CAPITAL UNIVERSITY OF SCIENCE AND
TECHNOLOGY, ISLAMABAD

Genome-wide Identification and Analysis of Amino Acid Permeases in *Physcomitrella patens*

by

Taimoor Abdul Sattar

A thesis submitted in partial fulfillment for the
degree of Master of Science

in the

Faculty of Health and Life Sciences
Department of Bioinformatics and Biosciences

2024

I dedicate my work to the love of my life, Maham Imran, and my brother and mentor, Imran Afzal, both of whom I hold very close to my heart.

CERTIFICATE OF APPROVAL

Genome-wide Identification and Analysis of Amino Acid Permeases in *Physcomitrella patens*

by

Taimoor Abdul Sattar

(MBS-221010)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Ishtiaq Hassan	Hazara University Mansehra
(b)	Internal Examiner	Dr. Muhammad Asad Anwar	CUST, Islamabad
(c)	Supervisor	Dr. Sohail Ahmad Jan	CUST, Islamabad

S A Jan

Dr. Sohail Ahmad Jan
Thesis Supervisor
March, 2024

Dr. S. Marriam Bakhtiar
Head
Dept. of Bioinfo. and Biosciences
March, 2024

Dr. Sahar Fazal
Dean
Faculty of Health and Life Sciences
March, 2024

Author's Declaration

I, **Taimoor Abdul Sattar** hereby state that my MS thesis titled "**Genome-wide Identification and Analysis of Amino Acid Permeases in *Physcomitrella patens***" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

(Taimoor Abdul Sattar)

Registration No: MBS-221010

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Genome-wide Identification and Analysis of Amino Acid Permeases in *Physcomitrella patens***" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

(Taimoor Abdul Sattar)

Registration No: MBS-221010

Acknowledgement

All praise and thanks to Allah Almighty, who never left me alone, despite all of my weaknesses especially when I felt a need to be listened to. Indeed, to him belongs all power and knowledge, and without him, we are nothing but helpless travelers on a very dark and lonely road of life.

I'd like to start by thanking my dear professor and friend Dr. Hamid Saeed, who has always been a source of motivation and guidance. His wisdom and kindness have helped me in times when I felt most helpless and without comprehension. I am also thankful to my dear sister Maham Imran, who was always there for me when I needed her the most. Her unconditional love and compassionate nature has made me the man I am today, and my achievements will forever reflect my gratitude and love for her.

My special thanks to my supervisor, Dr Sohail Ahmad Jan, for making this work possible and for all the valuable lessons that I learned during this period. I'd also like to thank Dr. Sami Ullah Jan, for his guidance whenever I needed it. Finally, my warmest thanks to Maha Ghias, for her assistance in formatting my work.

(Taimoor Abdul Sattar)

Abstract

Amino Acid Permeases (*AAPs*) belong to a class of Amino Acid Transporter (*AAT*) protein family, which play a pivotal role in the transportation and selective transit of various biomolecules such as Nitrogen, Potassium, Sodium and various amino acids based on their size, structure, charge and though their specialized binding sites. Cells are unable to synthesize some amino acids and must obtain them from the environment, and thus, *AAPs* are critical players in this process. *AAPs* have been widely studied in vascular plants but they have not been explicitly reported in non-vascular bryophytes, to date.

In the present study, *In-silico* tools have been employed for the identification and characterization of *AAPs* in the non-vascular bryophyte model *Physcomitrella patens*. A total of 16 *P. patens AAPs* (*PpAAPs*) were identified that shared physical and chemical attributes with *AAPs* of *Arabidopsis thaliana* (*AtAAPs*). The selected *PpAAP* sequences shared a common domain with scale *AtAAPs*, confirming they belong to the same gene family. Furthermore, the average gene lengths of *PpAAPs* were found to be significantly higher than those of *AtAAPs* while the average protein lengths of the two were almost similar with the average of *PpAAPs* slighter higher than the latter. Similar was the instance for GRAVY (Grand Average of Hydropathicity) values where *PpAAPs* were higher as compared to *AtAAPs*. However, the average molecular weight (MW) and Theoretical Iso-electric point (pI) of *AtAAPs* was found higher than those of *PpAAPs*. Online tools suggested that all *PpAAPs* are hydrophobic and localized in the Plasma membrane, and share a significant degree of homology in their gene structures and protein motifs with *AtAAPs*. Phylogenetic analysis showed that *PpAAPs* possess evolutionary divergence and variation among them while substantial evolutionary linkage was observed with *AAPs* of several other vascular plants, confirming common ancestry. The closest neighbors of *PpAAPs* were observed to be the *AAPs* of *Cocus Nucifera*, *Vicia faba*, *Glycine max*, *Eucalyptus grandis*, *Zea mays*, *Cannabis sativa*, *Brassica rapa*, *Brassica napus* and *Arabidopsis thaliana*. Protein-protein interactions show that the interacting proteins of the majority of *PpAAPs* were involved in the role of transportation of biomolecules. Results proposed that the

PpAAPs indeed belong to the Amino Acid Permease (*AAP*) gene family and shared significant structural and functional homology with *AtAAPs*.

