosophy, there are still concerns: What happens to minority concerns? Is the majority always well informed?

John Rawls

The modern philosopher John Rawls believed in an egalitarian society and presented a theory of a "hypothetical social contract" (Rawlsianism). John Rawls claimed that "Justice is the first virtue of social institutions, as truth is of systems of thought."

Rawls argued that principles of justice are the outcome of a special kind of agreement that we would all agree to if we had to choose rules for society and no one had any unfair power or advantage. According to Rawls, the only way to ensure ethical decisions is to imagine a scenario where no one knows his or her age, sex, race, intelligence, strength, social position, economic status, religion, or even his or her goals. Rawls calls this hypothetical situation a "veil of ignorance."

What principles would we agree to behind the "veil of ignorance"? And would these principles be fair?

John Rawls says that we should answer these questions by asking what principles you would choose to govern the distribution of wealth (and other services such as health

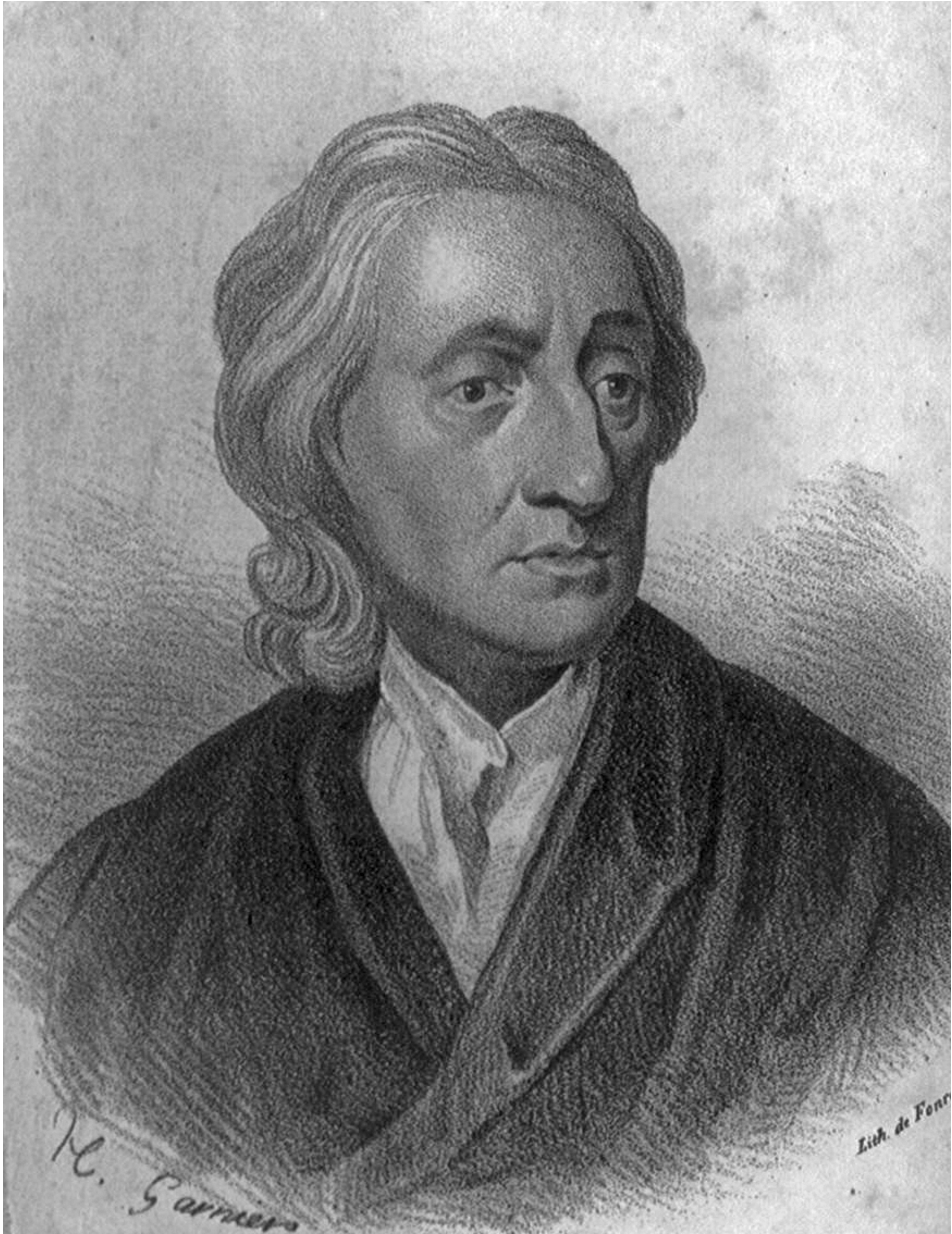


Figure 10.1 John Locke (1632–1704) was an English philosopher. His political theories helped shape the Declaration of Independence ideas on law and private property, and the U.S. Constitution's separation of church and state.

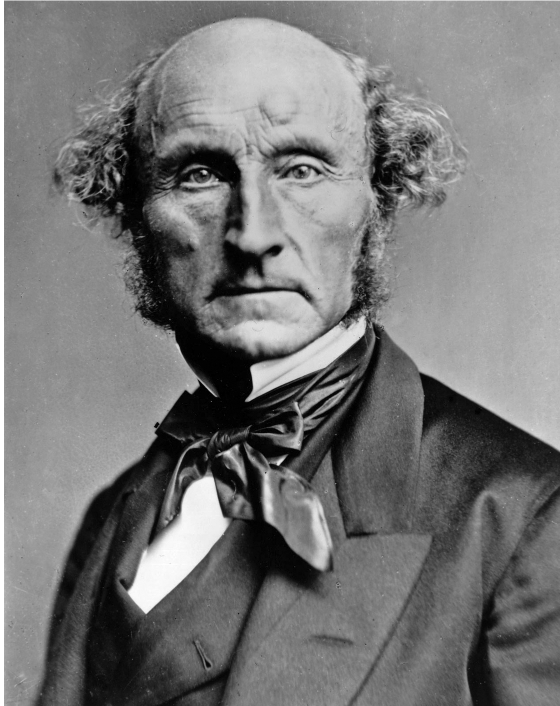


Figure 10.2 John Stuart Mill (1806–1873) was an English philosopher and a student of Jeremy Bentham. He is considered one of the most influential English-speaking philosophers of the 19th century. (*Source*: London Stereoscopic Company (Hulton Archive), https://commons.wikimedia.org/wiki/File%3AJohn_Stuart_Mill_by_London_Stereoscopic_Company%2C_c1870.jpg, public domain.)

care) if you did not know who you were, whether you grew up in privilege or in poverty. Wouldn't you want an equal distribution of resources, or one that maximally benefits whomever happens to be disadvantaged? Rawls argues that even meritocracy doesn't go far enough in leveling the playing field because those who are naturally gifted will get ahead, and the naturally gifted can't necessarily claim credit because sometimes their success depends on factors beyond their control (genetic inheritance and/or environment). Rawls argues that distributed justice is the just thing to do.

Rawls suggested that the best methodology for making ethical decisions is the idea of reflective equilibrium that takes into account both fundamental and case-based approaches by stating that theories can be modified and that intuition can sometimes be wrong.

Shouldn't people be able to use their natural gifts to benefit themselves and their family? How do these beliefs affect capitalism and the economy? What is the incentive to better yourself in society or to develop new technology if you make judgments without regard to how they will affect your outcome?

Robert Nozick

A modern-day libertarian philosopher Robert Nozick argued that government shouldn't have the power to enact laws that (1) protect people from themselves, (2) impose individual moral values on society as a whole, or (3) redistribute income from the rich to the poor. He believed that the redistribution of wealth for housing, health care, and education for the poor is a form of coercion. Libertarianism (neoclassical liberalism) promotes individual rights, and only a minimal government.

Don't most poor people need social services in order to survive? By agreeing to live in a society, don't you have to do what is best for the society as a whole? Don't many rich people acquire their wealth through luck (lottery or inheritance) and do not earn it?

Michael Sandel

A Harvard professor, modern-day philosopher, and communitarian. Communitarians argue that, in addition to voluntary and universal duties, we also have obligations of membership, solidarity, and loyalty. These obligations are not necessarily based on consent. We inherit our past, and our identities, from our family, city, or country. Because we have an obligation to respect these ties, the idea of a "veil of ignorance" is not plausible. But what happens if our obligations to our family or community come into conflict with our universal obligations to humanity? Do we owe more to our country than to citizens of other countries?

Is patriotism a virtue, or a prejudice for one's own kind? How do we define universal human rights?

Broad questions arise when reflecting on the philosophical doctrines of ethics and justice. How should income, wealth, health care, technological advances, and opportunities in life be distributed? Is it necessary to determine what constitutes a "good life" in order to decide what rights people have and what is just? If so, what is the good life? What is the relationship between the law, science, religion, and morality? Should we use the principles of libertarians, egalitarians, utilitarians, or the meritocratic system when making decisions? Each system has something to offer, but each is also wrong insofar as each places absolute values on their beliefs and

no single philosophy covers all situations [8]. However, some people do ascribe to a single theory, such as utilitarianism, and live their lives accordingly.

Evolution of ethics and morality

A major academic dilemma is the fundamental nature of ethics and morality. The search for the nature of ethics and morality falls under two academic disciplines, biology and philosophy, that both try to explain the moral actions of individuals. Philosophers are typically concerned with the justification for morality and the principles behind decision making; biologists, on the other hand, are concerned with causality and the biological mechanisms responsible for morality. The broad questions addressed in the quest to understand the development of ethics and morality in humans are as follows: Why should we be good and not be evil? What is the nature and source of morality? Why do we appear to be the only animals capable of developing ethical principles and making moral decisions? Do environmental pressures on our ancestors explain morality or is it a recent phenomenon? Was morality just something that helped humans make more babies [10]?

Many philosophers explain morality as part of human nature, feelings dependent on experiences, and the result of revelation, whereas many biologists see it as a result of natural selection influenced by environment.

As previously described, ethics involves a set of principles guiding behavior and distinguishing right from wrong. The evidence for a behavior being biological is a consistent specific behavior that is transmitted from generation to generation regardless of culture or experience. The evidence for behavior being a result of nurture is the ability of behavior to vary with culture, personal interactions, and experience [11]. One subdiscipline in biology that tries to address the source of morality is behavioral genetics.

Behavioral genetics is a combination of biology and experimental psychology that has found that certain behaviors are constant across species. Several behaviors have been shown to have a biological source and function. Examples in humans include risky behavior, altruism, and even depression. Depression, for example, has been associated with an adaptive biological function of helping to fight off existing infections and avoid new ones [12]. Altruistic behavior toward family members and altruistic behavior in species that congregate are both easily explained by natural selection and long-term survival mechanisms (protection of the family or group is related to self-protection). However, how can traits such as morality or self-sacrifice, which go beyond altruism, be explained by natural selection?

One of the major arguments against natural selection being a force for morality is the question of how could early humans judge an action such as morality as part of a long-term survival mechanism? Because of this, some philosophers argue that ethics and morality have no direct link to survival and so there is no evolutionary or genetic foundation for morality, instead it is just part of human nature and directed by experience. The experiences driving morality can include religious beliefs, cultural experiences, reflection on past events, and the ability to think critically. Reflection, reasoning, and judgment are all autonomous behaviors and have no evolutionary basis except for the development of intellect. Even though evolution may provide the ability to reason, moral decisions are made autonomously. One of the major dilemmas is that if evolution had a role in developing morality what role does it play in current reality and a person's or group's perception of morality today? How could a trait such as morality change with other evolutionary advances to deal with current reality and feelings that can change? Rachels [13] argues that you cannot rely on a form of special evolution and special evolutionary principles to create a moral status for humans, since no other animals have this special aptitude. Since there is no other biological evidence that provides a special separation between humans and other animals, morality must be a case of a human-specific nonbiological phenomenon.

Edward O. Wilson [14] argued that morality arose from a time when it was essential for survival of the group, and that immoral behavior or egoistic behavior would have been removed by natural selection (through removal of the group). He stated that morality is deep seated in evolution and genetically wired so that we have no choice but to act morally (or at least consider it). Wilson argued that there was a biological basis for hate, love, and fear, and because these emotions come from the brain's limbic system and hypothalamus this proved that they have an evolutionary basis. Philosophers such as Michael Ruse [15] also see morality as something fashioned by natural selection as an adaptation that ensured our survival in a societal context.

Philosopher Herbert Spencer also argued for the evolutionary development of morality. He argued that the human drive was toward a gain of pleasure and the avoidance of pain. This egoistic tendency drove evolution and provided self-gratification to the individual, as well as gratification and pleasure from giving to others. Society and cooperation among individuals for the good of the group (and self-interest) developed the principles of altruism and equity and held egotism in check [16].

If morality and ethics are a result of natural selection, how could they and other complex behaviors have evolved? Neuroscientists are finding that many behaviors have a neurological component and that some complex behaviors that we do not share with other animals may be the result of our larger brains. Several criteria have been proposed that could have influenced the evolution of brain expansion. Changes in the development of jaw muscles associated with mutations in the MYH16 gene that rendered it inactive have been proposed to allow for encephalization. The ASPM, MCPH1, CDK5RAP2, and CENPJ genes have also been implicated in brain size [17]. Increased group size requiring complex relationships and interactions (moral behavior) would have favored increased brain size to accommodate these cultural changes. Complex subsistence patterns that consisted of different plant and animal products collected over a wider range and differing seasons would have favored a larger brain size to identify edible food sources.

The Human Genome Project (HGP), next-generation sequencing, and genome-wide association studies have led to a better understanding of the role of genes in behavior. The complexity of behavior has led to a systems approach to understanding the mechanisms behind it. The systems approach looks at behavior as a whole from neuroscience, biology, bioinformatics, psychology, and philosophical perspectives. This approach has allowed for the study of behavior at the molecular level and an understanding of how animals can process information and translate it into behavior [6]. Genes are responsible for the production of neurotransmitters, neuronal circuits, and their ability to interact with each other, and various organs and tissues.

Epigenetic experiments in rats by Weaver [18] and Zhang and Meaney [19] showed that parental behavior could influence progeny behavior through epigenetic modifications. Monozygotic twin studies showed epigenetics involved in personality traits, and other studies have shown epigenetics involved in learning and drug addiction. Epigenome-wide association studies (EWAS) have been suggested to look for epigenetic variation that may be responsible for a variety of complex traits.

Today, most would agree that there appears to be a biological component to morality and that behaviors can be learned and shaped by natural and artificial selection. Just like morality and ethical behavior, most behaviors show a continuous distribution and are difficult to specifically define. Research suggests that behavior has a genetic and an environmental component with genes imparting physiological and psychological limits on expression. Behavior can be dramatically influenced by environment but is not purely environmental. There is growing acceptance that ethics and moral behavior have a biological component and were most likely shaped by natural selection. If natural selection was able to shape our perceptions of objects and to guide our understanding of complex truths (chemistry, physics, and evolution), then why could it not have shaped our grasp of moral truths? Perhaps evolution did not shape our moral truths but instead shaped our emotions and behavioral instincts that helped develop our moral beliefs, or perhaps evolution was more focused on reproduction and survival and morality was a by-product of cultural experience [17]. There is no firm acceptance of which positions are most acceptable.

The history and beginning of modern-day bioethics

Historically, social evolution has always included some form of perceived ethical and just behavior; however, the definitions and practices of ethics, morality, and justice have changed dramatically over time. It is easy to reflect on past condoned practices and behaviors that included the Roman gladiators fighting for the amusement of the populace, the Christian crusades, and slavery in the United States and elsewhere, and realize these were gross misinterpretations of justice and morality. More recently, we have encountered acts of injustice that question our moral values. The eugenic movement of the early 1900s, the Japanese and other interment camps, and the Tuskegee experiments are just a few examples.

The rediscovery of Mendel's work on inheritance and Darwin's theory of evolution through natural selection contributed to the discovery of several human genetic diseases and traits in the early 1900s. These included

hemophilia, color blindness, and ABO blood type. These initial discoveries led to the idea that genetics could be responsible for most behavioral and disease traits that could be perpetuated through inheritance. With this in mind, geneticists began to perform pedigree analysis on families that had a history of alcoholism, mental retardation, depression, pauperism, and criminality. The fact that some of these conditions were seen to appear more often in these families than the general population led to the idea that these traits could be prevented and even eliminated by not allowing families with these traits to reproduce. Several scientists argued that for centuries humans had been selecting plants and animals for specific genetic characteristics directing the development of current agricultural practices, and that those practices led to advances in agriculture, and social structure, so why not apply the same selective processes to humans? This was not a new idea. Francis Galton proposed the term eugenics (meaning good at birth) in 1883, as a way to improve human society. He proposed what today is known as positive eugenics, in that he encouraged those people with talent and good characteristics to mate. He felt that eugenics was justified based on scientific knowledge and could help to solve a variety of social problems. Galton did not endorse the idea of negative eugenics (forcing those without good traits to not reproduce), but instead suggested that people think about the greater good and practice eugenic principles [20]. The term eugenics took on a negative connotation after practices in the United States and Nazi Germany.

In the United States, a number of scientists concluded that limited reproductive practices could improve the human condition and direct future human evolution. In 1910, Charles Davenport led the eugenic movement stating that it was "the science of the improvement of the human race by better breeding." Because of the eugenic movement, laws were passed in many states requiring the sterilization of individuals with certain genetic disorders, those convicted of certain crimes, and those deemed to be mentally unfit. In addition, after hearing statements from eugenic experts Congress passed the Immigration Reduction Act of 1924 restricting the immigration of southern and eastern Europeans, Russians, and Asians (deemed to be genetically inferior). In 1927, when the constitutional rights of individuals were questioned, the U.S. Supreme Court upheld the right of states to use eugenic sterilization. After the Supreme Court decision, 33 states passed laws to perform forced sterilization of individuals diagnosed as genetically unfit, and prohibit marriage between "social misfits." The social misfit diagnosis included alcoholism, promiscuousness, depression, and criminality [21]. This especially affected the poor and uneducated who had very little recourse when facing accusations (Table 10.1).

In 1933, Nazi Germany used the philosophy of eugenics to justify medical experimentation and extermination of non-White and Jewish citizens who were deemed unfit. The impetus was the purification and preservation of the Aryan race and the justification was based on practices already in use in the United States and Britain. Further justification was described as mercy killing for those regarded as having lives not worth living. These included individuals with physical deformities, disabilities, mental illness, and disease. By the late 1930s, recognition of the atrocities of the German government led to reconsideration of eugenic laws in the United States and a decline in eugenic practices.

Modern bioethics in the United States and Europe began shortly after World War II as a result of concerns over medical experimentation by the Germans and the eugenic laws and practices in the United States. Even with the acknowledgement that eugenic practices were immoral, and that the Jewish population had been harmed by discriminatory practices, society did not change immediately as seen in the infamous Tuskegee syphilis experiment from 1932 to 1972 (see Box 10.1).

Since its inception, the field of bioethics has struggled to keep up with technological advances in genetics and molecular biology [24]. Reproductive technologies, gene therapy, vaccines, proteomics, genomics, directed evolution of proteins, evolutionary medicine, and personalized medicine are already a focus of medical research and will continue to develop to eventually become part of common medical practice.

Although technologies such as genome-wide association studies (GWAS), next-generation sequencing, proteomics, and genomics promise to bring relief to many by discovering genes and treatments for complex diseases, the long-term impacts on human evolution remain unclear. Even the immediate social and ethical impacts of GWAS and genomics are in question since many studies are revealing the targets of natural selection. Vitti et al. [25] discuss the potential negative impacts and potential discrimination that could ensue unless scientists are careful to steer the results of human

Table 10.1 History of eugenics and the association of genes with behavior in the United States.