Contents

List of Figures

List of Tables

Abbreviations

Aa_Trans	Amino Acid Transporters domain
A. thaliana	Arabidopsis thaliana
AAAPs	Amino Acid/Auxin Permeases
AAPs	Amino Acid Permeases
AATs	Amino Acid Transporters
APC	Amino Acid Polyamine-Organocation
AtAAPs	Arabidopsis thaliana Amino Acid Permeases
AtCATs	Arabidopsis thaliana Cationic Amino Acid Transporters
AtLHTs	Arabidopsis thaliana Lysine Histidine Transporters
AtProTs	Arabidopsis thaliana Proline Transporters
B. napus	Brassica napus
C	Celsius
CATs	Cationic Amino Acid Transporters
CD	Candela (unit to measure light intensity)
CHASE	Cyclases/Histidine kinases Associated Sensory Extracellular domain
GC	Guanine-Cytosine
GRAVY	Grand Average of Hydropathicity
GW	Grain Weight
IGPS	indole-3-glycerol-phosphate synthase domain
LHTs	Lysine-Histidine Transporters
MW	Molecular Weight
NLPs	Nodule Inception-like Proteins
NUE	Nitrogen Use Efficiency
O. sativa	Oryza sativa

OsAAPs	Oryza sativa Amino Acid Permeases
pI	Isoelectric point
PlSc	Phosphate acyltransferases domain
PpAAPs	Physcomitrella patens Amino Acid Permeases
ProTs	Proline Transporters
P. patens	Physcomitrella patens
qPCR	quantitative Polymerase Chain Reaction
RI	Recombinant Inbred
Tg	Teragram
TM	Transmembrane Domain
UMAMIT	Usually Multiple Amino Acids Move In And Out
V. faba	Vicia faba
VfAAPs	Vicia faba Amino Acid Permeases
ZnMc	Zinc-dependent metalloprotease domain

Chapter 1

Introduction

The amino acids, which constitute the fundamental blocks of proteins, are essential in many metabolic activities that occur within cells. While cells can synthesize some amino acids, others must be obtained from the environment. Amino acid transporters, which are integral proteins that are found in cell membranes, play an important role in the ingestion of these amino acids. They are specialized proteins that assist transportation across cellular membranes. They are a group of proteins that recognize and bind to certain amino acids and help them move into or out of the cell based on physiological demands and concentration gradients [1]. Amino acid transporters have been reported in several vascular plants and many important bio-molecules such as Nitrogen (N), Potassium (K) and Sodium (Na) are in direct interaction with these transporter families which are responsible for permeability [2]. Research suggests that these nutrients play an essential role in the reproduction, development, and growth of plants [3].

Amino acid transporters ($AATs$) can be further divided into several divisions according to the specificity of their functions such as uptake property. $AATs$ are further classified in two families which are the Amino Acid-Polyamine-Organocation Family ($APCs$) and the Amino Acid/Auxin Permease family ($AAAPs$) [3]. Amino acid Permeases ($AAPs$) are a sub-family within the Amino Acid/Auxin Permease ($AAAP$) family. Generally, $AAPs$ play important roles in the uptake of amino acids from the soil, long-distance transport of amino acids from soil to other parts

of the plant through the phloem, efficient nitrogen use, synthesis or development of proteins, reproduction, and stress response. In the model plant *Arabidopsis thaliana*, 8 *AAPs* have been identified. These *AAPs* (*AtAAP1-AtAAP8*) play different roles as reported by several studies for instance, *AtAAP1* is known to import nitrogen for seeds in the development stage by loading amino acids in the phloem. *AtAAP2* is known to be expressed in the vascular tissue and is involved in the long-distance transport of amino acids, ensuring a steady supply of nutrients. *AtAAP3* is involved in the export of amino acids from leaves and is expressed in the phloem while *AtAAP4* is involved in pollen development. Similarly, *AtAAP5* is expressed in the roots and is involved in the uptake of amino acids from the soil while *AtAAP6* plays a vital role in the redistribution of amino acids during stress conditions. Some *AAPs* have fewer known functions as in the case of *AtAAP7*, which is known to play a role in nutrients and ion transport. *AtAAP8* has been reported to supply amino acids in developing embryos [3].

In a similar study on rice (*Oryza sativa*), researchers focused on the function of four *AAPs* of *O. sativa* in particular, namely *OsAAP1, OsAAP3, OsAAP7* and *OsAAP16* using electrophysiology and found that *OsAAP1, OsAAP7* and *OsAAP16* could transport a wide range of amino acids across the plasma membrane and their functional patterns were found similar to those studied earlier in *A. thaliana* [4]. Interestingly, *OsAAP3* was found to display selective transport against aromatic amino acids, a trait not observed previously even in *A. thaliana*. However, it effectively transported specific amino acids such as lysine and arginine. Researchers also concluded that these *OsAAPs* could transport both, positively charged and neutral forms of these amino acids effectively. The localization of these transporters was also established and plasma membrane was found to house these transporters, similar to *Arabidopsis* as reported in previous studies [4].

Identically, another study aiming to functionally characterize *AAPs* in broad bean (*Vicia faba*) was conducted. Researchers achieved this by isolating full-length cDNAs responsible for encoding three AAPs, namely *VfAAP1, VfAAP3* and *VfAAP4*. It was found that *VfAAP1* and *VfAAP3* were responsible for the

transport of a wide range of amino acids such as cysteine and arginine/lysine respectively. It was also found that *VfAAP1* played a significant role in supplying amino acids for storage protein synthesis, and that *VfAAP3* was mainly expressed in maternal tissues while no detectable transcripts for *VfAAP4* were found. Similarly, expression regulation and localization were done for these permeases with the conclusion that these were localized in parenchyma cells while different developmental stages affected the pattern of expression [5].

1.1 Problem Statement

In the present study, *Physcomitrella patens* was chosen as the organism of choice for studying *AAPs* in a non-vascular plant model. Transporter proteins have been extensively studied in vascular plants over the years, but less study has been conducted on the structural and functional characterization of *AAPs* in non-vascular plants.

1.2 Research Gap

AAPs have been identified and characterized both functionally and structurally in many vascular and non-vascular plants. However, no study has been conducted on the structural and functional characterization of *AAPs* in *P. patens* to date.

1.3 Scope

This study is an attempt to bridge the gap of structural and functional homology between vascular and non-vascular plants and predict any unknown functions in members of *AAPs* of *P. patens*. Less studies exist on the comparison between vascular and non-vascular plant gene families and the reason such studies should exist is because they provide valuable ground and notable predictions for future studies. Hence, this study may ease *in-vivo* functional characterization of *PpAAPs*

in any future studies, which, in turn, may prove beneficial for crop yield, efficient use of nitrogen, and various genetic engineering techniques.

1.4 Aim and Objectives

1.4.1 Aim

- This study aims to identify and structurally characterize the *AAP* gene family in *P. patens*.

1.4.2 Objectives

- To identify and characterize *AAP* gene family structurally in *P. patens* using computational tools.
- To study the physicochemical properties of *AAP* gene family in *P. patens*.
- To study the evolutionary relationship of *AAP* gene family between vascular and non-vascular plants.
- To study the interacting proteins of *AAP* gene family in *P. patens*.

Chapter 2

Literature Review

2.1 General Overview of *Physcomitrella patens*

P. patens is a moss (bryophyte) that is used as one of the model organisms for research on the physiology, development, and evolution of plants. It is a non-vascular plant which means that it does not have a well-established xylem and phloem, and this is a reason why it is a good model organism for comparative studies with vascular plants. In modern research, it was first documented by Engel in 1968 in which the moss was grown in the laboratory, and various biochemical and morphological mutants were introduced in the plant [6]. According to that study, under specific circumstances, the moss *P. patens* can complete its full life cycle in between seven and eight weeks. When cultivated at temperatures between 15°C and 19°C, it can reproduce sexually. Researchers used a variety of methods to cause mutations in this moss, including X-rays, ethyl methane sulfonate, and N-methyl-N'nitro-N-nitrosoguanidine. They produced a variety of mutants, including those with altered color (yellow mutants), modified shape or structure (morphological mutants), and the inability to generate specific compounds (such as thiamine, para-aminobenzoic acid, niacin, and fungal extract). Compared to typical moss, the yellow mutants contained 35–65% less chlorophyll [6].

P. patens has received a great deal of interest in botany and plant biotechnology. It is a tiny species of moss that can be found in temperate regions everywhere.

Its ease of development and manipulation, its thoroughly sequenced and well-annotated genome, and its distinctive evolutionary position, all contribute to its attractiveness as a model organism [7]. *P. patens* is quite simple to grow and manipulate in laboratory settings. It is a cost-effective option for many laboratories since it needs little room and resources to thrive. Importantly, this moss demonstrates a high level of genetic adaptability, including an elevated level of homologous replication, a characteristic that is uncommon in plants. This makes it a perfect model for the study of the function of genes and genetic pathways and enables precise genetic modification. Sequencing and comprehensive annotation of the *P. patens* genome have been completed. This has created new possibilities for comparative genomics research, allowing scientists to follow the development of numerous genetic features and functions. *P. patens*, which was a member of the first class of plants to colonize land some 450 million years ago, holds a special place in the evolution of plants. Studying *P. patens* can help us understand how plants evolved to live on land from aquatic environments, including how they adjusted to abiotic stresses such as water scarcity changed how they reproduced, and developed complicated structures like organs and tissues [8].