1883	Francis Galton proposes the term eugenics as a way to improve human society. He proposed positive eugenics, encouraging those people with talent and good characteristics to mate.
1910	Virginia passes a state law requiring sterilization of poor women. Charles Davenport led the eugenic movement stating that it was "the science of the improvement of the human race by better breeding."
1914	National Conference on Race Betterment sponsored by the American Eugenics Society. Babies and school-aged children were judged based on characteristics (weight, height, teeth, etc.) and awarded prizes.
1920	Kansas State Fair has a poster announcing "How long are we to be careful about the pedigree of our animals and leave the ancestry of our children to chance."
1922	William Sadler, Professor at the University of Chicago Medical School and Head of the Eugenics Society, states that inherited traits are responsible for criminality and feeblemindedness. He states that eugenics is a benevolent practice keeping the unfortunate from society.
1924	Immigration Reduction Act restricts the immigration of southern and eastern Europeans, Russians, and Asians (deemed to be genetically inferior).
1927	The U.S. Supreme court upheld the right of states to use eugenic sterilization.
1932	The Tuskegee syphilis experiment was conducted by the U.S. Public Health Service. The experiment started in 1932 and lasted until 1972. The experiment exposed poor African American subjects to syphilis without treatment in order to study the effects of the disease.
1960s	Patricia Jacobs associates criminal behavior with males carrying an extra Y chromosome (XYY).
1961	At Macy Conference, scientists discuss the inefficiency of natural selection due to modern technology and modern medicine. Some state that inbreeding may be good in order to expose recessive lethal alleles and eliminate them from the population.
1974	Frederick Osborn, President of the Society for Social Biology, states that progress is being made because families of stature were having more children.
1990s	Numerous magazine articles talk about "better babies" being highly desirable, born criminals are considered a product of the underclass, DNA provides personal identity and determines fate (genetic essentialism).
1992	National Academy of Sciences and the National Research Council propose the Violence Prevention Initiative. The agencies issued a report calling for more attention to "biological and genetic factors in violent crime."
1992	The director of NIMH (Goodwin) proposes using genetics to detect biological markers for violence in at-risk inner-city children and treating them with drugs before they become criminals.

Adapted from Nelkin & Lindee 2004 [22].

evolutionary genomic research in the right direction and temper media attention.

Globalization has made modern bioethics contentious and complex. Attempts to establish principles and practices for global ethics have led to conventions and international ethics committees established by governments. Many times these conferences have tried to standardize bioethical practices with poor unsubstantiated arguments, and made authoritative decisions without public or scientific input. Because of pressure to reach consensus, most decisions have led to abstract and ambiguous principles that are difficult to apply to specific situations. Many decisions are ultra-conservative, banning certain types of scientific research and medical practices without

Box 10.1 The Tuskegee syphilis experiment [23].

The Tuskegee syphilis experiment was conducted by the U.S. Public Health Service (PHS) from 1932 until 1972. The experiment exposed human subjects to syphilis in order to study the effects of the disease. In the study, 600 impoverished African American male sharecroppers (201 controls and 399 experimental subjects) living in Alabama were recruited to take part in the experiment under the deception that they were receiving free medical treatment for "bad blood." The experiment proceeded for four decades despite the World Health Organization's Declaration of Helsinki in 1964, which required informed consent for any experiment using human subjects. By the time the experiment was brought to media attention in 1972, 28 men had died of syphilis, over 100 had died of complications associated with the disease, 19 children had been born with the disease, and subjects had infected 40 of their wives with the disease. In the end, the PHS denied that the experiment was akin to the studies in Nazi Germany and stated the men were volunteers, and implied that the experiment was justified by providing a greater understanding of the disease. No novel scientific information from the study was ever released. In 1997, President Clinton apologized to the eight remaining survivors.

specific justification [26] in order to accommodate the various political perspectives.

In addition to medical ethics, national practices such as mandatory DNA testing of crime suspects and military personnel, direct-to-consumer genetic testing (DCGT), and DNA dragnets have led to concerns over privacy. Questions concerning access to genetic information, how information will be used, and the privacy rights of individuals have only recently been addressed. Ethical, political, and religious conflicts over the teaching of evolution, biological warfare, death penalty, abortion rights, right-to-life decisions, immigration, and economic inequities continue to be a source of debate. All of these have had and in many cases continue to have an impact on social and human evolution.

Reproductive technologies and the new eugenics: unnatural selection?

Today reproductive technology has resulted in the ability to overcome disease and infertility and the ability to choose the characteristics of future generations through prenatal genetic diagnosis, the use of sperm/egg banks, *in vitro* fertilization, and pre-implantation genetic diagnosis. Many view these technologies as a natural process by which we can better society, but others see this as another approach to eugenics. Many times the use of these technologies results from increasing cultural pressures for parents to have children of a specific gender, children without disabilities, or children of a specific height or weight, increased intelligence, and athletic ability.

Prenatal genetic diagnosis is a common practice especially for mothers over the age of 35 and for families that have a history of genetic disease. Many of the genetic diseases that are tested for are due to single genes inherited according to Gregor Mendel's laws. Diseases that have been linked to genetic component or a single gene are defined in an online database called OMIM, which stands for Online Mendelian Inheritance in Man. Each disease has a number associated with it that explains the symptoms of the trait and its chromosomal location.

There are three forms of prenatal diagnosis that can be used to determine the genetic state of the fetus. Each has drawbacks as well as advantages [27]. (1) Amniocentesis is the most familiar form of prenatal genetic diagnosis. The procedure is done between 16 and 22 weeks into the pregnancy. A needle is used to draw amniotic fluid (surrounding the fetus) from the uterus for testing. The fluid can be tested for infection or illness or more commonly the cells in the fluid are used to test for genetic abnormalities. The fluid contains cells that have been sloughed off of the developing fetus. The cells are extracted from the fluid and grown in culture for examination. Chromosomal abnormalities or sex determination can be detected in a few days. The cells can also be used to detect genetic diseases or other traits using the polymerase chain reaction (PCR). (2) Chorionic villi sampling is the second most common technique used for prenatal genetic diagnosis. The procedure takes place within the first 10–12 weeks of pregnancy and involves scraping cells from the fetal chorion (early placental tissue). Because the cells are already dividing, they do not need to be cultured and can be examined for chromosome abnormalities immediately. The cells can also be used to detect genetic diseases or other traits using the polymerase chain reaction. (3) Maternal blood sampling is the least invasive of the prenatal diagnosis techniques but has technical challenges not associated with the other two. It involves drawing blood from the mother and using flow cytometry or magnetic cell sorting to separate the fetal cells (bearing paternally inherited surface antigens) from the maternal blood. The cells can be used to detect chromosome abnormalities, genetic diseases, or other fetal traits. Recently, fetal DNA has also been seen in maternal blood. The cell-free DNA (cfDNA) can also be used to detect genetic abnormalities in the fetus using next-generation sequencing. It may someday be possible to perform whole fetal genome sequencing or whole fetal exome sequencing. It is currently possible to provide an accurate genome sequence from the fetus in 18–19 weeks using cfDNA and DNA samples from the parents [28].

When prenatal genetic diagnosis is used and a gross chromosomal abnormality or lethal genetic disease is detected, abortion is often requested. Parents who are not opposed to abortion usually consider this to be a moral alternative, especially for lethal diseases such as Tay–Sachs disease (OMIM#272800) and Lesch–Nyhan syndrome (OMIM#300322) due to the pain and extreme suffering associated with these diseases. On the other hand, diagnosis of Down syndrome (OMIM#190685) or Huntington disease (OMIM#143100) is sometimes less clear choice because of the range of expressivity associated with Down syndrome and the late onset of

Huntington disease [20]. When faced with these decisions, there can be disagreement among the couple that can lead to marital problems. Nevertheless, the termination of a pregnancy due to a genetic abnormality, disability, or disease fits the definition of eugenics and a decision of what lives are worth living. Although this is currently an acceptable legal and social practice, many view it as a new form of eugenics.

Other examples of what some perceive as the new eugenics include techniques that do not involve abortion but have become increasing acceptable to society.

Infertility and subinfertility in the United States affects one in six couples, which have difficulty conceiving. Changes in fertility are enough to limit reproduction in most species, but in humans this has been overcome by the use of assisted reproductive technologies (ART). The reasons for infertility range from physical or developmental problems to hormonal. Damaged or missing ovaries, uterine problems, hormone-related ovulation problems, and age in females are a common source of infertility, while low sperm count, low sperm motility, no sperm, and varicocele are common causes in males [21]. Some of these abnormalities can be due to developmental problems leading to individuals who are intersex (have a mixture of both male and female reproductive systems).

Individuals or couples who cannot conceive naturally for various reasons or who have a history of genetic disease can use ART. ART can involve the use of eggs and/or sperm isolated from a variety of sources, followed by fertilization and embryo implantation. ART consists of a variety of procedures including *in vitro* fertilization (IVF), gamete intrafallopian transfer (GIFT), zygote intrafallopian transfer (ZIFT), intracytoplasmic sperm injection (ICSI), and pre-implantation genetic diagnosis (PGD). Although all of these techniques overcome the reproductive barriers of natural selection, we will only discuss basic IVF and PGD, and their association with evolution and ethical concerns.

IVF followed by PGD can be used to detect genetic abnormalities and genetic traits. IVF and PGD involve the comingling of sperm and eggs in a Petri dish to form eight-cell embryos that can be tested prior to implantation. PGD can be used to screen embryos for genetic diseases, gender, skin pigmentation, and a variety of single-gene characteristics.

ART are sometimes supported through the use of sperm and egg banks. If the female cannot produce eggs, or the father cannot produce sperm, or if parents have undergone testing and know their genetic makeup is likely to result in genetically diseased embryos, they can purchase eggs and/or sperm from a bank. Prospective parent(s) scan the catalog of donors for desirable traits. Detailed information is provided on donor education, hobbies, favorite colors, ethnicity, height, eye color, athletic ability, artistic ability, success in business, assets, and so on. In 2013, the company 23andMe was granted a patent that allows parents to view the hypothetical child given the gametes from a potential donor using its Family Traits Inheritance Calculator [28].

Once embryos are produced, a single cell can be removed from the resulting eight-cell embryos and tested using the polymerase chain reaction to detect a variety of traits or diseases. Embryos that are deemed to be free of diseases or undesirable traits can be selected for implantation. Since this involves selection for desirable traits and elimination of undesirable traits, many people considered this to be a form of eugenics.

In all of the above cases, most would agree that these procedures are not the eugenics of the past and do not involve coercion, or the elimination of other human rights, unless you feel that the rights of the fetus are being infringed or that life begins at conception. Even if you do not believe humanity begins at conception, there remain several problems and moral dilemmas associated with these technologies.

Prenatal genetic diagnosis is commonly used because of the fear of genetic diseases; however, many parents are deciding to test for nondisease traits and selecting to terminate on the basis of the genetic screen. A recent survey in the United States found that 42% of clinics reported doing PGD for nonmedical reasons, and 47% reported using PGD in all IVF cases [28]. In some countries, the tests are used to determine gender and normal embryos are discardedon the basis of gender. Some countries restrict the number of traits that can be screened, but the United States has no restrictions. In addition, not all individuals have access to these services due to costs, and because of this they are not able to make the choices that others are accustomed to.

In the case of IVF and PGD, these procedures are expensive and are not covered by insurance. Because of cost, only a select group of people can choose this alternative. The costs of these procedures sometimes go beyond the medical expenses. In the case where female and/or male donors are required, eggs and/or sperm must

be purchased. Eggs are priced according to the cost and complexity of isolating eggs and the background of the donor. The isolation of eggs from a female donor requires hormone injections (to induce superovulation) followed by surgery to remove the eggs. Eggs from a tall, college educated, musician with athletic ability can cost thousands of dollars per egg and semen from an individual with similar characteristics can cost hundreds of dollars. Gametes are chosen from a catalog based on characteristics and cost to the consumer. This raises another problem associated with expense where the wealthy can afford the gametes that potentially have the most desirable traits. The costs of IVF alone involve screening the donor, donor compensation, and medical and legal fees. The typical procedure costs between $15,000 and $20,000. From social and ethical perspectives, many see this as treating life as a commodity to be sold to the highest bidder, and an unfair advantage to those who can afford these services raising the concern that we are moving toward a social dichotomy and isolation by selection of the wealthiest.

In some cases, the biological mother or the mother who has purchased gamete(s) is not capable of carrying a child. In this case, a surrogate mother must be found who agrees to carry the child to term and give up the child at birth. In 2014, the cost of IVF and a surrogate in the United States was between $100,000 and $150,000, including agency fees, legal fees, screening fees, surrogate fees, and medical and insurance costs; $70,000 excluding medical and insurance costs [29]. Because of costs in the United States, many parents seek the same services in developing countries. In India, for example, the cost can be less than 50% of the cost in the United States, and surrogate mothers can be employed for $6000–8000 [30]. Because this is a significant amount of money for those women in developing countries, some people have described this as a form of coercion.

Beyond conventional IVF and PGD, there are other controversies associated with ART: PGD has been used to select embryos carrying genes for deafness and dwarfism so that children have the same disability as their parents, IVF has sometimes resulted in multiple births of seven and eight children that need intensive medical attention at birth, surrogate mothers have refused to give up the child they are carrying, sperm and egg donors have been identified by their biological children, sperm and egg fraud where prospective parents are not provided with the gametes they paid for, and sperm and eggs have been isolated postmortem in order to use their sperm for future IVF [31,32].

In a controversial case, a deaf lesbian couple Sharon Duchesneau and Candy McCullough used a deaf friend as a sperm donor to ensure that their child would be deaf. Duchesneau and McCullough did not consider deafness to be a disability and wanted a child that would share their cultural identity. Initially, they had trouble convincing IVF clinics to perform the procedure but eventually found one that was willing to use the sperm [33].

A 2008 survey by Baruch et al. [34] at Johns Hopkins University found that 3% of PGD clinics had used PGD to intentionally select an embryo for a disability. Most physicians refuse to choose embryos with disabilities because they feel it is not the norm and it is unfair to the child.

Another very controversial issue involves romantic couples who have been identified unknowingly as half-siblings due to the use of IVF by their parent(s). As more children are conceived through ART, there is an increased chance of half-siblings meeting and becoming romantically involved without their knowledge. Most prospective parent(s) seek ART from a regional source and many are looking for similar traits in their children. Because of this, many females will choose the same sperm or eggs for the procedure. Regional sperm banks have been known to provide over 150 females with the same sperm, and in at least one case a British doctor provided his sperm to between 300 and 600 females in his clinic. This could obviously lead to incest if sibling marriages take place without the couple's knowledge. Laws in the United Kingdom now prohibit bulk donations (more than 10) of sperm in order to prevent unintended incest. The United States has not adopted a similar law [35].

Another controversy is children conceived through IVF for the purposes of using them as donors for their sibling. These so-called savior siblings have been conceived through IVF and tested by PGD for the proper genotype match in order to provide stem cells, tissues, or organs for their sibling [36,37]. One argument against this practice is that once the savior siblings are born they do not provide consent for the biological donation used to save their brother or sister. The donations can go on for years with only the consent of the parents. Another argument against this practice is using a child as a means to an end when a couple did not originally plan on having another child.

The obligations of sperm donors to children conceived through ART have recently become an issue. Typically,

the sperm donor is provided complete anonymity after donation but in the case where the sperm donor is identified questions have arisen as to the biological father's obligations. In a recent case (2013), William Marotta [38] provided sperm to a lesbian couple free of charge; he signed an agreement with the couple giving up financial responsibility for the child. However, when the mother applied for financial assistance in Kansas, the state sued Marotta for child support. Marotta lost the case and was ordered to pay child support. The court decided that since the mother did not go though a licensed physician and inseminated herself, and the father could be identified, the financial agreement between the mother and sperm donor was invalid and the state had the right to demand child support.

Fertility treatments have increased the number of multiple births raising both social and biological issues [39]. From a social perspective, there has been an outcry of disapproval for certain families that have multiple children through IVF. These are considered nonconventional families that have gained media attention by being from a lower class or having children out of wedlock (Nadya Suleman also known as Octomom who gave birth to eight children in 2009 is a good example). Multiple births are also a biological concern since the children are born unhealthy and often have developmental problems.

Arguments for the use of ART point out that the technology has been associated with the elimination of many genetic diseases, and the fulfillment of many couples who could not conceive naturally. The fact that this is seen as eugenics is not an issue for those desiring children and they see eugenics as a problem only when it is state regulated. ART is described as reproductive liberty in addition to other natural rights. Proponents also argue that technology is a part of nature and human evolution and that this is just another step in the social and biological evolution of humankind. Proponents say that arguments against the cost of ART describe social–economic forces that will, just like any other technology, eventually make ART affordable to the masses. They argue that there is already an unfair advantage for those children who have access to private education and other costly resources but those inequities are part of the market economy and our capitalistic society.

From a biological evolutionary perspective, ART is a dramatic change from natural selection. Individuals who biologically cannot conceive are being allowed to conceive, and choices are being made to eliminate certain traits, while encouraging the development of other traits. This could lead to the predominance of certain alleles and the elimination of other alleles. As mentioned above, these choices are not available to everyone but instead the decision to bear children is based on capitalism and the free market, where gametes are for sale to those who can afford them.

As technology advances, will we shun individuals who do not or cannot take advantage of genetic technologies and have children with genetic abnormalities? Are we creating two different classes of humans through ART?

Enhancement through IVF, PGD, and CRISPR

Research on quantitative trait loci and genome-wide association studies suggest that testing for hundreds or thousands of traits and conditions will be possible in the future [25]. In addition to screening for enhancement, recent advances in gene therapy could also provide new methods of enhancement through changing the genetic makeup of human embryos using a process known as CRISPR (a method of editing the genome by removing specific nucleotide sequences discussed further in the gene therapy section). This leads to the controversial discussion of screening for genetic enhancements as opposed to therapeutic screening. There are two opposing views to the use of enhancements but all agree that enhancements are already available to individuals in the form of vaccinations, nutritional supplements, medical care, special educational opportunities, and hormonal treatments (e.g., in 2003 the FDA approved the use of human growth hormone, for social reasons, in children with idiopathic short stature). The question is whether these should be extended to genetics to do what is best for our children?

Many who are in favor of reproductive liberty and the use of ART for conception and the elimination of diseases draw the line at enhancement [40]. They argue that enhancement threatens equality, is a misuse of medical technology, and is unethical. The equality argument is similar to the one discussed above for ART and the growing discrepancies between the resources provided to the rich and the poor. The use of enhancements would provide an unfair advantage to those who could afford them. Medical technology is for the purpose of

disease prevention and treatment for ailments not for enhancement [20].

Several philosophers such as Harris [41] and Savulescu [42] promote enhancement and argue that there are already socially acceptable discrepancies in the level of medical treatment and other resources provided to the rich and the poor (e.g., enhanced medical insurance coverage for those who can afford it). They argue that the distinction between treatment and enhancement is arbitrary and insignificant. Since enhancement is a medical treatment, there is no moral justification against it. Proponents also point out that many technologies that were once thought to be immoral are now accepted as routine (e.g., heart surgery and organ transplants once considered immoral are accepted today by most Christians as moral medical practice).

The best arguments against the use of enhancements are those that focus on inequality and the long-term effects of allowing one class to use the technology. Because current medical technology is not equally available to everyone, there is no guarantee that enhancement would be available to everyone. Most current medical research is focused on the common and most profitable diseases and less research is focused on rare less profitable disease. If enhancement fell under the same guidelines, this could create different genetic classes of individuals and in the most extreme case create a different species (an example would be if individuals who were genetically enhanced decide that enhancements should include genes that would impart reproductive isolating characteristics from those unenhanced individuals). Even if enhancement became available to the general public over time, there would still be degrees of enhancement available (just as in any capitalistic society when technology does become available it is still not equally distributed). There would be individuals who could afford one level of enhancement and others who could afford an elevated form.

In general, people argue that capitalism is a great social structure that drives creative thinking and productivity but in the case of genetic technology it may instead drive genetic inequality and a different form of eugenics. As Selgelid [20] points out, “If enhancement's threat to equality was sufficiently great, then this could provide grounds for restricting reproductive liberty in the context of enhancement.” This may be the only way that restrictions on enhancement would be enforced but the threat may only be realized after the genetic dichotomy has already occurred.

Even philosophers Robert Nozick and Michael Sandel who don't agree on basic theories of justice agree that there are some things we should probably not be pursuing because of their possible consequences. Nozick [43] says that restrictions on liberty might be justified when it is necessary to prevent a disaster. Sandel suggests that genetic enhancement seems more intrusive and sinister than other forms of enhancement that we provide to our children, and he stresses the inequity that could ensue and threaten our appreciation of life as we currently know it [40].

Ethical issues associated with medical technology

Natural selection is relatively slow in comparison with cultural, technical, and environmental changes imposed by humans. Disease will continue to exert strong selection pressures but many of these pressures will be eliminated by modern medicine. Current and evolving medical technologies will continue to ensure a longer life span and the ability to reproduce. Therapeutic interventions for genetic diseases are already being used to treat or cure diseases at the prenatal, neonatal, childhood, and adult stages of life. These interventions include gene therapies (mRNA, gene insertion), stem cell therapy, biosimilars, pharmaceuticals, and surgical procedures. Many of the conditions would be fatal if not treated and so treatments or cures allow the detrimental alleles responsible for the disease phenotype to persist in the population, and may eliminate resistant alleles.

Gene therapy

Gene therapy brings together molecular biology and clinical medicine and is the most controversial of the genetic treatments. Human gene therapy is the process of introducing a “normal” gene into an individual's cells or tissues for the purpose of treating or curing a genetic disease. The gene inserted into the patient is usually the wild-type form of the gene that is used to compensate for the patient's mutant allele(s). The genes are usually delivered through a vector that usually is an attenuated

virus. Virulent genes are removed from the virus and the gene to be delivered is placed inside the viral vector [44].

There are two types of somatic cell gene therapies used to treat cells or tissues, *ex vivo* and *in vivo*. *Ex vivo* involves the removal of tissue from the patient. The gene is then inserted into a vector and the vector is placed in the tissue, thus inserting the new gene. Once the tissue has been treated, it is placed back into the patient. The *in vivo* gene therapy approach packages the gene into a vector and then the vector is given to the patient directly, with the hope that the vector delivers the gene to the proper cells, tissues, or organs.

The process is complicated by a number of factors: (1) the normal form of the gene must be identified and available; (2) the regulatory sequences needed to regulate the gene's activity must be identified and placed next to the gene; (3) the vector used to deliver the gene must be attenuated (to not cause disease) and the gene inserted into the vector; (4) the virus transfects the patient's cells delivering the gene, but because gene placement is random it cannot interrupt other normal gene activity, without detrimental effects; and (5) once the patient has been treated, he/she must be monitored for the gene's activity and any signs of detrimental effects caused by the therapy.

The first gene therapy trial was conducted in 1990 on Ashanti DeSilva who was suffering from severe combined immunodeficiency (SCID) caused by her inability to produce adenosine deaminase (ADA) (OMIM#102700). Ashanti underwent *ex vivo* therapy where her blood cells were removed, treated with the adenovirus containing the wild-type ADA gene, and placed back into her. The treatment was a success and since then there have been several reported success stories for treating SCID, adrenal failure, cancer, cardiovascular disease, HIV infection, and a form of genetic blindness [21].

Unfortunately, many other gene therapy trails have not been as successful. In 1999, Jesse Gelsinger was treated for ornithine transcarbamylase deficiency but died as a result of a massive immunological reaction to the viral vector. Around the same time, two children treated for SCID developed leukemia as a result of the ADA gene inserting near an oncogene. There have been a number of cases where the therapy did not work or resulted in detrimental effects to the patient.

Although all patients for gene therapy are volunteers who give their consent, and there are strict ethical and medical guidelines, there are still ethical concerns. Currently, only somatic cells can undergo gene therapy in the United States, but there is concern that gene therapy could involve germinal tissue or be used for enhancement. Germline gene therapy transfers genes to gametes that would ensure that the genes were present in all of the progeny somatic cells [44]. This would then allow those genes to be transferred to the next generation without consent. Research involving the replacement of defective mitochondria in gametes is already underway. This involves removing the nucleus from an egg with defective mitochondria and placing it inside a donor egg with healthy mitochondria. Daughters produced by this method would pass on the healthy mitochondria to their offspring. This technique known as three-parent embryos was recently approved in the United Kingdom. The FDA is considering approving the procedure for IVF in the United States [45].

Another more pressing concern is the use of gene therapy for enhancement. Similar to the discussion of using PGD for enhancement, gene therapy would only be available to those who could afford it. Gene therapy would be a more direct way to ensure that specific genes were inserted into an individual. Currently, scientists are looking for the genes, and eventually the specific alleles, for polygenic traits such as height, athletic ability, intelligence, behavior, and artistic ability. Once these genes and their allelic combinations are identified, they could be inserted into individuals or gametes. Psychometric, genetic, and neuroimaging studies have already begun looking at the source of intelligence with the eventual goal of using this information for enhancement [46]. A recently developed technique known as CRISPR/Cas9 is able to edit the genome by removing specific nucleotide sequences through the use of matching RNA constructs. It is not capable of adding genes but can modify or remove existing genes (https://www.youtube.com/watch?v=2pp17E4E-O8). It has been used to alter gene activity in mice, monkeys and the modification of human embryos [47]. CRISPR/Cas9 has recently raised ethical concerns over its potential use to modify human germline cells. Several scientists including Jennifer Doudna (the inventor of the genome editing protocol) have raised concerns that human genetic modifications pose a serious threat and that it is not clear that the therapeutic benefits outweigh the risks [48].

There is concern that enhancement may already have taken place among some athletes. Over the last decade, over 20 genes related to athletic ability have been identified. One example is the EPO gene that codes for erythropoietin, a hormone that increases the production of red

blood cells. Some athletes have taken the hormone to increase their oxygen uptake. There is fear that some athletes may have undergone gene therapy to insert the EPO gene along with a gene regulatory element that increases the production of red blood cells when oxygen levels drop under strenuous activity. This process is known as "gene doping." The difference between taking the hormone and gene doping is that it is much more difficult to detect gene doping [49].

Proponents of gene enhancement see it as just another advancement in technology that should be used to better society. The argument is that the legalization of enhancement would be much more effective than trying to stop it. Similar to the arguments for PGD, proponents say that the technology will only improve individual performance and that it will benefit society in the long run.

Stem cell therapy

Stem cell research and therapy hold the promise of curing a variety of genetic diseases. Stem cells are defined as cells that are undifferentiated and have the ability to differentiate into a variety of cells, tissues, and organs. There are three general types of stem cells. These include adult stem cells, induced pluripotent stem cells (iPSCs), and embryonic stem cells. Embryonic stem cells are isolated from the inner mass of a blastocyst and are pluripotent (capable of differentiating into any type of cell, tissue, or organ). Adult stem cells are found in a variety of organs and tissues (umbilical cord blood, bone marrow, etc.) and are used naturally to regenerate cells within the tissue or organ. Adult stem cells are multipotent (capable of regenerating a specific tissue type). Induced pluripotent stem cells are adult cells that have been induced to become undifferentiated by transferring four genes into adult skin cells [50]. The use of embryonic stem cells is the most controversial of the practices, because it involves the destruction of a human embryo.

Stem cells have proven to be a useful treatment for several disorders; however, because the stem cells are genetically different from the recipient, they can induce an immune response in the patient. To overcome this, two independent laboratories have produced cloned human embryonic stem cells [51]. The procedure is similar to that used to clone animals. Somatic cells were taken from adults and the nuclei were removed. The nuclei were placed inside anucleate eggs and the resulting cells induced to divide into embryos. The cloned human embryos were grown to the blastocyst stage where they were disrupted to form embryonic stem cells. These cells are genetically identical to the donors and would not cause an immune response if placed inside the donors. The research is ethically controversial because it is expensive, technically difficult, and the developing embryos are clones of the donors and if allowed to develop would become a human clone. In addition, the isolation and use of eggs, and the cloning of embryos just to harvest cells, is ethically dubious. Currently, federal funds cannot be used in the United States to support this type of research; however, there is no federal law in the United States forbidding human reproductive cloning.

Biosimilars

Biosimilars are medical products made by a living organism (plants, bacteria, animals, or human cells) either by controlling gene expression in the organism or through recombinant DNA technology. Because the drugs are made in living cells, there is no way to ensure that a biosimilar will be identical each time it is produced even though all are produced in a highly controlled environment. Differences in nutrition coupled with small environmental variation could have significant impacts on the cells and the protein products they produce.

Because many of these drugs are biologically complex and require strict purification protocols, there is concern over their consistency and the ultimate effect on health. Protein structural changes, lack of modifications, impurities, and so on could go undetected if not closely monitored.

The advantages of biosimilars are that the drugs may offer a decreased treatment rate and a large opportunity for expansion to a variety of diseases, while the disadvantages include a high cost of manufacturing and difficulty with drug preservation. There is some concern that genetically modifying plants, bacteria, and animals to produce human proteins is unethical and there is no way to predict the long-term consequences of the genetic modifications.

Genetic privacy

Scientists involved with the Human Genome Project recognized that sequencing the genome would raise

concerns over the use of genetic information for discrimination. To safeguard against this, the HGP set up the Ethical Legal and Social Implications (ELSI) program to help direct policy guidelines for the use of genetic information. ELSI focuses on four topics: (1) genomic research as it relates to the protection of medical records and their distribution; (2) how research will affect health care and ownership of genetic information; (3) the legal issues associated with genomic research; and (4) how research information affects reproductive technologies, the societal perceptions of genomic research, and the regulation of genetic testing. Even with ELSI in place, mandatory DNA testing of crime suspects and military personnel, DNA medical testing, direct-to-consumer genetic testing, and DNA dragnets have led to concerns over privacy. Questions about who will have access to genetic information, how will it be used, the privacy rights of individuals, and treatment of individuals with disabilities have only recently been addressed. The Office for Civil Rights is responsible for enforcing the anti-discrimination laws including the Americans with Disabilities Act (ADA), the Health Insurance Portability and Accountability Act (HIPAA), and Genetic Information Nondiscrimination Act (GINA).

The ADA (1990) prohibits discrimination and ensures equal opportunities for people with disabilities. The act covers discrimination in employment, government services, public accommodations, commercial facilities, and public transportation. The law was updated and clarified in 2011.

HIPAA (1996) regulations protect health insurance coverage for workers when they change or lose their jobs, and required national privacy standards for health care transactions.

GINA (2008) protects Americans from discrimination in health coverage and employment based on genetic information. The bill was meant to prevent health insurance companies from charging higher premiums or denying health coverage based on genetic information, and prevent employers from using genetic information when making job placement decisions. The bill should allow people to take full advantage of personalized medicine and employment without fear of discrimination.

The 2014 Affordable Care Act (ACA) ends preexisting condition as a means for excluding health insurance, ends insurance coverage withdrawals for honest mistakes, and provides individuals flexibility about health insurance coverage.

Although laws such as ADA, HIPPA, GINA, and ACA protect health insurance, employment, and disability, they do not apply to life insurance or long-term care and there is still concern that genetic information will still be accessed and misused by insurance companies, employers, medical professionals, and law enforcement.

Genetic testing

Recent advances in genome sequencing have made the process cheaper and faster and have provided more opportunities for the collection of genetic data from a variety of groups. The amount of data collected promises to make it easier for doctors to diagnose diseases and predict responses to medication. Adults can be tested for predisposition to genetic diseases or as carriers of recessive alleles. Adults who have a family history of genetic disease (e.g., breast cancers, OMIM#113705 and OMIM#600185) can be tested and the results can be beneficial by alerting individuals to the possible onset of disease and/or making plans for prevention or disease maintenance. Heterozygote screening for prospective parents can be very useful for detecting hidden recessives and family planning.

Genetic testing for 29 treatable genetic conditions was mandated in 2005 by the U.S. government's Maternal and Child Health Bureau upon the recommendation of the College of Medical Genetics. Since that time several untreatable diseases have also been added to the list. Testing for up to 54 disorders can be done with a simple blood test using mass spectrometry or the polymerase chain reaction. Testing for treatable diseases such as phenylketonuria (OMIM#261600), congenital hypothyroidism (OMIM#218700), and sickle-cell anemia (OMIM#603903) is common and these diseases can be treated by dietary restrictions, hormone therapy, and prophylactic antibiotics, respectively.

Many genetic tests are also administered to adults who have a family history of chromosome abnormalities, a single-gene disorder, or a multifactorial disorder. These tests can predict elevated risks or susceptibilities to disease. Ethical concerns over these tests are (1) maintaining privacy for the patient and (2) maintaining privacy for family members since many family members do not want to know their genetic information (brothers and sisters share 50% of the patient's DNA and so have an associated risk that they may not want to know about).

Individuals and/or their children diagnosed with or suspecting a genetic abnormality are referred to a genetic counselor. Genetic counselors are well-trained health care professionals who work in hospitals, pharmaceutical companies, clinics, and diagnostic laboratories. They explain the genetics, epidemiology, and treatment of the disease in question. Counselors explain basic genetics, the risks of the disease, the tests that are available, and the results of the test. If the patient is diagnosed as having a disease or predisposition to a disease, the counselor can refer him/her to a support group or direct him/her to original research literature on the subject. The diagnosis for late-onset diseases either can be comforting to the patient in helping him/her prepare for the disease or in some cases can be devastating when the patient was not prepared to hear the results. Results can also be confusing to an uneducated public and so the counselor can many times provide some directive to the family or the patient to help make decisions. This is especially helpful since many doctors do not have the psychological training to approach some of these delicate issues.

The bioethical consequences of testing are discussed below and include privacy, discrimination, and health care costs.

Over the last decade, DCGT has become popular. Direct-to-consumer profiling is being done by a variety of companies that solicit samples from customers with the promise of identifying genetic risks. These tests are available through the Internet or at local drug stores, to anyone who can afford the procedure and wants to investigate his/her genetic background. The procedure involves a simple cheek swab that is sent to the company along with the customer's basic information. The list of potential genetic information that can be gleaned from these tests grows each year, and includes single-gene traits, multifactorial traits, paternity testing, and ancestry. Clients not only receive this information but also have the option of providing their genotypic and phenotypic information to the companies' research group, or they can share their information through social media. The companies use the data to look for novel markers for disease and other characteristics.

Consumer groups, geneticists, and health officials have raised concerns about DCGT. The concerns range from genetic information sharing to the lack of follow-up by the companies. Some companies offer tests for disease traits but do not provide appropriate information or genetic counseling [28]. Most customers are not familiar with statistical analysis, penetrance, expressivity, genetic variants, genetic background, and environmental effects and are not equipped to interpret the results. There are also concerns that companies may not accurately perform the tests or overstate the impact of the results. There is growing concern over confidentiality of the results and the possible unexpected consequences of sharing genetic information. Genetic results involve relatives who may not want to know the information. Disease alleles that are revealed in one individual may affect that person's sibling or children in ways they were not prepared for. Paternity testing can reveal unexpected results that can disrupt family life. Many states do not regulate DCGT and only recently has the federal government questioned some companies' practices and the use of genetic data to create new drugs (www.engadget.com/2015/03/13/23andme-drugs-dna/).

Advocates of genetic testing point out that concerns are overblown and that current laws such as GINA, HIPPA, and ADA protect individuals from discrimination. They state that because customers give their consent to share their information they are fully informed of the consequences. Proponents argue that the information gathered by these companies has already helped to uncover rare genetic markers for a variety of diseases and that further disclosure of genetic information will only help the Personal Genome Project and show long-term gains for genetic and evolutionary research.

DNA profiling

DNA profiling began with forensics and is most commonly used in the justice system, the military, anthropological studies, and increasingly in biobanks and DCGT. In forensics, population genetics, and anthropological studies, DNA profiling takes advantage of rapidly evolving alleles that can vary from individual to individual and can be used to specifically identify individuals and family members. Biobanks and DCGT companies can act as repositories for genetic information, to be used by researchers to study population dynamics and locate genes of interest. DCGT is also popular for ancestral studies and providing medical information to the public. More than a dozen nations are using genetic information to screen their populations in "biobank" projects.

DNA profiling can be used to identify genetic ancestry and is currently being used to discriminate between

Native and non-Native Americans on reservations in the United States. Since the 1980s, many reservations have developed gambling casinos and discovered natural gas and oil on their land. This has lead to an economic boom for some Native American tribes and has led to a controversy over who can share the profits. DNA testing is now routinely used to discriminate between those who are considered Native American (by heritage) and those who are non-Native American, even though many of the so-called non-Native Americans have lived on the reservations for many generations. Many of the non-Native Americans are former slaves who sought refuge on the reservations but are now being shunned and in some cases have been asked to leave the reservation because of their non-Native status.

Molecular geneticists study different populations from around the world using DNA profiling. Some of the populations sampled are isolated and many are uneducated or not well educated. Scientists typically collect blood or saliva from individuals in order to isolate DNA. Although all of these people are volunteers and required to sign a consent form (that has been translated into their native language), it is not always clear that they understand what they are signing, the purpose of the research, what rights they have to learn about the outcomes of the research, and what rights they have to any medical products produced as a result of the research. In addition, many times these people are provided with small incentives (gifts or money) to compensate for the inconvenience and their participation. Ethical concerns over coercion, exploitation, and informed consent have come up with many of these groups. Other concerns are control over what can be done with the DNA once the study has finished and anonymity.

In addition, improper or inappropriate descriptions by researchers of the studied populations as they relate to race and ethnicity are a concern because they can lead to discrimination and a reinforcement of prejudices. A 2012 workshop in Japan brought together academics from the humanities, social sciences, and genetics to discuss these issues and the indiscriminate use of descriptors such as Mongoloid, European, and Asian by scientists in their publications. The participants concluded that scientists need to consider the views of the populations they are studying and describe them as accurately as possible. The issue surfaced because of past discrimination and unethical practices within the biomedical community and a growing concern over ethnic and racial issues [51,52].

DNA evidence can be used to convict criminals, exonerate falsely accused suspects, and identify remains. Individuals who are suspected of a felony must provide their DNA sample in 23 states and the DNA is kept on file regardless of whether the person is found innocent or guilty. In 2013, the U.S. Supreme Court ruled (in a split decision) that taking a cheek swab to collect DNA was akin to fingerprinting and was not intrusive or a violation of constitutional rights. In 2014, a San Francisco federal appeals court ruled that the California law requiring felony suspects to submit their DNA to investigators was justified.

In 2014, a federal court upheld the law allowing federal law enforcement officials to collect DNA samples from any person entering the federal criminal justice system, regardless of their innocence or guilt. The law allows the data from the sample to be kept in perpetuity.

Advances in our understanding of genetics and its relationship with behavior have entered the legal profession [53]. Molecular genetic studies have associated a number of genes and epigenetic modifications with antisocial and aggressive behavior [54–56]. Sibling and twin studies that indicate that criminal and antisocial behavior may have genetic as well as environmental components have been used as evidence for the defense [57]. This approach is a double-edged sword. Since the justice system has seriously considered genetic evidence to be sound, it could be used to vindicate criminals based on the idea that their DNA made them commit the crime. On the other hand, the genetic evidence could also be used to discriminate against individuals who have a certain genotype (based on their predisposition) even if they have not committed a crime. This would be similar to the Supreme Court decision of 1927 upholding Virginia's sterilization laws based on heritability.

There are a large number of ethical and social concerns about the collection of DNA for use in the criminal justice system. These include taking DNA before conviction and keeping it on file, the fact that DNA reveals more than just a person's identity, taking DNA without consent (from a cup, hairbrush, toothbrush, garbage, etc.), the potential for current laws to become incentives for police to randomly accuse or arrest someone to collect their DNA, and the understanding or in some cases misunderstanding of DNA evidence by the jury.

From an evolutionary standpoint, the concerns are not so much with the ethical issues of collection but the disproportionate number of minorities in the United

States who are accused of a crime, incarcerated, and have their DNA on file.