2.1.1 Physiology & Biochemistry of *Physcomitrella patens*

P. patens has a three-month life cycle and does best in open, disturbed environments. Although it can survive in some dry environments, it needs water for reproduction just like other mosses do. The other phases throughout its life cycle react to light, the force of gravity, and various forms of stress, but its spores need light to germinate. Researchers have carefully examined *P. patens*' responses to various kinds of light. They have discovered that it exhibits a range of growth reactions when exposed to elevated amounts of light and that its development can vary depending on the direction of polarized light [9]. Due to *P. patens's* easy visualization and manipulation of its cells, it was additionally utilized to explore cell polarity. Cytokinins, auxins, and abscissic acid hormones that are typical of mosses and flowering plants have been examined in *P. patens*, and various mutants

have been developed. Hormones and other variables can cause the moss to transition from a two-dimensional to three-dimensional growth. Calcium is essential for *P. patens* cell polarity, which affects how the organism's cells branch and develop. The ability of cytokinin to induce three-dimensional development in moss cells provides a novel method for investigating the function of these hormones. It is simple to follow the moss' development from spores to sophisticated structures, a trait not frequently found in other examples of model organisms. Rhizoids, or the roots of moss plants, separate from the plant's base, and the leaves develop in a spiral shape. Reproductive organs develop when the appropriate circumstances are met, and fertilization happens when water is supplied. Additionally, moss may interact with other living things, such as fungi, but nothing is known about how *P. patens* respond to pathogens. Last but not least, although biochemical research on *P. patens* is not nearly as sophisticated as genetic study, scientists have discovered several genes involved in metabolism and anticipate that *P. patens*' biochemical nature may be fairly similar to that of other terrestrial plants [10].

2.1.2 Life Cycle of *Physcomitrella patens*

A dominating haploid gametophyte as a morphologically distinctive, diminished diploid sporophyte alternates throughout *P. patens* life cycle, which is typical of all bryophytes. By apical tip growth, a germinating meiospore develops into filamentous protonemata that contain two distinct cell types that represent the juvenile gametophyte: chloronema cells, which have more chloroplasts and cell walls that are parallel to the growth axis, and caulonema cells, which have fewer chloroplasts as well as oblique cell walls. A gametophyte having basal rhizoids and an upright, leafy stem that may develop sexual organs (gametangia) is produced by caulonemal cells as meristematic buds having three-faced apical cells. due to *P. patens*' monoecious nature [11]. The life cycle of *P. patens* is depicted in Figure 2.1.

The life cycle of *P. patens* was well observed by Engel in a study in which morphological and biochemical mutants were induced [6]. According to the study, this

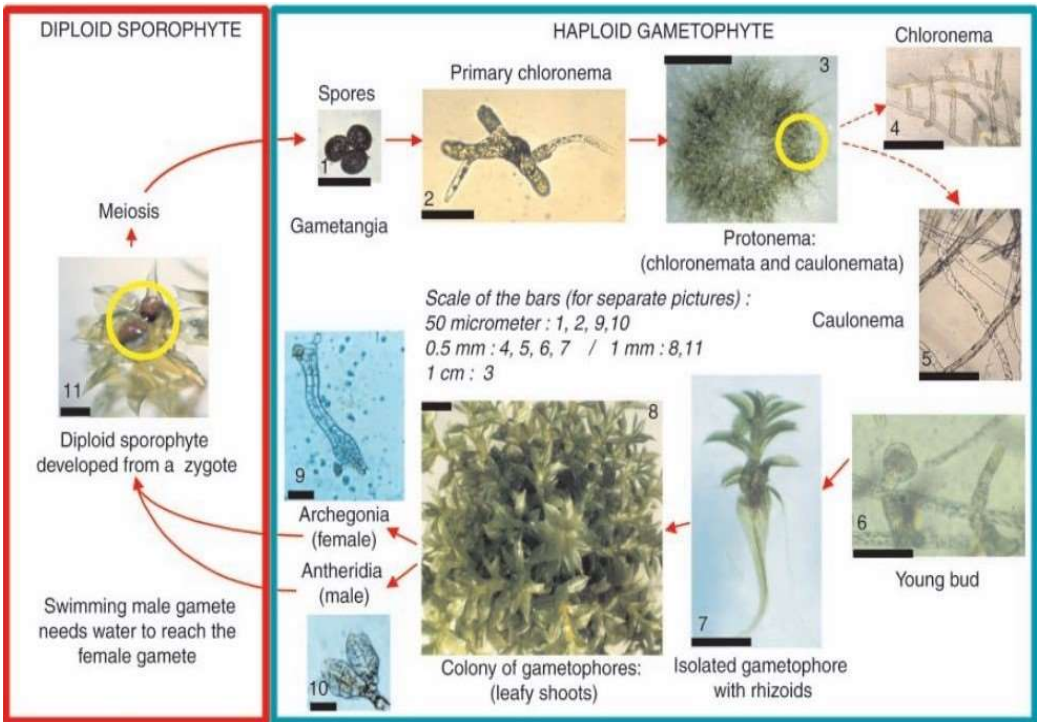


FIGURE 2.1: Life cycle of *P. patens* Source: Adapted from [10].