According to the Sentencing Project, racial and ethnic minorities in the United States make up about 30% of the population, and over 60% of the prison population is composed of Hispanics and African Americans. Among women, African Americans are three times more likely, and Hispanics 69% more likely, to be incarcerated than White women. Because DNA is taken from individuals accused of a crime and those incarcerated, there is a disproportionate amount of DNA collected from minorities and a disproportionate amount in the FBI system when looking for criminals based solely on DNA evidence [58]. Some people see the evolutionary implications as associating criminal behavior with specific ethnic groups. This association is a double-edged sword since the idea that DNA could influence behavior could vindicate individuals from their criminal actions or alternatively promote a feeling of intolerance for certain groups, thereby isolating those groups from society and ultimately from the gene pool.

In addition, there are environmental factors involved in behavior that are not often recognized. One of these are the evolutionary consequences to parenting. In addition to the epigenetic and other consequences of proper parenting, when a parent is removed from the home, statistics have shown that children (raised in a fatherless environment) grow up to show delinquent behavior at a higher rate than families with fathers [2,59].

Proponents of DNA profiling argue that DNA identification is a natural progression of criminal technology, and current laws are enough to protect the rights of people. They point out that the public wants crimes solved at all costs, and that DNA has helped vindicate hundreds of individuals who were falsely accused.

According to some social scientists, the new eugenics has been present in American society for some time and continues to grow. Unlike the eugenics of the 1930s, the new eugenics is more cryptic and taking shape through subtle political action, social media movements, and the advance of new technology. As Nelkin and Lindee [22] point out, determinism has led to the public's need to seize control of the future of evolution in America. Nelkin and Lindee call this movement "genetic futurism" and say that it is expressed in four ideas that are promoted in American society and involve some of the ideas we have discussed previously. (1) The idea that the poor and certain ethnic groups reproduce at a higher rate has concerned the public and politicians since the 1980s. Although a 2010 survey showed that welfare mothers have an average of 1.9 children (similar to the national average), there is still a feeling that they are flooding society with children who will grow up to be dependent and therefore welfare should be limited [60]. Although many Americans have more than two children, the focus restricting reproduction is on the poor and certain ethnic groups. (2) There are lives that are not worth living due to disabilities, disease, and behavioral problems. There is growing anxiety over the cost of caring for individuals with disabilities and social problems and a feeling that it is cruel to bring a child into the world with a disease or disability. This has increased the level of prenatal and IVF screening to avoid having children with "genetic" problems. (3) Environmental, social, and economic problems are the result of immigration and reproductive practices among certain ethnic groups. According to the Southern Poverty Law Center, the number of hate groups in America has risen from 273 in 1990 to over 1000 in 2012. The groups want to limit the population in the United States by race and class. Some groups point out the Department of Justice 2012 report that the lifetime likelihood of imprisonment is 1 in 3 for African American men, and 1 in 6 for Hispanic men, compared with 1 in every 17 for White men. The limits on immigration into the United States are directed toward Third World countries and immigration has become a national political issue. In 2014, citizens in California protested the immigration of children from Central America and Mexico with signs reading "go home don't ruin my children's dreams" and "we don't want your diseased children." In 2015, presidential candidate Donald Trump called for a ban on Muslim immigration into the U.S. and 44% of his supporters in North Carolina stated Islam should be illegal. Selective dating services based on ethnicity, religion, and economic status, among others, try to ensure selective mating. This is unlike past practices where cultural identity was preserved through assortative mating and intermarriage occurred among those who found each other physically or socially appealing. (4) The threat of current reproductive practices requires states to impose reproductive rights. Welfare reform in 1996 put a cap on the calculation of cash grants given to welfare mothers based on the number of children, sending a message about reproductive rights. In the 1990s, 13 state legislators approved offering Norplant contraceptives to women on welfare and some would offer money as incentive. In 2003 the

New Jersey Supreme court upheld a law limiting the number of children that could receive welfare. In 2014, Senator Ron Paul proposed penalizing poor women for having children out of wedlock. The 1994 book "The Bell Curve" and a statement in 2007 by James Watson suggesting that people of African descent are not as intelligent as those of European descent fueled the idea that there are inferior ethnic groups, although from a genetic standpoint there are no inferior groups.

Conclusions

Social evolution and biological evolution have shaped who we are today. Current political, medical, economic, technological, and cultural practices are shaping our evolution. Philosophers of the past and present have tried to describe just, moral, and ethical practices, but it is difficult to develop practices that can apply to everyone and the complexity of different situations. Attempts to globalize ethics have resulted in ambiguous policies with few to no directives. Because of modern medicine, almost everyone in developed countries reaches reproductive age, and current and evolving medical technologies will continue to ensure a longer life span. Infertility and subinfertility in the United States affects one in six couples and although change in fertility is enough to limit reproduction in most species humans have overcome this by the use of ART. Gene therapy, proteomics, directed evolution of proteins, evolutionary medicine, personalized medicine, and genomics are already a focus of medical research and will continue to develop as therapeutics, and may eventually lead to human enhancement. Natural selection is relatively slow in comparison with cultural, technical, and environmental changes imposed by humans. Diseases will continue to exert a strong selection pressure, but many of these pressures will be solved by modern medicine before natural selection. These practices could lead to an increase in detrimental alleles or a decrease in resistance alleles in the population. Ethical issues will most likely continue to lag behind modern technology and social and political decisions that are playing more of a role in biological evolution than ever before. An ever-increasing feeling that some cultural, religious, or ethnic groups threaten our way of life and that some lives are not worth living due to physical and mental disabilities is leading to an increase in the number of hate groups and new more cryptic forms of eugenics.

Review questions and exercises

1. Jeremy Bentham's principle of utilitarianism suggests doing the greatest good for the greatest number. Would that mean that it would be justified to use the organs of a death-row prisoner (without consent) to supply organs to save four other people? Would it matter if the four people were prominent figures in society? Would it matter if the organs came from an illegal immigrant?

2. Does altruism provide an intrinsic reward?

3. Principles of justice depend on moral or intrinsic worth, so how should we deal with the fact that people hold different ideas and conceptions of what is good and what is valued?

4. Once your DNA is in the CODIS system (used by the FBI to identify individuals), it can be used to identify you or your relatives in a partial match (your siblings share 50% of your DNA, so would show a partial match to your DNA profile). If your DNA reveals (through a partial match) that a family member may have been involved in a crime, should that data be used to question your relatives even though the partial match may also identify several other suspects?

5. Does collecting the DNA of a suspect violate the principle that you are innocent till proven guilty?

6. Because of technological advances, do you think our species has stopped evolving from natural selection?

7. What are the difficulties in determining whether behavior is genetically based?

8. How have scientists and the government dealt with the ethical issues arising from sequencing the genomes of thousands of individuals?

9. Many years ago, a sperm bank in California had collected sperm from Nobel laureates and highly accomplished males from around the world. The owner of the sperm bank required females interested in the sperm to fill out an application that included a questionnaire asking for information on education, income, artistic and athletic ability, and so on. The

owner would then go over the application and make a decision to accept or deny the client. He said he did not want to just waste the sperm on anybody. What are the social and biological concerns with this type of policy?

10 If a group of African tribesmen signed consent forms for blood donation and during subsequent studies a medical product was developed from their blood cells, do they have a right to any further compensation from the scientists who developed the product? What was their contribution to the design and development of the product? If you feel the tribesmen do deserve compensation, what do you think is fair compensation?

11 A family with a history of Alzheimer disease comes to tell you that they have heard there is a genetic basis for the disease and they want to know the probability of having children with the disease. What do you tell them about the complications of determining the probability?

12 Over the years, people have begun to use the Internet for genetic and medical diagnosis. What are the advantages and disadvantages of virtual genetic counseling?

13 Why would you test newborns for genetic diseases that have no cure or treatments?

References

1 Thareja P. 2012. Ethics and morality. Goodreads.com.
2 Sanderson SK. 2014. *Human Nature and the Evolution of Society*. Westview Press, Boulder, CO.
3 Campbell F. 2014. Early childhood investments substantially boost adult health. *Science* **343**:1478–1485.
4 Freitas D. 2013. *The End of Sex: How Hookup Culture Is Leaving a Generation of Unhappy, Sex Unfulfilled, and Confused About Intimacy*. Basic Books, New York.
5 Time Magazine. 1989. Watson.
6 Phelan JC et al. 2013. The genomic revolution and beliefs about essential racial differences: a backdoor to eugenics? *Am Soc Rev* **78**:167–191.
7 Stanford CC, Conner VJ. 2013. *Ethics for Professionals*. Jones & Bartlett, Burlington, MA.
8 Sandel M. 2009. *Justice: What's the Right Thing to Do?* Farrar, Straus and Giroux, New York.
9 Denis L. 2014. Kant and Hume on morality. In: Zalta EN (ed.), *The Stanford Encyclopedia of Philosophy*. Stanford University. http://plato.stanford.edu/archives/spr2014/entries/kant-hume-morality/.
10 Joyce R. 2006. *The Evolution of Morality*. MIT Press, Cambridge, MA.
11 Ben-Shahar Y. 2014. Genes, brains, and behavior. In: Losos JB et al. (eds), *The Princeton Guide to Evolution*. Princeton University Press, Princeton, NJ.
12 Anders S et al. 2014. Depression as an evolutionary strategy for defense against infection. *Brain Behav Immun* **31**:9–22.
13 Rachels J. 1990. *Created from Animals: The Moral Implications of Darwinism*. Oxford University Press, Oxford.
14 Wilson EO. 1975. *Sociobiology: The New Synthesis*. Harvard University Press, Cambridge, MA.
15 Ruse M. 1995. *Evolutionary Naturalism*. Routledge Press, London.
16 FitzPatrick W. 2014. Morality and evolutionary biology. In: Zalta EN (ed.), *The Stanford Encyclopedia of Philosophy*. Stanford University. http://plato.stanford.edu/archives/fall2014/entries/morality-biology.
17 Bates TC. 2008. Recently derived variants of brain-size genes *ASPM, MCPH1*, *CDK5RAP*, and *BRCA1*, not associated with general cognition, reading or language. *Intelligence* **36**:689–693.
18 Weaver ICG. 2007. Epigenetic programming by maternal behavior and pharmacological intervention. *Epigenetics* **2**:22–28.
19 Zhang TY, Meaney M. 2010. Epigenetic and environmental regulation of the genome and its function. *Annu Rev Psychol* **61**:434–466.
20 Selgelid MJ. 2014. Moderate eugenics and human enhancement. *Med Health Care Philos* **17**:3–12.
21 Cummings MR. 2014. *Human Heredity: Principles and Issues*, 10th edition. Brooks and Cole, Belmont, CA.
22 Nelkin D, Lindee SM. 2004. *The DNA Mystique: The Gene as a Cultural Icon*. W.H. Freeman & Company, New York.
23 Allen GE, Baker JJW. 2001. *Biology, Scientific Process and Social Issues*. Fitzgerald Science Press, Bethesda, MD.
24 Beauchamp TL, Walters L. 2003. *Contemporary Issues in Bioethics*, 6th edition. Wadsworth-Thomson Learning, Belmont, CA.
25 Vitti JJ et al. 2011. Human evolutionary genomics: ethical and interpretive issues. *Trend Genet* **28**:137–145.
26 Harris J. 2001. In: Harris J (ed.), *Bioethics*. Oxford University Press, New York.
27 Pierce BA. 2012. *Transmission and Population Genetics*, 4th edition. W.H. Freeman & Company, New York.
28 Murry TH. 2014. Stirring the simmering "designer baby" pot. *Science* **343**:1208–1210.
29 Circle Surrogacy. 2014. www.circlesurrogacy.com.
30 James SD. 2013. Infertile Americans go to India for gestational surrogates. ABC News, November 7.
31 Blomfield A. 2011. Family given permission to extract eggs from ovaries of dead daughter in world first. The Telegraph, August 8. www.Telegraph.co.uk.

32 Leiboff M. 2006. Post-mortem sperm harvesting, conception and the law: rationality or religiosity? *QUT Law J* **12**: 193–2003.

33 Savulescu J. 2002. Deaf lesbians "designer disability" and the future of medicine. *Br Med J* **325**:771–773.

34 Baruch S et al. 2008. Genetic testing of embryos: practices and perspectives of US *in vitro* fertilization clinics. *Fertil Steril* **89**:1053–1058.

35 Findlay S. 2012. Toronto man claims to have up to 1,000 siblings through single sperm donor. Toronto Star, April 9.

36 Spriggs M, Savulescu J. 2002. Savior siblings. *J Med Ethics* **28**:289.

37 Sheldon S, Wilkinson S. 2004. Should savior siblings be banned? *J Med Ethics* **30**:533–537.

38 Chuck E. 2013. Hey, sperm donor, don't answer that Craigslist ad! www.NBCNews.com.

39 Kulkarni AD et al. 2013. Fertility treatments and multiple births in the United States. *N Engl J Med* **269**:2218–2225.

40 Sandel MJ. 2004. The case against perfection. *Atlantic Monthly* **293**:1–11.

41 Harris J. 2011. Enhancements are a moral obligation. In: Savulescu J, Bostrom N (eds), *Human Enhancement*. Oxford University Press, New York.

42 Savulescu J. 2001. Procreative beneficence: why we should select the best children. *Bioethics* **15**:413–426.

43 Nozick R. 1974. *Anarchy, State, and Utopia*. Basic Books, New York.

44 Bourne H et al. 2012. Procreative beneficence and *in vitro* gametogenesis. *Monash Bioethics Rev* **30**:29–48.

45 Cohen G et al. 2015. Transatlantic lessons in regulation of mitochondrial replacement therapy. *Science* **348**:178–180.

46 Gray JR, Thompson PM. 2004. Neurobiology of intelligence: science and ethics. *Nat Rev Neurosci* **5**:471–482.

47 Cyranoski D, Reardon S. 2015. Embryo editing sparks epic debate. *Nature* **520**:593–595.

48 Doudna J. 2015 Genome editing revolution: My worldwind year with CRISPR. *Nature* **528**:469–471.

49 Pray LA. 2008. Sports, gene doping, and WADA. *Nat Educ* **1**:77.

50 Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**:663–676.

51 Baker M. 2014. Stem cells made by cloning adult humans. Nature News, April 28.

52 Takezawa Y et al. 2014. Human genetic research, race, ethnicity and the labeling of populations: recommendations based on an interdisciplinary workshop in Japan. *BMC Med Ethics* **15**:33.

53 Appelbaum PS. 2014. The double helix takes the stand: behavioral and neuropsychiatric genetics in court. *Neuron* **82**:946–949.

54 Strous RD et al. 1997. Analysis of a functional catechol-*O*-methyltransferase gene polymorphism in schizophrenia: evidence for association with aggressive and antisocial behavior. *Psych Res* **69**:71–77.

55 Crockett MJ et al. 2010. Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proc Natl Acad Sci USA* **107**:17433–17438.

56 Kaminski Z et al. 2008. Epigenetics of personality traits: an illustrative study of identical twins discordant for risk-taking behavior. *Twin Res Hum Genet* **11**:1–11.

57 Tuvblad C, Beaver KM. 2013. Genetic and environmental influences on antisocial behavior. *J Crim Justice* **41**:273–276.

58 Mehler B. 1994. In genes we trust: when science bows to racism. Reform Judaism, December, pp 10–14.

59 Kolliker M et al. 2014. Evolution of parental care. In: Losos JB et al. (eds), *The Princeton Guide to Evolution*. Princeton University Press, Princeton, NJ.

60 Agenor M. 2010. Women, welfare, and human rights. Harvard Public Health Review, Summer, pp 12–15.

CHAPTER 11

Future of human evolution

> We do not rest satisfied with the present. We anticipate the future as too slow in coming, as if to hasten its course; . . . so imprudent are we that we wander in times which are not ours, and do not think of the only one which belongs to us; . . . For the present is generally painful to us. We conceal it from our sight, because it troubles us; and if it be delightful to us, we regret to see it pass away. We try to sustain it by the future and think of arranging matters which are not in our power, for a time which we have no certainty of reaching.
>
> —*Blaise Pascal (Pensees 1665)* [1]

Summary

It is difficult to predict the future without reflecting on the past, and without considering the present. Historians often refer to the present and the future as a repeat of the past, or as George S. Patton stated "Prepare for the unknown by studying how others in the past have coped with the unforeseeable and the unpredictable." Although the story of the genus *Homo* began about 2.5 million years ago, the story of modern-day humans began less than half a million years ago. According to recent Y chromosome data, the most recent common ancestor from whom all human males descended occurred about 350,000 years ago [2], and the earliest fossil evidence for anatomically modern humans around 200,000 years ago [3]. Signs of a complex culture including burial rituals, fishing, and modern behavior are evident between 70,000 and 150,000 years ago [4]. Around 50,000–60,000 years ago, the migration out of Africa is clearly identifiable and *Homo sapiens* that left during this period began to interbreed with Neanderthals [5] and Denisovans [6]. Over the next 25,000–35,000 years, Neanderthals and Denisovans went extinct leaving behind modern-day humans.

Our cultural and evolutionary history has led to an ever more advanced and complex species that has developed consciousness, language, reason, morality, and advanced technology. Over the last several hundred years, we have increased our life span, developed social and political organizations, increased our scientific knowledge, developed technology to overcome reproductive barriers, increased our rate of mobility, occupied a wide range of environments, and brought about changes in our climate, diet, and susceptibility to disease.

During the early evolution of humans, there were a number of environmental, genetic, and cultural forces that led to the current physically and geographically diverse groups that we call modern-day humans. Although the forces that shaped our evolutionary history (genetic variation, natural selection, and adaptation) still exist, in many cases they have been modified, begging the question where do we go from here? This question partially arises because of the apparent human isolation from natural selection and the feeling from some philosophers and scientists that humanity has reached a state of evolutionary stasis, where humans are buffered from both positive and negative selection and are able to adapt phenotypically but without significant genetic change [7]. Evolutionary stasis promotes the idea that humans modify their environment to match their genes rather than natural selection shaping our genotype. Proponents of evolutionary stasis point out that because of increased admixture, increased mobility, the increased lack of isolated populations, and reduced drift, significant beneficial mutations will fail to become established in the gene pool. In addition, they point out that humans are becoming more homogeneous, and because of technology more humans

Genomes, Evolution, and Culture: Past, Present, and Future of Humankind, First Edition.
Rene J. Herrera, Ralph Garcia-Bertrand, and Francisco M. Salzano.
© 2016 John Wiley & Sons, Ltd. Published 2016 by John Wiley & Sons, Ltd.

who would have not lived in the past due to natural selection are now surviving beyond reproductive age.

The proponents of stasis fail to recognize three important facts: (1) natural selection is not the only force driving evolution, the random forces of mutation, recombination, and drift still occur along with adaptive evolution influenced by environment; (2) the majority of people do not live in affluence and are subject to selection through the lack of technology, malnutrition, and disease; and (3) if selection is relaxed, the mutation rate to neutral alleles will increase, and evolution proceeds [8].

This chapter will focus on some of the cultural and microevolutionary (drift, selection, and mutation) pressures that are still effective and driving our evolution, and discuss how these pressures might affect our future as a species. Some of the forces affecting the future of human evolution include gene and cultural coevolution, longevity and population size, mutation rates and the formation of new genes, diet, climate change, and artificial selection. All of these factors will have some effect on the future of evolution and will be discussed below. Although it is impossible to predict how random variation and our environment (cultural and ecological) might change allele frequencies in the future, we can still examine the past, and with the tools from the present speculate on how this might affect the future of human evolution.

Gene and culture coevolution

Culture is typically defined as a set of beliefs, ideas, values, and knowledge that can be passed from generation to generation (Chapter 7). Each year humans seem to increase their control over health, reproduction, technology, and their environment. Today, we appear to be evolving principally through cultural and technological means, and value a variety of activities and traits that our ancient ancestors would have thought frivolous (social networking, game playing, fashion, recreation, slenderness, economic status, etc.).