moss is cleistocarpous (creating a specific kind of moss spore capsule that fails to open when mature) and monoecious (containing reproductive organs from both, male and female organs within the same plant). The gametophore, or sexual form of the plant, has a small seta (stalk) that measures between 0.25 and 0.50 mm and connects the spore capsule of the moss to it. This capsule lacked an operculum, a lid-like appendage commonly found in other moss species, in keeping with the distinctive features of cleistocarpous mosses. A single fertilization event produced 3–4 thousand spores in the capsule, which a viable count supports. *Physcomitrella's* spores are roughly 20 x 40 micrometers in size and have an elevated viability rate of 90–95%. Given conditions of 25°C to 30°C and 200 to 300-foot cd (a unit that measures the intensity of light) under constant illumination, they can germinate in 48 to 72 hours. On solid or liquid media, the germinated spores develop into filamentous colonies of size 7–10 mm in 2–3 weeks [6].

After 2-3 weeks, each colony of these filaments, referred to as green protonemata, bears 10-15 gametophores. The upright gametophores were 0.5-1.0 millimeters

long and bear 5–6 bracts that resemble leaves after this phase. It is worth mentioning that protonemal and gametophoric cell inoculum can also grow into a colony. The sexual reproduction of *P. patens* had been observed in earlier investigations, but the crucial significance of low temperature had not been shown by these scientists. According to the results of this study, 90–95% of the *P. patens* gametophytes cultured at 15–19°C develop male (antheridia) and female (archegonia) reproductive organs, which lead to the development of spore capsules. Only cultures kept at cooler temperatures (15-19°C) for 4-6 weeks were shown to have sexual reproduction. The experiment showed that light-dark regimes had no discernible impact on sexual reproduction. Between 19°C and 21°C, the production of capsules is reduced by 80%, and at 23^oC, no capsules are formed. The researcher also noted that growing colonies produced from spore or vegetative inoculum on minimum media for 2-3 weeks at 25–30°C is the standard approach for promoting sexual reproduction. The colonies are then transferred to a colder environment that is kept at 15-19°C. 7-10 days after lowering cultures to cooler temperatures, antheridia, and archegonia start to grow on gametophores. At the reduced temperature, mature spore-containing capsules begin to form after 3–4 weeks. 150–300 capsules can be produced from an aseptic culture of 15–30 colonies on a solid medium in a 25 X 150 mm culture tube [6].

2.2 *Arabidopsis thaliana* as a Model Organism

The scientific community initially concentrated a great deal of its attention on a procedure known as mediated cell transformation. *Arabidopsis*, a little plant from the mustard family, was introduced, and the story started to shift. The first person to recognize its genetic potential was the European researcher Laibach, whose baton was later picked up by Re'dei in the United States [12].

With the creation of its genomic map and the opportunity for further genetic investigation, the 1980s were an important period for *A. thaliana*. Additionally, its small genome provided a benefit for thorough genetic and molecular analysis, distinguishing it in research of plant physiology. It is impossible to overstate 1987's

	P. patens	*A. thaliana*
Genome size	487 Mbp	157 Mbp
No. of chromosomes	27	5
Average intron length	252 bp	146 bp
Average intron number	5	5
Intron G/C	40%	33%
Exon G/C	50%	40%

FIGURE 2.2: Comparison between genomes of *P. patens* and *A. thaliana*. Source: Adapted from [13]

importance in the context of the *A. thaliana* study. An important event conducted at Michigan State University's campus served as a stimulus for the launch of an online discussion forum for *Arabidopsis* research. Ambitious goals were established in 1990 as this concept started to gain popularity among scientists. These included in-depth genetic analyses that ranged from substantially mutating its genome to completely decoding its genetic makeup. The larger scientific community anticipated that knowledge gained from *A. thaliana* may have repercussions, helping industrial applications, medical sciences, as well as agriculture. The fact that the *Arabidopsis* Genome Initiative originated in 1996 only serves to emphasize its significance.

In addition to serving as a monument to the plant's importance, this international cooperative effort signaled an essential change regarding how substantial genetic projects were undertaken. The project allowed for the cataloging of a sizeable amount of the plant's genomic DNA sequence in open databases. The trajectory of *A. thaliana*, from an underappreciated weed to a pillar in the field of genetic model organisms, represents a scientific miracle, that had an objective to completely sequence the entire 120-Mb genome by 2000. In summary, *Arabidopsis*' journey has cemented its status as a crucial tool for illuminating complex features of plant physiology and genetics [12]. Figure 2.2 shows the genomic differences between *P. patens* and *A. thaliana*.

2.2.1 The Genome of *Arabidopsis thaliana*

The genome of *A. thaliana* is divided into five chromosomes and contains about 20,000 genes [12]. Analyzing the chromosome structure was difficult because of the tiny meiotic chromosomes and lack of polytene chromosomes. However, subsequent developments have improved visualization, such as in *in-situ* hybridization techniques. Each chromosome has been visualized using three different mappings for instance, based on recombination frequencies, the classical genetic map displays the locations of mutant genes. By looking at segregating phenotypes in self-pollinating plants, the initial map was generated. Currently, there are about 460 mutant genes on this map, which is available online (http://mutant.lse.okstate.edu). Notably, over 110 of these genes have been cloned. The positions of genes that were cloned and molecular markers are shown on the Recombinant Inbred (RI) Map. This map was created using a particular population of plants that frequently self-pollinated. This map, as is also accessible online (http://nasc.nott.ac.uk/new_ri_map.html) now displays more than 790 markers. Online access to physical maps of the *Arabidopsis* genes provides current information on the chromosomes. It is also available online (http://genome-www.stanford.edu/Arabidopsis/) and is updated regularly as more studies are being conducted on the subject [12].

2.3 Nitrogen Use Efficiency (NUE) in Plants

Nitrogen use efficiency (NUE) measures a plant's ability to produce grain in relation to the quantity of nitrogen in the soil. It is defined mathematically as GwlNs, when 'Gw' refers to grain weight, showing the total quantity of grain generated, and 'Ns' represents nitrogen availability, representing the quantity of nitrogen available to the plant. This efficiency is divided into two major components:

- **Absorption (Uptake) efficiency**: This refers to the extent to which the plant can soak up or absorb nitrogen accessible from the soil.

- **The efficiency of utilization**: After absorption, this indicates how well the plant can use the ingested nitrogen to produce grain. It measures the crop's metabolic and physiological efficacy in turning ingested nitrogen into productive grain yield [14].

A significant quantity of nitrogen fertilizer is sprayed globally each year, primarily on cereal crops. Despite the vast quantities applied, its use is poor, with just 33% of total nitrogen collected in grain. Unutilized nitrogen can pollute the environment, cause greenhouse gas emissions, and be costly due to variations in nitrogen fertilizer costs. In agriculture, there is an urgent need to improve nitrogen fertilizer efficiency [15]. Nitrogen is carried through the root to the shoot via its xylem, and from high in nutrients leaves to nutrient-demanding regions via the phloem. Sink organs that require nutrients typically have limited xylem intake due to their poor transpiration activity. Phloem loading is the process of supplying nutrients to the phloem, which happens in the small veins of leaves. Unloading, which occurs in sink organs, is the act of removing or dispersing such nutrients from the phloem. The transport phloem is the fundamental component of the phloem system, connecting the loading and unloading processes. This network also provides for the interchange of nitrogen compounds in plant components such as stems, roots, and leaves between the xylem and phloem [16].

2.3.1 Effects of Nitrogen Use on Environment

Nitrogen fertilizer consumption has increased worldwide, from roughly twelve Tg in 1960 to around 113 Tg in 2010 [17]. This trend is projected to continue as the global population grows. Nitrogen fixation, nitrification, assimilation, ammonification, and denitrification are all phases in the nitrogen cycle. Various substances escape into the atmosphere during this cycle, influencing the climate. These chemicals, which include greenhouse gases (GHGs) such as carbon dioxide, carbon monoxide, and nitrogen dioxide, are contributing to global warming, which can have a negative impact on the security of food and agriculture. N_2O is particularly problematic because of its extended atmospheric lifetime and strong Global

Warming Potential (GWP), which is 310 times that of CO_2. Agricultural nitrogen consumption increases NH_3 and NO_x emissions, which are contributing to climate change indirectly. These compounds can either warm or cool the atmosphere by eliminating methane or generating light-scattering aerosols. Furthermore, greater nitrogen usage can result in greater retention of carbon in soils due to better primary productivity and litter generation [17].

2.3.2 Improving NUE by Altering Transport Mechanism of Nitrogen

Determining the rate of nitrogen absorption in plants, especially cereals, is critical in agronomy for optimizing nitrogen usage efficiency (NUE). Recent discoveries have shed light on the mechanisms behind nitrogen transportation in vegetation, but the optimum tactics for increasing NUE remain unclear. Increasing the capability or selectivity of transporter proteins at the root epidermis may theoretically boost nitrogen uptake, however, the existing uptake capacity is significant [15].

Complex molecular networks control absorption, assimilation, and intraplant dispersion in the nitrogen transport process. External influences, like changes in the environment that affect soil nitrogen levels, add an additional level of complication. Although high-affinity transport systems (HATS) are primarily responsible for nitrogen absorption, low-affinity transport systems (LATS) may play important roles in particular circumstances, such as specific soil temperatures.

The lack of empirical data on the real contribution of these carriers in field conditions makes it difficult to identify viable genetic targets for increasing NUE. According to some studies, rather than manipulating individual components, changing complete regulatory cascades may be more beneficial. For example, in *A. thaliana*, amplification in the *Dof1* transcription factor resulted in a 30% rise in plant nitrogen content, evading the plant's typical regulatory constraints [15].