We define much of our environment by our culture, and cultural change has driven adaptive evolution and played a significant role in shaping who we are by helping to develop new genotypes in response to cultural changes. Culture can affect both selection and the response to selection. Humans demonstrate some level of phenotypic plasticity (the ability for a genotype to produce various phenotypes under different environmental conditions) allowing for the accumulation of variation that can arise under different environmental conditions and increase the rate of adaptability in an atypical environment. Some past examples of cultural and adaptive evolution affecting our genomic makeup include lactose tolerance [9], the ability to digest high amounts of starch [10], sex selection (mating preferences), resistance to malaria, adaptation to high altitudes [11], and the modification of our genomes through molecular technology (Chapter 7). Gene–cultural coevolution has certainly been shielded from natural selection by some technological advances, and this condition will certainly continue, possibly leading to even more dramatic effects on shaping our future evolution.

Today, cultures change so rapidly that it is hard to measure the effect on selection. Still some cultural adaptations are obvious and will continue into the future. Large populations survive in limited environments that would have challenged even the fittest of our prehistoric ancestors; we have overcome many diseases that plagued our ancestors, developed technology that overcomes genetic and reproductive barriers, increased our population size to levels thought unsustainable 100 years ago, and even begun to affect our climate. These changes, however, come with costs and as humanity moves forward there will be more and more difficult choices to make.

Life expectancy and population growth: past, present, and future

Human life expectancy is the mean number of years that individuals live within a population from birth until death. Humans have evolved the greatest life expectancy among primates. Estimates of human life expectancy during the Paleolithic and Neolithic periods range from 20 to 39 years [12]. Life expectancy since the early 1800s has doubled. According to the 2015 United Nations' World Population Prospects report, between 2010 and 2013, average worldwide life expectancy at birth was 68.5 years for males and 73.5 years for females. The average was calculated from 223 countries where life expectancies range from 49.4 years in Chad to 89.6 years in Monaco [13]. Women tend to live longer than men and comprise a larger share of the older population. In 2015, women accounted for 54 per cent of the global population aged 60 years or over and 61 per cent of those aged 80 years or over. Improvements in survival at

advanced ages in the older population that is aged 80 years or over is projected to grow from 14% in 2015 to 21% in 2050. According to the United Nations' report the main cause of overall population aging is a decline in fertility (young couples having fewer children than in previous decades), followed by improved longevity. Part of the increase in life expectancy is the result of environmental changes that have resulted in lower mortality in developed countries due to improvements in medicine, nutrition, and hygiene. Although genetic factors may have played a minor role (20–30%) in life expectancy [14,15], they along with environment [16] have contributed to some of the longest life spans ever recorded [17].

The increase in life expectancy has led to a significant increase in population. It was estimated that the anatomically human population 300,000 years ago was around 1 million, and by 25,000 years ago, the modern-day human population had risen to over 3 million. The world population has risen dramatically since the Upper Paleolithic [18] as a result of cultural innovation. The increase occurred in surges along with the development of stone tools, agriculture, the domestication of animals, and the industrial age. The growth resulted from cultural adaptations altering our environment each time the population increased. These adaptations continue even today as we try to develop technologies (genetically modified foods, agricultural innovations, water purification systems, etc.) to sustain our populations. Although an increase in population is a measure of evolutionary success, it is still too early to declare our species as a success. Many species have evolved and become numerous only to become extinct, and human history is very short in comparison with geological time.

The increase in the human population has led to greater and greater use of resources questioning the limits of population growth. Limiting factors to growth include potable water, oil and minerals, soil conditions, and the amount of energy that can be delivered by wind, water, and sunlight. Even with increases in technology, we cannot sustain the current levels of growth because of limits on the amount of land that can be used for agriculture and the sources of fresh water, which limit the Earth's carrying capacity. These limits will put pressure on population growth and hopefully lead to its stabilization, the ultimate goal, without running the risk of extinction [19].

Current and future increases in population size are linked to the rate of mortality before reproductive age, the size of families, and age at first reproduction. To the extent that these are genetic traits (e.g., risk-taking behavior and disease susceptibility for mortality), or cultural norms for average age of marriage and for having children, these traits will be selected for or against. For example, in populations where parents have three to four children at a young age (18–28 years) the population will grow more rapidly than a population where parents are older (30–40 years) and have two to three children, given the same mortality rate and assuming the children grow to reproductive age.

The rate of population growth will also depend on the interbirth interval, the average age of menopause, and the average age of the mother at last birth. Starting a family at a younger or older age (possibly because of cultural practices) would influence the rate of population growth. Population size has various genetic consequences on drift, mutation rate, and gene flow, associated with larger and/or smaller populations (discussed below). Those populations that have more children that live to reproductive age will have the largest genetic impact on future generations and evolution.

Some scientists have speculated that life expectancy and population size in the future will stabilize or even decrease due to lifestyle factors (cultural factors leading to a desire for fewer children, pressure on water and food production, increase of infertility or subinfertility, etc.). The increase in life expectancy brings with it age-related diseases that can affect the quality of life. Diseases such as coronary artery disease, diabetes, Alzheimer disease, Huntington disease, Parkinson disease, and certain types of cancer have risen as a result of an aging population. Many individuals are choosing to resist treatment by having living wills that have "do not resuscitate" orders and there is an increase in couples screening their unborn offspring for age-related diseases in an attempt to eliminate them. In addition, the increase in obesity, sustained alcohol and drug use, psychological disorders related to stress, risk behavior, the increased acquisition of weapons, and their complications could reduce life expectancy especially in developed countries [20,21].

Mutation rates and future evolution

Because the human population is not at genetic equilibrium, along with the increase in the population come genetic consequences. Understanding the different

classes of mutations and mutation rates is essential for understanding the genetics of evolution and its future implications. Population genetics predicts that with a large population drift is reduced and gene flow increased, accompanied by increases in the number of random mutations. Small populations have relatively very few new random mutations. For example, if a mutation has a probability of one in a billion per gene per generation (10^{-9}/gene/generation), then with 7 billion people in the world the likelihood of a mutation at this locus is 14 occurrences per generation. The steady increase in population size predicts that soon every single-step mutation will occur at least once every generation. This increase in mutation will create a pool of new variants that will increase heterozygosity and human's ability to adapt to new environments [22]. Most mutations will be neutral or deleterious, though a few may be beneficial. As the population increases, so will the number of deleterious mutations and those mutations will be more likely to increase in number as the population continues to grow. Assuming the deleterious alleles are under purifying selection, the effectiveness of natural selection should also increase as the effective population size increases removing deleterious (disadvantageous) mutations, thus maintaining and driving novel advantageous mutations to higher frequencies. Even if an advantageous allele is lost, the high population size should ensure that it will be reintroduced at some point. This could lead to a substantial increase in the frequency of an advantageous allele. For example, if a new mutation conferred a selective advantage that increased the number of offspring by 5%, in 423 generations (approximately 10,000 years) 99.5% of the population would carry the allele [23].

The increase in the number of mutations due to increased population size is also predicted to increase the number of deleterious mutations. Even today, many of these deleterious mutations are under purifying selection and this is expected to continue into the future. For example, prenatal and gametogenic mechanisms are capable of removing deleterious mutations. One purifying prenatal mechanism is the rate of spontaneous abortions (which constitute 10–30% of pregnancies). In addition, gametic mechanisms that occur during pre- and post-gametogenesis reduce the number of viable gametes. In females, oocytes are reduced from about half a million at birth to about 400 pre-ovulation. In spermatogenesis, 86% of fertile mobile sperm are not capable of binding to the zona pellucida for penetration of the oocyte. In both these processes, the gametic mutation load can be reduced purging deleterious mutations [23].

Gazave et al. [24] simulated how purifying selection operates during population growth. They found that while population growth increases the number of deleterious mutations, it only slightly increases the number of deleterious mutations carried by each individual. Their computer simulations also showed that a higher proportion of deleterious alleles were eliminated during each generation and that natural selection was most effective at eliminating the most deleterious mutations. They concluded that in a growing population the risk of complex disease might be distributed across a larger number of weakly deleterious and rare variants [24].

Although you might conclude that with a current human population of 7 billion, random mutations would occur at every site in the genome, evidence from whole-genome sequencing suggests that single-nucleotide variations (SNVs) and copy number variations (CNVs) are not random. Transitions outnumber transversions in SNVs, the distribution of *de novo* SNVs has been shown to be nonrandom, and CNVs have been shown to be associated with nonallelic homologous recombination sites. Mutation rates are not random at nucleosome occupancy sites, and at CpG sites where mutations can be 10–18-fold higher than non-CpG sites. Mutation rates appear to be associated with replication timing, transcription, and repeat content [25]. Some recent evidence suggests that 76% of new mutations originate in the paternal lineage and that the number of mutations increases with paternal age. The increased rate is measured by the increased number of progeny carrying mutations from the father and these increased mutations result in an increased effect on the Y chromosome. Studies suggest that the increase in mutant sperm may be driven by selfish genes that confer growth advantages to the mutant sperm [26]. Other studies suggest that there is considerable variation in the mutation rate among different families [27].

There is evidence that mutation rates differ between somatic and germinal tissues. While germinal mutations are passed on to future generations, somatic mutations have the potential to contribute to a variety of genetic diseases and increase with age. Somatic mutations can affect population size through increased mortality. Recent studies suggest that mutation rates in humans

have not remained constant over evolutionary time and most likely changed numerous times during our evolutionary history possibly due to variability in generational time [28–30]. Recent genome sequencing of a 45,000-year-old human suggested that one to two new mutations arose each year over the past 45,000 years, providing an estimate of mutation rate over a relatively long time period [31]. The affect of mutation and mutation rates will continue to be a force in future evolution in both the modification of existing genes and the development of new genes.

The evolution of new genes

The modification of existing genes will no doubt have an effect on evolution (as discussed above); however, the evolution of new genes and their elimination or fixation will also play a role in the future of human evolution. In addition to the formation of new genes, old genes may be lost from the genome. Although it may be assumed that young genes play an insignificant role because they have not reached optimization through evolutionary mechanisms, studies have uncovered young genes with important and in some cases essential molecular and cellular functions [32]. New genes have been found to function in biochemical pathways, gene networks, and development. It is important to understand the mechanisms of new gene origins in order to make predictions on how they may affect the future of evolution. New genes arise by several different mechanisms including gene duplication, chromosome duplications, exon/domain shuffling, RNA-based duplication (retroposition), transposon insertion, lateral gene transfer, frameshift mutations, and gene fission and fusion mechanisms (discussed in Chapters 4 and 5). All of these mechanisms require new genes to be developed from preexisting genes [33], although *de novo* origins may also exist.

The sequencing of thousands of genomes in different species has provided data to estimate the rate of new gene origination. Studies suggest that the origin of new genes is relatively common, and that new genes have been modified to perform new functions [34]. Although gene duplications can sometimes cause deleterious effects, estimates of the rate of new gene origination through gene duplication mechanisms have established a rate of 0.01 per gene per million years, or 100 new duplicates per million years per 10,000 genes. A gene created by a structural change (fission, fusion, transposon insertion, etc.) will most likely take on a function that differs from the original source gene. This could be as simple as a different temporal or spatial expression or the gene forming a completely different protein product. The majority of retrogenes (genes copied from a RNA by reverse transcription) in humans are chimeric genes formed by exon regions from surrounding sites. Estimates for the rate of retroposition are 1 per million years per genome, and the rate of formation of chimeric genes through retroposition is 0.01 per million years per genome [34].

De novo genes (those not arising from previous genes and found in a single species) have also been found. These genes have developed from previously noncoding DNA or noncoding RNA. These genes develop when stretches of noncoding nucleotides take on mutations that allow them to form open reading frames and promoter sequences. A number of complex *de novo* genes have been found in humans [35,36]. Studies in *Drosophila* suggest that *de novo* gene origination occurred approximately 23 times per genome per million years suggesting that new genes evolve about every 50,000 years in the species [34].

Climate change

Climate is defined as the regional measure of variation in temperature, humidity, atmospheric pressure, wind, and precipitation. Climate includes the biosphere, small local environments, and the overall surroundings, which vary according to geographic location. As described previously (Chapter 6), climate changes in the past were most likely responsible for multiple human migrations and for driving adaptive evolution. Scientists have speculated that evolutionary mechanisms needed for humans to adapt to the broad range of climatic conditions may have triggered increases in brain size and cognition, and changes in social structure and locomotion.

Over the last century, we have visibly transformed 40–50% of the Earth's land surface for production and settlement [37]. In addition, we have increased our use of fossil fuels, and increased our need for fresh water and food, leading to disturbances in our climate. Deforestation, increased energy consumption (through agriculture, transportation, and development), pollution, and increased population size have contributed to

increased CO_2 levels, rising sea surface temperatures, and modifications of our atmosphere. Unfortunately, these changes are global and there are no places left on Earth that have not been influenced by humans [37].

Over the last several decades, scientists have concluded that climate change is occurring at a rate unseen in the past and have predicted that it will continue into the future. Those who do not believe in climate change point to the fact that climate change has occurred since the Earth began; however, the rate of change is what concerns scientists today. Over the next 100 years, an increase in global temperature of 2–4 °C centigrade is predicted, with a resulting rise in sea levels of up to 50 cm. We have seen how past environmental changes affected human adaptation, and predicted climate changes will impact different regions of the Earth and different cultures in a variety of ways. Sea level changes from melting glaciers, change in weather patterns, and continued use of fossil fuels will further alter the current landscape and atmosphere. The dramatic changes could challenge the survival of many species including our own, as large populations are forced to migrate from their existing settlements. Atmospheric change could further increase UV levels and toxic pollutants (acid rain, etc.). The greatest concerns could be the effects on landscape changes and mobility. Landscape changes are predicted to alter agricultural practices and the availability of fresh water. Landscape changes forcing the mass migration of people due to weather extremes could also increase the spread of exotic diseases [38]. Overall, climatic changes could affect human adaptation through nutritional changes and the ability of populations to adapt to new diseases. Those with stronger immune systems may be selected for under these conditions.

It is difficult to speculate on exactly how future changes will affect human adaptive evolution because so much is dependent on human behavior (how humans will alter or adapt to the new environment). Many feel that technology will help to avert any dramatic consequences; however, there is still concern about how the poor and developing countries will cope with dramatic climate change and the adverse effects, since not all countries will be able to use the same response or have the resources to adapt. Despite some predictions of mass extinction [38], there are alternatives given the proper choices. Dramatic changes could be slowed or reversed if governments agree to a reversal of habitat destruction, a stabilization of population size, global agreements on the use of renewable energy sources, reduction in fossil fuels, and continuous impact assessments. Scientists will have a major role in convincing government officials and the populace to alter many cultural and environmental practices, in order to ameliorate the current situation.

Diet

The response of agriculture to climate change may increase or decrease food production depending on the changes in weather patterns. Water for crop production and animal husbandry may be a major concern for regions that suffer from extreme weather conditions. This would be especially true for regions that have no infrastructure for irrigation or to control flooding. The types of crops grown and domestic animals may also change depending on weather conditions.

Our present metabolic mechanisms evolved early in our history and were conducive to the Paleolithic diet of 40,000 years ago. The relatively recent advance in agriculture, animal husbandry, and the production of processed foods over the last 10,000 years has had a dramatic influence on the human diet. An increased level of trans fats, refined sugar, sodium, and low fiber has led to a variety of changes in our metabolism, physiology, and disease [39] (see Chapter 8).

Agriculture, lifestyle, and cultural food preferences have already been shown to influence human adaptive and nonadaptive evolution. Examples of diet and adaptive evolution were provided previously (lactose and starch digestion). A more recent example of adaptive evolution and its relationship to diet can be seen in the increased occurrence of type 2 diabetes in the United States due to changes in diet and lifestyle. One theory is that the genes that currently predispose people to type 2 diabetes were initially favorable and adapted under conditions of famine (thrifty gene hypothesis). Dietary changes over the last century have led to the increase in diabetes among individuals with the genetic background for tolerance to famine. Nonadaptive evolution related to diet has been proposed to have led to changes in tooth formation, jaw development, and brain development [22].

When the dietary environment remains constant, stabilizing selection maintains the genes for genetic traits that provide optimal performance. Dietary

changes affect our environment, and future changes in diet will affect directional selection that will alter the genome [40]. Future changes in diet associated with climate change or cultural preferences may lead to further evolutionary discord and the possible onset of nutritional deficiencies and diseases, and the need for dietary adaptation. Those populations that cannot adapt may be subjected to increased mortality. The ability to adapt may be dependent on geographical location, poverty, and technological advances, especially in developing countries.

Sex selection

Mate choice has been shown to be a strong evolutionary force in many species, including humans, and may be responsible for many genetic and phenotypic changes of the past. Selection can occur at many different levels including attractiveness, behavior, income level, education, religious and political beliefs, personality, and skin color, and there is some evidence that body odor and olfactory senses may play a role [41]. Mate preferences vary among cultures (facial attractiveness, body shape, and social status) [42] and the reasons for mating (casual or long term, arranged marriage, etc.) [43]. Men and women have both evolved mate selection preferences, and both are engaged in competition for mates (intrasexual competition) [42]. This competition has been shown to go beyond phenotypic selection. There is recent evidence that there is competition among sperm and their ability to fertilize the egg (possibly due to meiotic drive), which may explain why spermatogenesis, fertilization, and olfactory and chemosensory genes show positive selection [23].

A study looking for selection mechanisms in humans revealed that selection is acting to reduce reproductive ages in both sexes (earlier puberty). In addition, the age of menopause and age at last birth has also increased leading to a broader range of reproductive time [44]. Studies looking at reproduction patterns in 2238 women determined that the women who had the most children were slightly shorter and portlier than average, and that selection has continued to act for reduced height in females in three postindustrial populations. The researchers predicted that if the trend continues in 400 years women would be 0.8 inches shorter and 2.2 pounds heavier [43]. In male populations, selection appears to be directional for increased height and stabilizing for intermediate height. Laland et al. [45] found that selection for male assets was weaker in industrialized countries and stronger in nonmonetary economies with polygamy.

The same traits in the different genders can encounter different selection pressures. One example is seen in the stronger fitness advantage for short women and tall men [46]. Because height is heritable, this would predict that future offspring would be short or intermediate in height, which may be beneficial for female offspring but not males. This continuous struggle will most likely last into future societies, especially those with short women and where early age of reproduction is favored [47].

As with other traits, humanity's future will depend significantly on behavior and the response to environmental and social changes. In mate choice and reproduction, the role of celibacy, intelligence, economic success, social success, and stature may play greater roles in the future. These traits could work to give two different outcomes. If people who are less intelligent and less successful economically and socially have more children, this may lead to these traits evolving downward. If instead intelligence and economic and social success become more desirable, and couples with these traits have more children, then there will be an upward trend in these traits [48]. The roles of intelligence and economic status may be especially troubling in the future due to artificial selection (genetic manipulation and artificial intelligence).

Artificial selection

When measuring lifetime reproductive success in developing countries, infant mortality is high and therefore mortality has a greater influence than fertility on selection of genetic variation. In developed countries, variation in fertility influences genetic variation rather than mortality, because most children reach reproductive age [46]. Today, however, in developed countries assisted reproductive technology (ART) has resulted in the ability to overcome most infertility problems. In addition, ART has led to the ability to choose the characteristics of future generations through prenatal genetic diagnosis, the use of sperm/egg banks, *in vitro* fertilization, and pre-implantation genetic diagnosis (see Chapter 10). One result of overcoming infertility through

the use of sperm and egg banks is the increased concern over consanguineous marriage resulting in increased homozygosity. Currently, about 10.8% of human couples are related globally as second cousins or closer [22]. The use of sperm and egg banks has increased the number of individuals who do not know their biological father and/or mother. Since sperm and egg banks are regional, there are a number of individuals in the same region who unknowingly are biological siblings. This has led to several cases of accidental incest among half-siblings [49]. This number could increase in the future as more infertile or subfertile couples, homosexual, or single parents use sperm and egg banks as a source of gametes for reproduction.