2.3.3 Improving NUE by the Use of Specialized Proteins

Research has shown that proteins such as *NLPs* (Nodule-Inception-like Proteins) may improve Nitrogen Use Efficiency. *OsNLP1*, a protein isolated from rice, or *Oryza sativa* is essential for nitrogen utilization efficiency (NUE) in the plant. This protein is found in the cell's nucleus and reacts swiftly to nitrogen deficiency. Increasing *OsNLP1* levels can improve the development of rice, yield of grains, and NUE under a variety of nitrogen circumstances. When nitrogen is scarce, deleting *OsNLP1* reduces both grain production and NUE. *OsNLP1* regulates several genes involved in nitrogen utilization, including those involved in nitrate and ammonium absorption and assimilation. The direct adherence of *OsNLP1* to the promoters of the genes promotes their expression. As a result, *OsNLP1* is critical for optimizing nitrogen usage in rice and represents a prospective path for increasing both nitrogen uptake and rice output [18].

2.4 Amino Acid Transporters and their Role

The exchange of nitrogen in plants happens through amino acids [19]. Nitrogen is essential for the growth of plants and reproduction. Plants acquire both inorganic (such as ammonia and nitrates) and organic nitrogen forms from the soil. Once ingested, nitrogen passes through an absorption process that involves transforming nitrate to ammonium and finally to amino acids. This can happen with either the root or the source leaves. This process normally produces glutamine or glutamic acid as the initial organic molecule, which then aids in the production of other amino acids. These amino acids can be synthesized in an array of cell compartments, which include plastids, mitochondria, and cytoplasm. Asparagine and glutamine are contained mostly in the xylem sap, whereas all amino acids are transferred by the phloem. These amino acid amounts can vary depending on the species of plants and environmental variables. The aforementioned amino acids are essential for nourishing various portions of the plant, such as the tips of the

roots and flowers. As a result, *AATs* are critical in ensuring the right amount of organic nitrogen across the plant [3].

2.4.1 Classification of Amino Acid Transporters

Plant *AATs* are essential for the transportation and dissemination of amino acids. According to their sequence homology as well as how they take up drugs, these types of transporters can be divided into two groups. They are as follows:

2.4.1.1 Family of Amino Acid/Auxin Permeases (*AAAP*).

- *AAPs* (Amino Acid Permeases): transporters that handle a wide variety of amino acids with no specialization.
- Lysine and Histidine Transporters *(LHTs)*: These specialize in the transport of lysine and histidine, as the name implies.
- Gamma-aminobutyric Acid Transporters *(GATs)*: For Gamma-Aminobutyric acid transport since they are specialized transporters, specific to their transport.
- Proline Transporters *(ProTs)*: They are in charge of proline transport.
- Indole-3-Acetic Acid Transporters *(AUXs)*: They are responsible for facilitating the movement of indole-3-acetic acid, a kind of auxin that is essential for plant growth.
- Aromatic and Neutral Amino Acid Transporters: They only transport aromatic and neutral amino acids.
- Amino Acid Transporter-like Proteins: These kinds of proteins may have structural similarities to existing amino acid transporters, but their functions may be different or as yet unidentified [3].

2.4.1.2 *APC* (Amino Acid-Polyamine-Organocation) Family:

- Cationic Amino Acid Transporters *(CATs)*: are in charge of transporting cationic amino acids.
- Amino Acid/Choline Transporters: Proteins that are responsible for the movement of both amino acids and choline.
- Polyamine H^+ Symporters *(PHSs)*: They're enzymes that transport polyamines in tandem with protons (H^+).

There is also a newer set of transporters known as 'Usually Multiple Acids Move In and Out Transporters' *(UMAMIT)*. This family was just discovered in *A. thaliana*. This points to the possibility of additional identification and classification of protein transporters in plants [3]. Figure 2.3 shows the classification of *AATs*.

2.4.2 Description of Various Amino Acid Transporters in *Arabidopsis thaliana*

Arabidopsis plants rely on a complex network of *AATs*, which are predominantly localized in their roots. These transporters are classified into three types: *AAPs, LHTs*, and *ProTs*. *AAPs*, such as *AAP1* and *AAP5*, are found in various regions of the root and aid in the absorption of certain amino acids. *LHTs* are recognized for their high-affinity systems of transport, and they specialize in carrying lysine and histidine, but they also transport additional neutral and acidic amino acids. *LHT1*, for example, has been found in both *Arabidopsis* and rice roots, highlighting its relevance in amino acid absorption. Finally, *ProTs*, particularly *ProT2*, are responsible for facilitating the movement of proline, glycine, and -aminobutyric acid (*GABA*). These transporters guarantee efficient nutrient uptake, which contributes to the general health and growth of the plant [3]. There are fourteen *APC* transporters in the *A. thaliana* genome [20]. Nine of them belong to cationic amino acid transporters (*AtCAT1-9*), that contain 14 putative transmembrane (TM) domains. They specialize in essential amino acid transport with great affinity. The