Many view ART and selective abortion as a natural process by which we can better society, but others see this as a cryptic approach to eugenics. In many cases, the use of these technologies is driven by cultural pressures for parents to have children of a specific gender, children without disabilities, or children of a specific height or weight, increased intelligence, and athletic ability.

Medical technology has already taken us down the path of artificial selection, and there is growing concern over science and technological advances. This has created a dichotomy of those who accept science and technology as a part of cultural and biological evolution versus those who consider it "playing God" or forbidden knowledge. One of the greatest concerns about future technology is affordability, and which groups will have access to ART. In addition, there is concern over the manipulation of germ cells using CRISPR technology, which could lead to altering the entire genome of individuals and their offspring (see chapter 10).

The future could bring about groups of individuals who can choose desirable traits or remove deleterious traits through technology, and those who cannot (see Chapter 10). As stated above, the increased population size will result in more mutations (most deleterious or neutral). The increase in mutations could lead to groups that are able to avoid deleterious alleles (through selective abortion or advanced technology) and those that cannot. This may result in the creation of two different societies, and some have speculated that over the long term it could lead to two different species. If future technology can provide certain people with more intelligent, successful, longer-lived children, then the consequences could be dramatic for those who cannot afford genetic alterations [48]. A recent advance in the development of a synthetic chromosome provides a possible scenario for future genetic modification. Annaluru et al. [50] recently designed a eukaryotic chromosome. The chromosome was developed for yeast, using advanced DNA synthesis techniques, and is composed of over 270,000 nucleotides. The long-term idea is that this type of synthetic chromosome technology could be used to create chromosomes that could eventually be placed into humans and provide characteristics such as disease resistance or other traits. This technology could also be used to provide traits to individuals and their offspring based on their ability to pay for the services. There could come a time when those who are genetically modified may choose to have reproductive barriers put in place to prevent admixing with those who are not genetically modified, which could result in significant political and cultural effects.

Some scientists have speculated that the alteration of genes could be our downfall since millions of years of evolution has brought us to this point, and the manipulation of genes could have dramatic unintended effects due to pleiotropy and a general misunderstanding of the complexity of epistasis. In addition, environmental and cultural changes can occur relatively quickly, making predictions on favorable traits more prone to error than good [22]. The technology of gene manipulation appears to be a part of our future and only time will tell whether the experimental manipulation of humans will be a benefit or detriment.

Transhumanism and artificial intelligence

Transhumanism is a cultural movement whose eventual goal is to transform humans by enhancing them through emerging technology. This movement is closely related to artificial intelligence or the symbiosis of humans and machines. We already use microchip implants in humans and animals for identification and medical purposes, and already have equipment that places our senses in a state of virtual reality. Paralyzed patients are able to use their brainwaves to control simple robotic tasks [51], and recently scientists have simulated the 302 sensory and motor neurons of *Caenorhabditis elegans* in a software program (built a functional computer model of the brain) and placed it into a small robot. The robot sensed and moved around objects demonstrating very simple

behavior [52,53]. The Human Brain Project started in 2009, although somewhat controversial and still in its infancy, promises to make a computer simulation of the human brain [54]. At this point, mapping and simulating the 85 billion neurons in the human brain is science fiction; however, 100 or 1000 years from now, it may not be. This type of technology raises the question of what it is to be human and whether a computer simulation of your brain is capable of thinking, learning, and reasoning.

As we continue to advance our computer technology and our dependence on computers, some scientist have predicted that the future will be a human–machine symbiosis uploading our minds into computers or linking our bodies to machines [48]. Although this sounds like a technological adaptation that would advance the species, there are many critics. Critics claim that increased dependence on technology especially machines would render humans obsolete. Nick Bostrom [55,56] predicts that a fusion with machines could ultimately lead to a situation where it would be difficult to define what is human? Fusion with machines he says would create unparalleled consequences and a life of drudgery where love, humor, sex, food, and drink would be unnecessary and become obsolete. Some scientists see Bostrom's visions as implausible and dramatizing. Several have suggested that a human – machine symbiosis would be more of a cooperative adventure and that machines would do continue to do the routine work while humans would remain as creative architects [57]. Some have even suggested that a symbiosis will create new economic opportunities [58].

Our dependence on computers has already led to some effects on health (carpal tunnel syndrome, vision problems, and a lack of outdoor activity). The artificial intelligence and trans- humanism scenario is played out in many science fiction films and at this point in time seems impossible, but many prior science fiction scenarios have been realized in the last century and only time will tell whether humans and machines will coalesce.

Conclusions

Modern humans have been molded by genetic variation, environment, and natural and artificial selection. This has led to successful reproduction, a significant increase in life expectancy, and large populations. Large populations and greater mobility weaken genetic drift, increase gene flow, increase genetic variation, and decrease the genetic differences between human populations. If this continues, it will lead to more heterozygosity among populations and a wider variety of gene flow and genetic variation. The continued mobility and gene flow will eventually result in a decrease in local adaptations, populations where genetic differences are reduced or eliminated, and only individual variation will be biologically meaningful due to the high levels of genetic variation in the common gene pool. Mutation and mutation rates will continue to be a force of evolution in both the modification of existing genes and the development of new genes. The increase in population will add to the overall number of mutations and while somatic mutations may increase mortality germinal mutations will be passed on to future generations for selection. In the future, the high levels of genetic variation may be important in helping to adapt to new environmental and cultural changes. Culture and gene coevolution will continue to have effects on human evolution and much of this may be dictated by future behavior. Climate change, diet, sexual selection, and technology will have an impact on future evolution and the level of impact will be dictated by the human response to these forces. Climate change could affect human adaptation through nutritional changes and adaptation to new diseases. Climate change could also affect our diet and our ability to adapt through directional selection to new foods and nutrition. Sex selection may also change in the future leading to new forms of intersexual and intrasexual selection. The increased heterozygosity due to increased mutation rate may be important in helping humans adapt to the new environments caused by climate change and culture.

Artificial selection has been around for many decades and will only increase as technology allows new methods to overcome reproductive barriers and genetic diseases, and the discovery and modification of genes for multifactorial traits. It is impossible to predict how new technology will be involved in the future, but the past tells us that technology will continue to play a significant role in human culture. Even today we are becoming more and more dependent on technology and it is not clear when our dependence will exceed our abilities to cope with environmental, genetic, and cultural changes, leading to a greater dependence on technology. Some scientists predict that the increased dependence on technology may lead to further human speciation. Speculations on a new human species encompass a wide range of

possibilities including genetic engineering resulting in modifications to a select group, habitation of other planets, isolation due to a catastrophic event that no longer allows the global exchange of genes, or a machine–human symbiosis. The predictions of mass extinction could also be possible. Mass extinction could result from overpopulation, the inability to adapt to climate changes, the rise in deleterious mutations already present in the population leading to high mortality, or a decline in human dispersal and isolation due to a catastrophe. Isolated populations may be subject to an inability to adapt and/or the disadvantages of homozygosity. Alternatively, the lack of dispersal and the isolation of populations (on Earth or in space) would increase genetic drift, which may present more adaptive innovations to overcome environmental insults. Either way, there is still hope for a long future to human evolution.

Review questions and exercises

1. It has been proposed that some day humans may inhabit other planets. What would be some of the potential problems associated with the transition from Earth to other planets?

2. Recently, scientists have developed algorithms that allow computers to play a simple 1970's computer game, developed by the company Atari. After many repetitions of the game using the algorithm, the computer has developed the most optimal approach to winning the game (learning). What does this say about the future of transhumanism, and what dangers might be involved with this type of technology?

3. We already live in a society of inequality where some people have more resources than others (material, economic, intellectual, etc.). If genetic manipulation leads to two or more classes of genetically modified individuals, what might be the problems associated with this type of society?

4. Humans are the only species on Earth who try to avoid natural selection through technology. Do you think that technology will continue to help us avoid some of the forces of natural selection into the future?

References

1. Hutchins RM (ed.). 1952. *The Provincial Letters, Pensees, and Scientific Treatises, by Blaise Pascal*. The Great Books of the Western World, Vol. **33**. William Benton Publisher, Encyclopedia Britannica, Inc., Chicago, IL.
2. Mendez F et al. 2013. An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree. *Am J Hum Genet* **92**:454–459.
3. Hopkin M. 2005. Ethiopia is top choice for cradle of *Homo sapiens*. *Nat News*, doi: 10. 1038/news050214-10.
4. Campbell BG et al. 2006. *Humankind Emerging*, 9th edition Pearson, Boston, MA.
5. Green RE et al. 2010. A draft sequence of the Neanderthal genome. *Science* **328**:710–722.
6. Reich D et al. 2011. Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am J Hum Genet* **89**:516–528.
7. Powell R. 2012. The future of human evolution. *Br J Philos Sci* **63**:145–175.
8. Brandon RN. 2006. The principle of drift: biology's first law. *J Philos* **103**:319–335.
9. Burger J et al. 2007. Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proc Natl Acad Sci USA* **104**:3736–3741.
10. Perry GH et al. 2007. Diet and the evolution of human amylase gene copy number variation. *Nat Genet* **39**:1256–1260.
11. Huerta-Sanchez E et al. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**:194–197.
12. Kaplan H et al. 2000. A theory of human life history: diet, intelligence, and longevity. *Evol Anthropol* **9**:156–185.
13. Central Intelligence Agency. 2012. Rank order, life expectancies at birth, 2014 estimate. The World Factbook. https://www.cia.gov/library/publications/the-world factbook/rankorder/2102rank.html.
14. Hjelmborg J et al. 2006. Genetic influence on human lifespan and longevity. *Hum Genet* **119**:312–321.
15. Budovsky A et al. 2013. LongevityMap: a database of human genetic variants associated with longevity. *Trends Genet* **29**:559–560.
16. Moore SC et al. 2012. Leisure time physical activity of moderate to vigorous intensity and mortality: a large pooled cohort analysis. *PLoS Med* **9**:e1001335.
17. Coles S. 2014. Current validated living supercentenarians. http://www.grg.org/Adams/E.HTM.
18. Deevey ES. 1960. The human population. *Sci Am* **203**:194–204.
19. Cohen JE. 1995. *How Many People Can the Earth Support?* W. W. Norton & Company, New York.
20. Olshansky SJ et al. 2005. A potential decline in life expectancy in the United Sates in the 21st century. *N Engl J Med* **352**:1138–1145.

21 Bains W. 2009. Evolution, hypotheses, and the question of whether we are still evolving. *Biosci Hypoth* **2**:193–197.
22 Templeton AR. 2014. The future of human evolution. In: Losos JB (ed.), *The Princeton Guide to Evolution*. Princeton University Press, Princeton, NJ.
23 Reed FA, Aquadro CF. 2006. Mutation, selection and the future of human evolution. *Trends Genet* **22**:479–484.
24 Gazave E et al. 2013. Population growth inflates the per-individual number of deleterious mutations and reduces their mean effect. *Genetics* **195**:969–978.
25 Campbell CD, Eichler EE. 2013. Properties and rates of germline mutations in humans. *Trends Genet* **29**:575–584.
26 Kong A et al. 2012. Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature* **488**: 47–475.
27 Conrad DF et al. 2011. Variation in genome-wide mutation rates within and between families. *Nat Genet* **43**:712–714.
28 Segurel L et al. 2014. Determinants of mutation rate variation in the human germline. *Annu Rev Genomics Hum Genet* **15**:1–24.
29 Scally A, Durbin R. 2012. Revising the human mutation rate: implications for understanding human evolution. *Nat Rev Genet* **13**:745–753.
30 Harris K, Nielsen R. 2014. Error-prone polymerase activity causes multinucleotide mutations in humans. *Genome Res* **24**:1445–1454.
31 Fu Q et al. 2014. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**: 445–449.
32 Chen S et al. New genes as drivers of phenotypic evolution. *Nat Rev Genet* **14**:645–660.
33 Long M et al. 2013. New gene evolution: little did we know. *Annu Rev Genet* **47**:307–333.
34 Long M. 2014. Evolution of new genes. In: Losos JB (ed.), *The Princeton Guide to Evolution*. Princeton University Press, Princeton, NJ.
35 Wu D-D et al. 2011. *De novo* origin of human protein-coding genes. *PLoS Genet* **7**:e1002379.
36 Xie C et al. 2012. Hominoid-specific *de novo* protein-coding genes originating from long non-coding RNAs. *PLoS Genet* **8**: e1002942.
37 Western D. 2001. Human-modified ecosystems and future evolution. *Proc Natl Acad Sci USA* **98**:5458–5465.
38 Myers N. 1987. The extinction spasm impending: synergisms at work. *Conserv Biol* **1**:14–21.
39 Frassetto L et al. 2001. Diet, evolution and aging. *Eur J Nutr* **40**:200–213.
40 Cordain L et al. 2005. Origins and evolution of the Western diet: health implications for the 21st century. *Am J Clin Nutr* **81**:341–254.
41 Zhou W, Chen D. 2009. Encoding human sexual chemosensory cues in the orbitofrontal and fusiform cortices. *J Neurosci* **28**:14416–14421.
42 Singh D, Young RK. 1995. Body weight, waist-to-hip ratio, breasts, and hips: role in judgments of female attractiveness and desirability for relationships. *Ethol Sociobiol* **16**: 483–507.
43 Buss DM. 2013. The science of human mating strategies: an historical perspective. *Psychol Inquiry* **24**:171–177.
44 Courtiol A et al. 2013. The demographic transition influences variance in fitness and selection on height and BMI in rural Gambia. *Curr Biol* **23**:884–889.
45 Laland KN et al. 2010. How culture shaped the human genome: bringing genetics and the human sciences together. *Nat Rev Genet* **11**:137–148.
46 Stearns SC et al. 2010. Measuring selection in contemporary human populations. *Nat Rev Genet* **11**:611–622.
47 Stulp TV et al. 2012. A curvilinear effect of height on reproductive success in human males. *Behav Ecol Sociobiol* **66**:375–384.
48 Ward P. 2009. What will become of *Homo sapiens*? *Sci Am* **300**:68–73.
49 Yoffe E. 2013. My wife is my sister. Slate Magazine, February 19, www.slate.com.
50 Annaluru N et al. 2014. Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**:55–58.
51 Zivkovic B. 2012. Paralyzed patient swills coffee by issuing thought commands. Scientific American Blog Network.
52 Boyle JH. 2013. Adaptive undulatory locomotion of *C. elegans* inspired robot. *Mechanotron Trans* **18**:439–448.
53 Dorrier J. 2014. Worm brain uploaded into Lego robot. SingularityHub.com.
54 Seung S. 2013. *Connectome: How the Brains Wiring Makes Us Who We Are*. Mariner Books, Boston, MA.
55 Tandy C, Bostrom N. 2004. In: *Death and Anti-death, Volume 2: Two Hundred Years After Kant, Fifty Years After Turing*. Ria University Press, Ann Arbor, MI.
56 Bostrom N. 2014. Superintelligence: paths, dangers, strategies. Oxford University Press, Oxford UK.
57 Simonite T. 2014. Software that augments human thinking. January 22, www.technologyreview.com.
58 Lewis C. 2014. Human-computer symbiosis, not Artificial Intelligence, will spur new jobs. January 22, www.roboteconomics.com.

Appendix

General information and bibliography

Before you proceed you should be aware that the sections designated by a number encased in parenthesis (for example (3)) indicates a question to be answer at the end of the exercise.

What is bioinformatics?

Bioinformatics is an interdisciplinary field involving biology, computer science, mathematics, and statistics to analyze biological sequence data, genome content, and sequence alignment, and predict the function and structure of macromolecules [1]. In order to perform these analyses, researchers attempt to find a particular sequence in an existing database.

To compare DNA or amino acid sequences with different lengths, the sequences must first be aligned. Consider the following example.

Example 1. Differences in aligning amino acid sequences.

. . . r v a g f r v d a a k . . .
. . . r v a g f r v d a a k . . .
. . . - v a g f r v d a a k . . .
. . . v a g f r v d a a k - . . .

In the above example you can see 1-10 differences depending on how you align the various sequences (*i.e.*, starting left to right with "v" or "r", or aligning the sequences right to left starting with "k" or alternatively starting in the middle and aligning the sequences at "f" and "r"). Notice that different alignments make a big difference in comparing the sequences. What is the best alignment? There is no clear answer since researchers still debate over this particular issue. Numerous computational methods have been developed to find the optimal alignment, but it is still difficult to say that a certain alignment is always better than others. There are many public and private databases available that often use different algorithms to find the optimal alignment. Some of these databases even provide the user with opportunity to choose an algorithm for a search. Refer to the section on a list of bioinformatics databases.

Even though you are comparing only three sequences, this exercise may take a long time. However, comparing more than three sequences is very common in many research problems. As the number of sequences increases, the number of necessary computations increases at a faster rate. Consider the following formula:

$$\text{\# of necessary computations} = \sum k \quad \text{for } 1 \leq k \leq \text{\#of sequences to be compared} - 1$$

This means that if there are 5 sequences to be compared, then 10 computations are necessary to make the complete comparison ($4+3+2+1=10$). Similarly, 45 computations are needed to compare 10 different sequences. Hence, comparing many sequences can take a very long time, and that is the reason why bioinformatics can be useful. In theory, how many computations are needed to compare 20 different sequences (1)?

Using a bioinformatics database

Bioinformatics databases hold information about numerous DNA, RNA, and amino acid sequences. The information can be accessed by several search methods.

Aligning sequences before comparing them is very important and because almost all DNA, RNA, and amino acid sequences are at least several hundred bases long, it is almost impossible to align them by hand.

Genomes, Evolution, and Culture: Past, Present, and Future of Humankind, First Edition.
Rene J. Herrera, Ralph Garcia-Bertrand, and Francisco M. Salzano.
© 2016 John Wiley & Sons, Ltd. Published 2016 by John Wiley & Sons, Ltd.

Computer instructions

Exercise 1

Below, you will find amino acid sequences of a protein isolated from different organisms (A–C). The sequences are not listed in any particular order.

The purpose of this exercise is to identify the variation in sequences A, B, and C and learn how close these organisms are related.

You will be using the more conventional method to align amino acid sequences. First, pick one of the three sequences from A, B, and C. Then, look at the beginning of your chosen sequence and locate this pattern in the other two sequences. It may or may not be a perfect match. Once you locate the region where it has similar amino acids (conserved region), align the entire sequence so that these regions are matched. Now you can compare the sequences and see how much they differ. Once you have documented the differences between sequences A, B, and C, you will construct a phylogenetic tree that represents the degree of difference found between the sequences. To do this, you will need to determine the number of amino acid differences between the sequences. Start with sequence A and compare it with sequence B. Then compare sequence A with sequence C. Then move to sequence B and compare it with sequence C. Record the number of differences in the sequence comparisons. Then, decide which sequences appear to be more related to each other. Draw out a tree to show how the sequences are related (2).

Amino acid sequences:

A NNNGVIKEVTINPDTTCGND . . . GRGNRGFIVFNN
DDWSFSLTLQTGLPAGTYCDVISGDKINGNCTGI

B NNNGVIKEVTINADTTCGND . . . GTGNRGFIVFNN
DDWQLSSTLQTGLPGGTYCDVISGDKVGNSCTGI

C NSDGTKSVTINADTTCGND . . . GRGDRGFIVFNND
DWYMNVDLQTGLPAGTYCDVISGQKEGSACTGK

Exercise 2. Identification of amino acid sequences.

1 Open an Internet browser.
2 Type http://www.ncbi.nlm.nih.gov/BLAST/ into the address box.
3 At the top of the page choose "SmartBlast" then copy and paste the A amino acid sequences into the box provided and choose BLAST.
4 Scroll down and under the heading "Best hits" look for the amino acid source and name. The top hits will be in blue and a phylogenetic tree will be provided showing the best matches.

Repeat for sequence B and C. Write down the source and name of the amino acid sequences [3].

Exercise 3. Alignment of two sequences.

1 Open an Internet browser.
2 Type http://www.ncbi.nlm.nih.gov/BLAST/ into the address box.
3 Choose "Protein blast".
4 Below the box check "Align two or more sequences".
5 Copy the amino acid sequence #1 (below) into the "Sequence 1" box. (These sequences are already in the supported format [>sequence1].) Be sure to copy everything from ">" on in the sequence.
6 Copy the amino acid sequence #2 (below) into the "Sequence 2" box.
7 Check the box "show results in a separate window" then Press "BLAST" button. In the next window choose "Analyze your query with SmartBlast".
8 The graphic representation of the alignment will show the sequence you placed in the first box as "Unknown" and call it unnamed protein product (yellow bar and dot). The sequence you placed in the second box will appear in blue (solid blue line and dot) and will be identified by the organism it came from, and the table below will show the percent identity for the match.
9 Repeat steps 1–8 with amino acid sequences #3, #4, #5, and #6 (below).
10 Compare "Score" and "Identities" for the trials. Higher numerical values for "Score" and "Identities" indicate that the sequences are closer. What is sequence #1 closer to? Sequence #3 or #6 (4)?

>sequence1
vagfrvdaakhmwpadlaviyxrlknlntdhgfssgskayivqevid
mggeaiskseytglgaitefrhsdsigkvfrgkdqlqyltnwgtawgfa
asdrslvfvdnhdnqrghgaggadvltykvpykmasafmlahpfgt
prvmssfsftdtdqgppttdghniaspifnsdnscsggwvcehrwrqi
ynmvafrntvgsdeiqnwwdngsnqisfsrgsrgfvafnndnydlns
slqtglpagtycdvisssgssctgktvtvgsdgrasinigsseddgvlaihvnak

>sequence2
yvrtkvadymnhlidigvagfrldaskhmwpgdikaildklhnlntkwfs
qgsrpfifqevidlggeavssneyfgngrvtefkygaklgkvmrkwdgek
msylknwgegwglmpsdralvfvdnhdrghgaggasiltfwdarlykm
avgfmlahpygftrvmssyywprnfqngkdvndwvgppnnngktke
vsinpdstcgndwicehrwrqirnmvafrnvvngqpfanwwdndsnq
vafgrgnkgfivnddwalsetlqtglpagtycdvisgdkvdgnctgikvyvg
ndgkahfsisnsaedpfiaihaeski

>sequence3
mflaksivclallavanaqfntnyasgrsgmvhlfewkwddiaaecenflg
pngfagvqvspvnenavkdsrpwweryqpisyklvtrsgneqqfasmv
krcnavgvriyvdvifnhmaadggtygtstaspssksypgvpyssldfnptc
ainnyndanqvrncelvglrdlnqgnsyvrdkvvefldhlidlgvagfrvd
aakhmwpadlaaiygrlknlntdhgfasgsrayivqevidmggeaiskse
yggaitefrhsdsigkafrgkdklqylsnwgtawgfaasdrslvfvdnhdnq
rghgaggadvlthkvprqykmatafmlahpfgtprvmssfsfsntdqgpp
ttdgqniaspvfnsdsscsggwvcerwqinnmvafrnavgsdaiqnww
dngsnqiafsrgsrgfvafnndnydlnsslqtglpagtycdvisgsksgssctg
ktvsvgsdgrasisvgsseddgvlaihvnakl

>sequence4
mkfflllftigfcwaqyspntqqgrtsivhlfewrwvdialecerylapkgfg
gvqvsppnenvaiynpfrpwweryqpvsyklctrsgnedefrnmvtrcn
nvgvriyvdavinhmcgnavsagtscgsyfnpgsrdfpavpysgwdfnd
gkcktgsgdienyndatqvrdcrltglldlalekdyvrskiaeymnhlidigv
agfrldaskhmwpgdikaildklhnlnsnwfpagskpfiyqevidlggeis
sdyfgngrvtefkygaklgtvirkwngekmsylknwgegwgfvpsdral
vfvdnhdnqrghgaggasiltfwdarlykmavgfmlahpygftrvmssy
rwprqfqngndvndwvgppnnngvikvtnpdttcgndwvcehrwrq
irnmvifrnvvdgqpftnwydngsnqvafgrgnrgfivfnnddwsfsltlq
tglpagtycdvisgdkingnctgikiyvsddgkahfsisnsaedpfiaihaeskl

>sequence5
mkfflllftigfcwaqyspntqqgrtsivhlfewrwvdialecerylapkgfg
gvqvsppnenvaihnpfrpwweryqpvsyklctrsgnedefrnmvtrc
nnvgvriyvdavinhmsgnavsagtscgsyfnpgsrdfpavpysgwdfn
dgkcktgsgdienyndatqvrdcrlvglldlalekdyvrskiakymnhlidi
gvagfrldaskhmwpgdikaildklhnlnsnwfpagskpfiyqevidlgg
eissdyfgngrvtefkygaklgtvigkwngekmsylknwgegwgfmps
dralvfvdnhdnqrghgaggasiltfwdarlykmavgfmlahpygftrv
mssyrwprqfqngndvndwvgppnnngvikvtnpdttcgndwvceh
rwrqirnmvnfrnvvdgqpftnwydngsnqvafgrgnrgfivfnnddw
tfsltlqtglpagtycdvisgdkingnctgikiyvsddgkahfsisnsaedpfva
ihaeskl

>sequence6
mqvllllaavglcwaqynpntqagrtsivhlfewrwadialecehylapng
fggvqvsppnenivitnpnrpwweryqpisykicsrsgnenefrdmvtrc
nnvgvriyvdavvnhmcgsmggtgthcgsyfntgtrdfpavpysawdfn
dgkchtasgdienygdmyqvrdcklsslldlalekdyvrstiaaymnhlid
mgvagfridaakhmwpgdirafldklhdlntqwfsagtkpfiyqevidlg
geigsqyfgngrvtefkygaklgtvirkwngekmaylknwgegwgfvps
dralvfvdnhdnqrghgaggasiltfwdarlykmavgfmlahpygftrv
mssyrwpryfengvdvndwvgppsnsdgstsvinadttcgndwvcehr
wrqirnmvifrnvvdgqpfsnwwdngsnqvafgrgdrgfivfnnddwy
mnvdlqtglpagtycdvisgqkegsactgkqvyvssdgkanfqisnsded
pfvaihvdak

Aligning multiple sequences

To align more than two sequences, you are going to learn how to use another bioinformatics database called ClustalW.

Exercise 4. Aligning multiple sequences and creating phylogenetic tree.

1. Open an Internet browser.
2. Type www.ebi.ac.uk/clustalw into the address box.
3. Copy and paste all amino acid sequences provided, #1 to #6. (These sequences are already in the supported format [>sequence1].)
4. Click on "Submit" and wait for the results.
5. Scroll down the page to see the subheading "Alignment". Notice how all six sequences are aligned.
6. Copy the multiple sequence alignment. Add this to your report to be handed in.
7. Go to the top of the page and choose "Phylogenetic Tree". At the bottom of the page choose Cladogram Tree.
8. Print out a cladogram and a phylogram tree. Attach this to your lab answer sheet.

Creating a phylogenetic tree

Exercise 4. Aligning multiple sequences and creating phylogenetic tree (continued).

9. Type www.ebi.ac.uk/clustalw into the address box of the new browser.
10. Choose WebPrank and Paste the sequences from step 6 into the box.
11. Submit.
12. Click on "Submit" and wait for the results.
13. Compare these results to those from step 5 above.

Do the results differ? [5] Looking at the phylogenetic tree from #7 above and the identifications of the sequences from exercise 3, do species relationships in the trees make sense? Why or why not? [6].

Now, use a BLAST search to find the source organism and common names for each of the amino acid sequences #1 to #6. On all phylogenetic trees, replace the numbers with the source organism common names.

Do species relationships in these trees make sense? Why or why not (7)?

Using protein structure repository

The function of proteins is determined by their structure. This is why it is important to study protein structure. Some bioinformatics databases contain visual representations of proteins, and they can be very useful in studying protein structures.

Exercise 5. Working with protein 3D structures.

1. Open an Internet browser.
2. Type http://bioinf.cs.ucl.ac.uk/psipred/ into the address box.
3. Choose "Predict secondary structure" under "Prediction Method".
4. Copy and paste amino acid sequence #4 from Exercise 3 into "Input Sequence". This sequence is also listed below:

```
mkfflllftigfcwaqyspntqqgrtsivhlfewrwvdialecerylapkg
fggvqvsppnenvaiynpfrpwweryqpvsyklctrsgnedefrnm
vtrcnnvgvriyvdavinhmcgnavsagtscgsyfnpgsrdfpavpys
gwdfndgkcktgsgdienyndatqvrdcrltglldlalekdyvrskiaey
mnhlidigvagfrldaskhmwpgdikaildklhnlnsnwfpagskp
fiyqevidlggeissdyfgngrvtefkygaklgtvirkwngekmsylkn
wgegwgfvpsdralvfvdnhdnqrghgaggasiltfwdarlykmav
gfmlahpygftrvmssyrwprqfqngndvndwvgppnnngvikv
tnpdttcgndwvcehrwrqirnmvifrnvvdgqpftnwydngsnq
vafgrgnrgfivfnnddwsfsltlqtglpagtycdvisgdkingnctgik
iyvsddgkahfsisnsaedpfiaihaeskl
```

5. Enter your e-mail address.
6. Enter "Exercise 5" as a short identifier.
7. Click the "Predict" button. The server will run PSIPRED and a new page pops up to inform you that the prediction job has been submitted and to check your e-mail for the results in approximately 5 min.
8. Check e-mail for PSIPRED results.
9. At the top of the e-mail, click the link to the results page.
10. Scroll to the bottom of the results page and click "Download PDF Version".
11. Print out the PDF results and attach this to your lab answer sheet.
12. What do the purple cylinders on the PDF printout represent (8)? How many helices are found in this 3D protein (9)?

Visualizing protein structure

You can now use a protein database to visualize the 3D structure, which should correspond with the results obtained from secondary protein folding.

Exercise 5. Working with protein 3D structures (continued).

13. Type www.rcsb.org/pdb into the address box.
14. In the search box, you can type in a protein name or its PDB ID. The PDB ID can be obtained from bioinformatics databases, such as the BLAST. The PDB ID starts with letters "PDB" on the results page of the databases. For the purpose of this exercise, type in 3OLD and click on "Search". This is the PDB ID for the protein that the above amino acid sequence codes for and should correspond with the predicted secondary structure from the first part of Exercise 5.
15. On the right-hand side of the results page, you will see a graphical representation of the protein.

Copy the structure and attach it to your lab.

16. Click on different "more images" to take a closer look at the protein structure.
17. Click on "Related PDB Entries" to see which other similar proteins are in the database. Write down the ID number of one similar protein (10).

Exercise 6. BLAST search.

1. Open an Internet browser.
2. Type http://www.ncbi.nlm.nih.gov/BLAST/ into the address box.
3. Go to "Nucleotide BLAST"—under the Basic BLAST heading.
4. **(a)** Copy and paste the following sequence into the FASTA sequence box:

```
aggaggcgcagcggagcccttggcctcagtcagtcaggcgctggggagc
gtttcggtttcacttccggtgaggggccgcgcctgagagggcgggcagtga
agcaaacggacggcgagcgcgggcggtcagtgacggcggcgccgctgcc
ggggggcgtgcggtaacgcggcggcggcggcggcggcgacggcggctg
ggcctcaagcgcctgcagcccacctcccggaggcgggctcccggcgcga
ggacggaggaagatggaggagctggtggtggaagtgcggggctccaatg
```

(b) Choose the "Mouse genomic and transcript" database.

(c) Under Program Selection choose "Highly similar sequences".

(d) Go to the bottom of the page and check the box "Show results in a separate window".

5 Before you use the search function, do you notice any interesting pattern in the sequence? If so, what is it (11)?

6 Now click on "BLAST!"

7 On the next page, you will see the message "Your request has been successfully submitted".

8 Another Internet browser will open automatically and it will give you an estimate of how long the search will take. (*Useful tip*: Search takes longer during the day due to high traffic in the website.)

9 BLAST search results provide a lot of information. The top line of the color diagram is the one that most closely fits your data. (You can put the cursor on that line and it will provide further information above the box.) The numbers in the box below the colored lines are statistical measurements used to determine the probability of a match.

The lower the value of E, the higher the probability of a match.

Below the box, you will also see a list of DNA sequences that are similar to the sequence you had put in the "Search" box. The best match will appear on the top of the list. Copy the identification number following "NM". Also copy the query coverage. Which gene is the sequence from (12)?

10 Click on "the identification number" (i.e., NM_008031).

11 This should provide you with the locus and the size of the mRNA.

12 Go back to the Nucleotide BLAST page and choose the "Human genomic and transcript" database and "Somewhat similar sequences", and redo the BLAST. Copy down the NR accession number and query coverage information as above (13).

13 Click on the NR accession number 024503.1. Write down the title of the article, authors, journal, and year of publication (14).

Exercise 7. RNA BLAST search.

You can also use a BLAST search for RNA and amino acid sequence searches.

1 Open an Internet browser.

2 Type http://www.ncbi.nlm.nih.gov/BLAST/ into the address box.

3 Go to "tblastx"—under the Basic BLAST heading.

4 Copy and paste the following sequence into the FASTA sequence box:
GGGGATATAGCTCAGTTGGGAGAGCGCTTGAATG
GCATTCAAGAGGTCGTCGGTTCGATCCCGATTATC
TCCACCA

5 Click "BLAST".

6 The page will refresh and continue refreshing until the results are formatted. This may take up to several minutes.

7 The results page will open. At the top of the page, you see basic information about the query sequence submitted. The color diagram shows how closely each hit fits your data. Scroll down to the "Description" section. The hit at the top of the list is the best match.

8 Click on the link to the first hit. Write down the NR accession number and the source organism of this RNA sequence (15). How many DNA base pairs make up this gene and is the DNA linear or circular (16)?

Using RNA secondary structures

RNA is single stranded, so bonds form internally between complementary base pairs, stabilizing RNA molecules. Bioinformatics can help us visualize these secondary structures.

Exercise 8. Predicting RNA secondary structure.

1 Open an Internet browser.

2 Type http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form in the address box.

3 Enter "Exercise 9" as a name for the sequence.

4 Copy and paste the same sequence from Exercise 8 into the sequence box.

5 Scroll down, keeping all default settings, and click on "Fold RNA".

6 On the next page, you will see the message "Your job is being processed".

7 When the output page appears, scroll down until you find the section entitled "View Individual Structures". Mfold looks for arrangements that yield secondary structures with the lowest possible energy and are therefore most stable. These structures are ranked according to their stability. Structure 1 is more stable than structure 2, and so on. Next to each structure, a measurement of free energy or maximum work attainable (dG) is listed. Write down the free energy associated with each structure (17).

8 Click on the PDF of structure 1. Print out the PDF results and attach this to the answer sheet. Now look at the PDF of structure 2 and look for differences between structures.

Exercise 9. Creating a contig and finding the sequence identity

Below are a series of sequences. Use the six sequence reads to create a sequence contig of this part of the human genome. Do this by hand and write down the sequence contig you found (18).

Read 1	ATGCGATCTGTGAGCCGAGTCTTTA
Read 2	AACAAAAATGTTGTTATTTTTATTTCAGATG
Read 3	TTCAGATGCGATCTGTGAGCCGAG
Read 4	TGTCTGCCATTCTTAAAAACAAAAATGT
Read 5	TGTTATTTTTATTTCAGATGCGA
Read 6	AACAAAAATGTTGTTATT

List of bioinformatics databases

Finding the right data

Name	Address	Description
Ensembl	www.ensembl.org	The human genome
GenBank/ DDBJ/EMBL	www.ncbi.nlm.nih.gov	Nucleotide sequence
PubMed	www.ncbi.nlm.nih.gov	Literature references
NR	www.ncbi.nlm.nih.gov	Protein sequences
SWISS-PROT	www.expasy.ch	Annotated protein sequences
InterProScan	www.ebi.ac.uk	Protein domains
OMIM	www.ncbi.nlm.nih.gov	Genetic diseases
PDB	www.rscb.gov/pdb	Protein structures
KEGG	www.genome.ad.jp	Metabolic pathways

Analyzing your DNA/RNA sequence

Name	Address	Description
Webcutter	www.firstmarket.com/cutter	Restriction map
PCR	biotools.umassmed.edu/bioapps	PCR primer design
GenomeScan	Genes.mit.edu/genomescan/	Gene discovery
blastn, tblastn, blastx	www.ncbi.nlm.nih.gov	Database search
The Genome Browser	Genome.cse.ucsc.edu	Browse the ultimate data
Mfold	www.bioinfo.rpi.edu	RNA structure prediction

Analyzing your protein sequences

Name	Address	Description
BLAST	www.ncbi.nlm.nih.gov	Database homology search
SRS	srs.ebi.ac.uk	Database search
Entrez	www.ncbi.nlm.nih.gov	Database search
InterProScan	www.ebi.ac.uk	Find protein domains
ExPASy	www.expasy.ch	Analyze a protein
ClustalW	www.ebi.ac.uk	Multiple sequence alignment
T-Coffee	igs-server.cnrs-mrs.fr/Tcoffee	Evaluate multiple alignment
Jalview	www.es.embnet.org	Multiple alignment editor
PSIPRED	bioinf.cs.ucl.ac.uk/psipred/	Secondary structure prediction
Cn3D	www.ncbi.nlm.nih.gov/Structure	Display and spin 3D structures

Questions

Name: ____

Answer questions (1) through (18) as you go on with the lab manual. Answer all the questions with complete sentences.

1 How many computations are needed to compare 20 different sequences? Show your work. (Bioinformatics exercise.)

2 Draw out a tree to show how the sequences are related. (Using bioinformatics database: DNA, RNA, or amino acid sequence search.)

3 Write down the source and name of the amino acid sequences A, B, and C.

A ____

B ____

C ____

4 Compare "Score" and "Identities" for the trials. Higher numerical values for "Score" and "Identities" indicate that the sequences are closer. What is sequence #1 closer to, sequence #3 or #6? Why?

5 Briefly explain how the results differ.

6 Do species relationships in the tree make sense? Why or why not?

7 Copy a tree and attach the common names (sequences 1–6). Do species relationships in these trees make sense? Why or why not?

8 What do the purple cylinders on the PDF printout of the secondary protein structure represent?

9 How many helices are found in this 3D protein?

10 Using the PDB ID provided, write down the name of one similar protein.

11 Do you notice any interesting pattern in the sequence?

12 From your blastn search, copy the identification number following "NM" and copy the query coverage. How long is the complete DNA sequence? What is the source organism of the DNA sequence?

13 Copy down the NR accession number from your search.

14 From the NR accession number, write down the title of an article, authors, journal, and year of publication.

15 Write down the NR accession number and the source organism for this RNA sequence.

16 How many DNA base pairs make up this gene and is the DNA linear or circular?

17 Write down the free energy measurement for both structure 1 and structure 2.

18 Write down the sequence contig you found from the series of six short sequences.

Reference

1 Mount DW. 2004. *Bioinformatics: Sequence and Genome Analysis*. Cold Spring Harbor Laboratory Press, New York.

Index

A Y haplogroup 186
A, B, O blood group 82
Abdur site 186
acculturation 59, 60, 79, 80, 183, 193, 196
acculturation changes 235, 236, 238
ACE 62
acetylation 85
Acheulean tools 126
admixture 49, 53, 59, 60, 62, 87–88, 98–99, 132, 133, 138, 192, 196, 197, 201
Afro-Asiatic languages 136
agricultural revolution 59, 79, 80, 82,148–151, 193, 194, 195, 199, 214, 232, 234–235
Aleut 194
allopatric speciation 89
alpha-globin 56
Altaic 56, 133, 136
altruism 212–213
Alu repetitive elements 45, 48, 55, 73, 80, 81, 184
Amerinds 193, 194
Ami 51, 54, 195, 197
amniocentesis 216
AMOVA 44, 49, 51
Anatolia 59, 79, 80, 82, 193, 194
ancestral state 44, 48, 130, 131, 184
ancient common polymorphism 132
Angelman Syndrome 85
animal markers 79, 198
  chickens 79, 199
  dogs 79, 136, 198–200
  pigs 79, 199
  rats 79, 199
anticipation, genetic 47
apomorphic 130
Arc1 dog mtDNA haplotype 198
Arc2 dog mtDNA haplotype 198
Ardipithecus ramidus 13, 130
Aristotle 208
art 151–153
artificial intelligence 236–238
artificial selection 213, 236–237
Ashkenazi Jews 62
assignment test 45
assisted reproductive technology (ART) 216–219
Atapuerca, Spain 133
Atayal 51, 54, 195
Australia, colonization 133, 186, 187, 190–194, 196, 198
Australian aborigines 191–193
australopithecines 12, 13, 125, 126, 130–132
Austrolopithecus gracile 131
Austrolopithecus robust 131
Austronesian expansion 194–196, 198–200
autosome variability 28–30
Ayu al-Buhaira site 136

B Y haplogroup 186
Bab-el-Mandeb, Strait 185, 186, 188, 190, 191
back to Africa 135, 136, 138, 184–186, 188–190
bacterial and molecular genetics 11
balance polymorphism 69, 77
Bantu expansion 48, 59, 60
Batan Islands 195
Bayesian statistics 53
behavioral differences 207, 212–213, 232, 236, 238
Bentham, J. 209
Beringia 56, 193
beta-globin 56, 69, 77, 78
bioinformatics exercises 241–247
biosimilars 220, 222
bipedalism 12, 124, 126, 127
Bonferroni correction 45, 49
bootstrap 44, 49, 50
bottlenecks 49, 58–60, 193, 201
brain 12, 125–128, 130–132, 134, 135, 213, 234–235, 237–238
Bruno's genetic distance 50
Bunun 51, 54, 195
burials 132, 138

cancer 47, 69, 70, 171–172
Candelabra theory 192

Genomes, Evolution, and Culture: Past, Present, and Future of Humankind, First Edition.
Rene J. Herrera, Ralph Garcia-Bertrand, and Francisco M. Salzano.
© 2016 John Wiley & Sons, Ltd. Published 2016 by John Wiley & Sons, Ltd.

canoes 195
carcinogens 71
Cavalli-Sforza's genetic distance 50
CCR5Δ32 61, 62
CD4 61
cerebral cortex 127, 128
Chi-square test 45
chloroplast DNA 49
chorionic villi sampling 216
chromatin 85
chromosome rearrangements 23, 24, 28–30, 71, 73. 74, 90, 96–97, 102–103, 108–110, 234
chromosomes 43, 44, 46, 47, 49, 72
chromosomes (synthetic) 237
climate changes 75, 126, 230–231, 234–236
clinal diversity 64
coalescent time 44, 45, 49
coastal route 191, 193
coding and non-coding sequences 46, 48, 71, 73, 77, 78, 81, 84
coevolution gene/culture 48, 91, 125, 142–146, 231
consanguinity 57
contour analysis 44, 49, 51, 52, 61
Cook Islands 198, 201
copy number variation (CNV) 30, 93–95, 101–108, 117
cosmic rays 70
CRISPR/cas9 219, 221, 237
culture 37, 59, 60, 80, 136, 141–157, 182, 183, 198, 199
  concept 141, 231
  origins and development 141–142, 231–232
  transmission and conditioning factors 142–146
cultural markers 79
cystic fibrosis 58, 61, 75

D chicken mtDNA haplogroup 201
Daic 194, 195, 197
Danube River 186
Dart, R. A. 12
Darwin, Ch. 5–7, 11, 15, 87–88
Darwinian medicine 160
de novo genes 30, 31, 105, 234
deamination 71, 74
degenerative diseases 172, 173
deletions, chromosomal 48, 69, 73, 75
Denisova 49, 130, 133, 134
dentatorubral-pallidoluysian atrophy 47
derived state 48, 130, 131, 184
Diamond, J. 133
diasporas, human 48, 59, 60, 79, 124, 193, 195
dideoxy nucleotides 83, 84
diet 61, 75, 125, 131, 235–236, 238
Dinaledi 131, 132
dingo 193, 198
directed evolution 206, 214, 227
disease classification 158–160
Dmanisi 134
DNA polymerase 44, 47, 70, 71, 74, 83
DNA profiling 60, 62, 206, 224–226
DNA repair 71, 72, 75, 163–165
DNA-based marker system 41–47, 52, 80, 82
Dobzhansky, Th. 11
domestication 48, 59, 79, 82,148–151, 194, 199
DUF1220 128, 129
duplicons 127, 128

E chicken mtDNA haplogroup 201
East African Ridge 125, 126, 184
Easter Island 195, 197–202
Ebu Gogo 132
ecocide 201
ecogenetics 174–176
Edward's genetic distance 50
effective population size 94–97
El Arenal 201
el Wad Cave 136
electrophoresis 82, 83
encephalization 126, 213
ENCODE Project 21, 22, 38, 78, 80–82
endogenous retroviruses 48
enhancers 46, 73, 84
entangled bank theory 196
environmental influences 55–58, 61, 68, 70, 71–78, 81, 82, 125–127, 134, 183, 184, 194, 200
epigenetics 84, 85, 91, 213, 225–226
Epipaleolithic Kebaran 136
epistasis 75, 77
Eskimo 194
ethics 207–208, 212–216
eugenics 214–217, 219–220, 226
evolutionary medicine, *see* Darwinian medicine
Ewens-Watterson test 45
exact test 45
excision repair 72
exomes and proteomes 30, 31, 92
exons 73
express train theory 196
eye color 75, 76

Fertile Crescent 194
Fiji 198
fitness 55, 71, 77
FNBP2 127, 128
foramen magnum 124, 126, 130, 131
founder effect 53, 56, 57, 62, 193

free will, morality, religion 153–155, 207–209, 212–213
F_{st} 44
F-statistics 45

GD (genetic diversity) 44
gene conversion 71, 73, 74, 88, 95, 112
gene family 78
gene flow 53, 59, 60, 79, 80, 183, 184, 187, 192, 196, 197, 201, 202
gene function 69–73, 76–78, 80–83, 85, 166, 167
gene pool 55–60, 62, 63, 69, 70, 75, 80, 133, 134, 136, 184, 186, 187, 196
gene therapy 179, 219–222
genetic counseling 176–178, 224
genetic diversity 42, 44–55, 57–60, 64, 68, 69, 71, 73–76, 78, 80, 82–84, 133, 134, 184, 193, 201
genetic drift 50, 56–58, 60, 88, 94, 108, 192, 193
genetic load 77, 81, 82, 85
genetic manipulation of animals 166, 167, 220–222, 236
genetic testing 223
genetic variability 13, 14, 28, 47, 49, 55, 57, 58, 69–71, 92–95, 193
genome wide association studies (GWAS) 102, 113–117
germ line 69, 70, 73, 85
Gibraltar, Strait 136
glacial maxima 185, 189
glacial periods 136, 185, 188, 189, 193
global warming 193
Goldstein's genetic distance 50
GpC islands 74
gradualism 88–90
Gravettian art 126
Greenberg's tripartite theory 194

Han Chinese 195
haplogroups 44, 52, 53, 60, 86, 188–192, 196, 197, 199–201
haplotypes 43–46, 49, 52, 56, 60, 93, 98–101, 186, 187, 189, 192, 198–201
HapMap 100
Hardy-Weinberg equilibrium 44, 49, 95
Hawaiian Islands 195, 198
health, concept 158
height 75
Herto fossils 135, 137, 138
HERV-K 48
heterozygosity 44
highly reiterated sequences 81
histones/non-histones 84
history 3–14
  up to the 16th century 3, 4
  17th, 18th century 4, 5, 13
  19th century 5–10, 12
  20th century 10–14
HIV 49, 61
holocene 192, 196
hominins 59, 124–128, 130–135, 184, 185
hominins, early 12, 124–131, 134–137
Homo antecessor 134
Homo erectus 13, 125–128, 130, 131, 132, 134, 135, 192
Homo floresiensis 131, 132
Homo habilis 12, 13, 126, 128, 131,132
Homo heidelbergensis 126, 132, 134, 135
Homo naledi 130–132
Homo neanderthalensis 12, 125, 130, 132, 134
  habitation 138
  Hunting strategies 138
Homo sapiens 63, 131–135, 138, 184
Horn of Africa 134, 136, 184–186, 188–191
Huang River Basin 194
Hula dancing 195
human genome 16–37, 42–48, 54, 55, 68, 69, 71, 73, 74, 76, 78, 80–82, 127, 196, 202
  functional aspects 21–23
  normal and abnormal phenotypes 19–21
  structural aspects 17–19, 107–113
  whole genome scans 53, 133, 192, 194, 197, 199, 201
Human Genome Sequencing Project 16, 17, 207, 213
human leukocyte antigens 100–102, 113, 194, 201
Hunter-gatherers 193
Huntington decease 47
hyoid bone 125
hypermethylation/hypomethylation of DNA 84
hypertension 62
hypervariable DNA 47, 74

imprinting 85
imputation 91, 115, 117
in vitro fertilization 217–219, 236
inbreeding 44, 45, 49, 57, 58, 168
individual identification 218–219
Indonesia 132, 194–200
infectious diseases, evolution 162–163
infertility 232, 236–237
influenza virus 115
information revolution 194
insertions/deletions (In/dels) 28, 29, 45, 48, 92, 96, 99, 109–113, 184
interbreeding 130, 133, 134, 138
intergenic DNA 71, 73, 81
introgression 132, 133, 138, 201
introns 46, 73, 79
inversions, chromosomal 71, 75

isolation 57, 58, 62, 63, 77, 78, 192, 193, 200
isozymes/allozymes 82

Jamal site 136
Jebel Faya site 136, 137, 185, 186
junk DNA 80–82

K mtDNA haplogroup 188, 189
K value 54
Kant, I. 209
Kebara site 136, 137
Khoisan 186

L1 mtDNA haplogroup 188
L1 repetitive elements 80
L2 mtDNA haplogroup 186, 188
L3 mtDNA haplogroup 186, 188
lake Victoria 125
Lamarck, J. B. 5, 6
language 59, 60, 63, 79, 125, 131, 136, 46–148, 182, 183, 194–196, 199, 207
Lapita culture 79, 195, 199
last glacial maximum 189, 193
Levallois technique 186
Levant corridor 59, 62, 134, 136, 138, 184–187, 189, 193
Lewontin, R. 64
life expectancy and population growth 220, 231–232
LINES 46, 48, 80
linguistic diversity 79, 195
linkage equilibrium and disequilibrium 95–102, 201
Linnaeus, C. 63
lithic tools 185, 186, 192
Locke, J. 208, 210
LTR 46

μ 44, 45, 49
M mtDNA haplogroup 192
M1 mtDNA haplogroup 187–190
M347 Y chromosome mutation 192
M78, Y chromosome mutation 188, 189
macroevolution 89
Madagascar 195
major genes 75, 76
major histocompatibility complex 100–102
malaria 55, 56, 69, 78
Malayo-Polynesian 195
malignancy 73
mantel test 45
marae 198
marker systems 183
Marquesas Islands 198, 200
maternal blood sampling 216
maximum likelyhood 45
Mayr, E. 1, 2, 5, 9, 11, 14, 16
MDS (multidimensional scaling) 44, 49, 51
medical applications 206
medical technology 206, 219–221, 237
medicine 60, 81
meiosis 43, 49, 73, 74, 85
meiotic drive 95, 113, 236
Melanesia 133, 186, 187, 190, 191, 194, 196, 197, 199, 200
Mendel, G. 5–8, 10, 49, 61, 75, 82
Mesolithic 189
metagenomics 91
metastasis 73
methods, population analyses 13, 14, 42–45, 49, 53, 83, 84, 94, 97, 182, 201
methylation 21, 22, 27, 38, 74, 84, 85, 92, 96, 109
microbiome 91
microevolution 87, 89, 231
Micronesia 196, 197, 199
microsatellites, also see short tandem repeats, variable number of tandem repeats 46, 50, 94, 102, 106
middle repetitive elements 80, 81
migration 48, 56–60, 62, 80, 87–89, 92–93, 98, 126, 132, 134–138, 184–200, 234–235
minisatellite 46
mitosis 73
modern humans 48, 59, 60, 125–128, 130–138, 184–187, 189, 191–193, 230
modifying genes 75, 76, 79
molecular clocks 80, 183
Monte Verde, Chile 56, 193
morality 153–155
morphometrics 42, 49, 75, 76, 79, 131, 183, 193
Mount Carmel Range 136
Mousterian 137
mtDNA haplogroups 189, 190, 192, 196, 201
mtDNA haplotypes I, II, III (rat) 200
mtDNA variability 28, 42–44, 49, 52, 58–60, 133, 136, 184, 186–190, 192, 194, 196–202, 221
multiregional theory 192
Mungo Lake 192
mutagens 71, 77
mutation (synonymous – nonsynonymous) 88, 93, 70, 71, 83
mutation rates 44, 45, 47, 49, 71, 73–75, 93,163–165, 183, 231–234
mutations 44, 45, 47–50, 52, 53, 55, 56, 60–62, 69–78, 81–83, 127, 132, 183, 184,186, 187, 190, 192
myelination 127

N mtDNA haplogroup 192
Na-Dene 194
Native Americans 193, 194, 201, 202

Near East 59, 133, 134, 136–138, 190, 194
Nei's Da genetic distance 50
Nei's genetic distance 50
neighbor-joining analysis 44
Neolithic 59, 136, 187–190, 193–195, 198–200
nephrite 195
nervous system and culture 37
network analysis 44, 49, 51–53, 81
neurons 56, 127, 128
new genes evolution 48, 73, 78, 107, 112, 127, 128, 234
New Guinea 133, 191, 192, 194, 196, 199
New Zealand 195, 198, 200, 201
next generation sequencing (NGS) 89–92, 104–105, 114–115
non allelic homologous recombination (NAHR) 55, 73, 95–96, 103, 107–108, 112
non-random mating 49, 57–59
Nozick, R. 211
NRY (non-recombining Y) 43, 49, 189, 196
nucleosomes 85

Oldowan tools 124–126
Omo fossils 135, 137, 138
Orchid Island 195
Out of Africa 48, 58–60, 124, 126, 131, 134–136, 138, 184–188, 190, 192, 194
Out of Taiwan 196, 199, 200

P60 Y chromosome mutation 192
Paiwan 51, 195
paleogenomics 25–28
Paleolithic 48, 56, 59, 82, 127, 136, 137, 187–190, 193, 194
Panini 130
Paranthropus robustus 12, 13, 130, 131
parent-offspring conflict 161
pathogen history 161, 162
PCA (principal component analysis) 44, 50
PCR 47
Peruvian Slave Trade 201
pharmacogenetics (pharmacogenomics) 62, 92, 118, 174–176
phenotype 68, 69, 75–79, 82, 85
Philippines 195, 196, 198, 199
phosphorylation 85
phylogenetics 44–46, 48–50, 52, 76, 78, 79, 80, 82, 85, 130, 131, 184, 192, 196, 199–201
Pithecanthropus 12
plants and animals dispersal evidence 79, 194, 195, 198–201
plate tectonics 125
Pleistocene 184, 194
plesiomorphic 130
polymorphic retroviral insertions 48
Polynesia 51, 54, 57, 79, 195–202
Polynesian motif 196
population variability 13, 14, 42, 44–64, 69, 74–79, 81, 82, 85, 133, 134, 138, 182, 183 184, 186, 188, 192, 193, 195–197, 201, 202, 231–233
Prader-Willi Syndrome 85
pre-HV1 mtDNA haplogroup 188–190
pre-implantation genetic diagnosis 216–217
Prevosti's genetic distance 50
primates 48, 55, 125, 128
privacy 206, 216, 222–224
probability of inclusion 62
promoters 46, 73, 81, 84
punctuated equilibrium 78, 88–90
Puyuma 51, 54, 195
pygmies 186

Qafzeh site 136–138, 185, 186
quantitative trait loci (QTLs) 61, 75–77, 79, 114–116

Ra'iatea 198
race 60–64, 132, 134, 199, 207, 214–215
racism 64
radiation 70–71
Rapa Nui 198, 201, 202
rats 198–200
Rawls, J. 209
recombination and evolution 43, 73, 88, 92–103, 105–108, 113
repetitive elements (sequences) 45–48, 55, 73, 78, 80, 81,108
replication slippage 47, 74, 81
reproductive fitness 167, 168
reproductive isolation 63, 74, 78
reproductive technologies 214, 216–218
retroviruses 48
reverse transcription (reverse transcriptase) 109, 111
Reynolds, Weir, Cockerham's genetic distance 50
RFLP 46
Rh system 82
RNA 46, 68, 77–79, 81, 83–85
  processing 71, 77, 81
  RNA, 7SL 48
  rRNA 68
  snRNA 68
  tRNA 68
rough endoplasmic reticulum 48, 81
Rukai 51, 54, 195

Saharan pump 135
Sahelanthropus tchadensis 126, 130
Sahul 192
Saisiyat 51, 54, 195
Samoa 197, 198, 200
Sandel, M 211
Sanger sequencing 83

Savanna thesis 126
science and philosophy 1–3
science, politics and ethics 16, 17, 206, 213–216
segmental duplications 46, 48, 73, 78, 81, 90, 95, 102–103, 107–108, 127, 128
selection 31–37, 45, 46, 48, 49, 55–57, 60, 61, 69, 71, 72, 76–83, 88–93, 100–102, 106–110, 126, 127, 130, 230–231
selection or drift? 31–37
 analyses 34–37, 100–102
 history 31, 32
 methods 32–34, 100–102
self awareness 127
selfish DNA 80, 82, 108–113
sequencing methods 46, 83, 84, 90–92, 134, 184
sex ratio 23–25
sex selection 236
short tandem repeats (STRs) 44–48, 74, 106, 194
Sima de los Huesos 133
simulation studies 200
SINES 46, 48, 55, 80, 81
single gene traits 61, 75, 76
single nucleotide polymorphism (SNPs) 45, 46, 48, 88, 91, 93, 105
 (blocks and haplotypes) 98–102
 (chips and tags) 114–117
sickle-cell 55, 56, 69, 75, 77, 78
Skhul site 136–138, 185, 186
slow boat theory 196
social evolution 207
Society Islands 198, 200
Socrates 208
Somali Plate 125
somatic 69, 70, 73, 85
South America/Polynesia contacts 200
stem cell therapy 206, 220, 222
Stone Age 127, 136
Strait of Hormuz 186
Strait of Sorrows 134, 136
structure analysis 52, 54, 201
Stuart Mill, J. 209
subspecies 63
sweet potato 79, 198, 200, 201
symbiotic relationship 42, 43
synthetic theory of evolution 11

Tabun site 136–138
Taiwan 51, 54, 55, 58, 79, 194–200
tattoos 195
telomeres 70, 74
teratogenesis 163–165
thermoregulation 127
thrifty gene hypothesis 235
thymidine bridges 70, 74
Tierra del Fuego 57
time to most recent common ancestor (TMCA) 196
Toba eruption 58
Tobias, P. V. 12
Tonga 197, 198, 200
Tor Faraj 136
trans-humanism 237–238
translocations, chromosomal 71, 75
transmission/disequilibrium test 99, 102
transposable elements 30, 109–113, 234
treatment of genetic diseases 178, 179
Tsou 51, 54, 195
tumors 69, 73
Tuskegee syphilis experiment 213–215

ultraviolet light 70, 71
UNESCO 63
uniparental inheritance 44, 49, 52, 59, 80, 85, 136, 186–188, 192, 197
UTRs 73

variable number of tandem repeats, VNTRs 47, 103, 106, 109, 111
violence 168–171, 207

Wallace, A. R. 7–9
Watterson test 45
whole mitochondrial genome (WMG) 199
world views of nature 1

Yami 51, 54, 195
Yangtze river basin 194
YAP insertion 189
Y-chromosome haplogroups 187–189, 191, 196, 197

Zhiren Cave 136, 138, 185