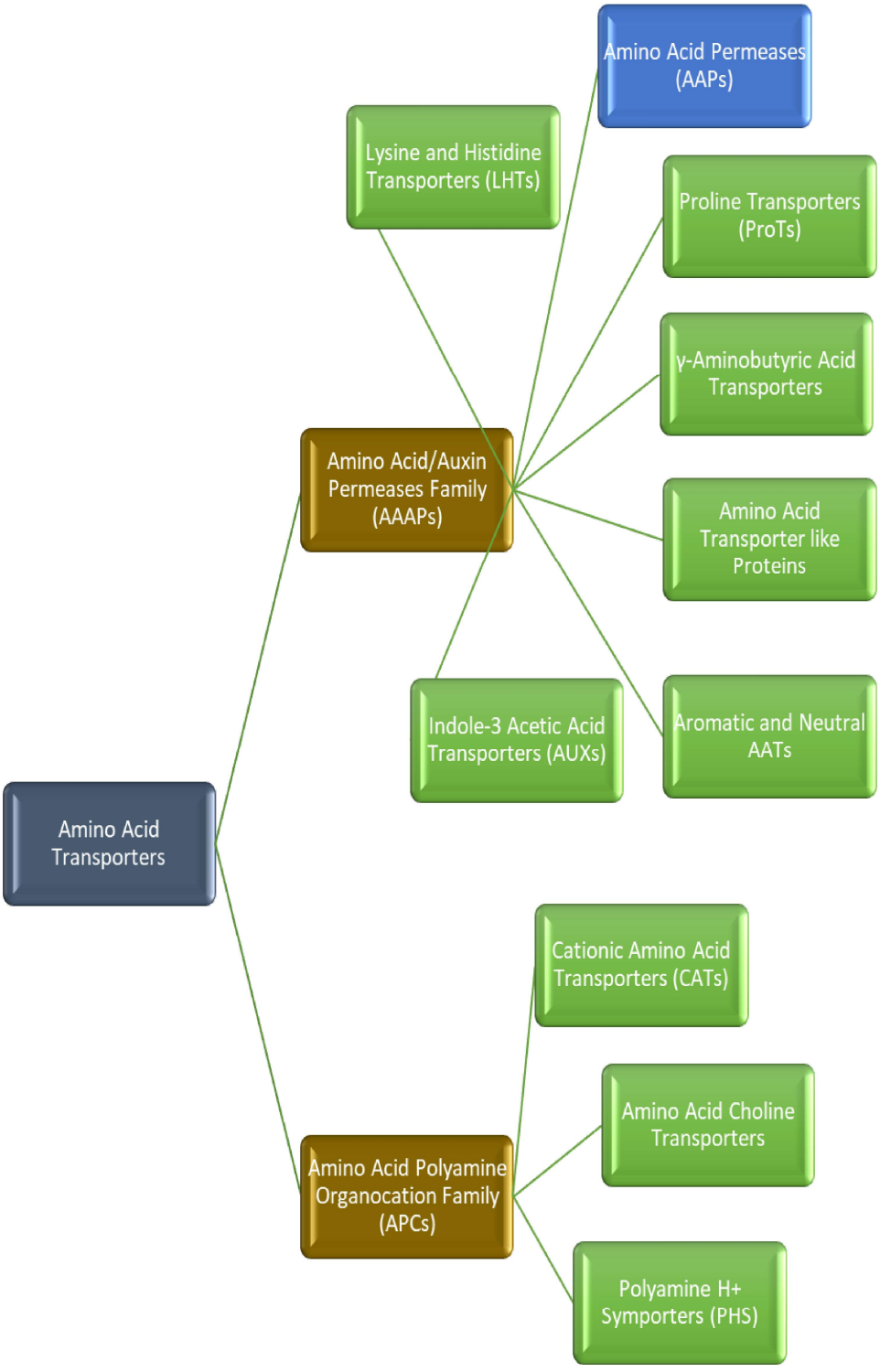


FIGURE 2.3: Classification of Amino Acid Transporters with reference to Amino Acid Permeases *(AAPs)* Source: Data modified from Yao et al. [3]

other five *APC* members have 12 transmembrane domains, one of which is a possible gamma-aminobutyric (*GABA*) transporter. Furthermore, the *Arabidopsis* genome contains the *ATF* superfamily of amino acid transporters, which has roughly 46 members. Amino acid permeases (*AtAAPs*), lysine and histidine transporters (*AtLHTs*), proline transporters (*AtProTs*), aromatic and neutral amino acid transporters (*AtANTs*), and potential auxin transporters (*AtAUXs*) are the five subclasses of this superfamily [20]. Given the abundance of amino acid transporters, it is clear that they serve a variety of activities that are controlled by factors such as substrate specificity, expression site, and environmental triggers. This complication motivates researchers to delve thoroughly into the intricacies within every transporter gene family.

2.4.3 Expression of Various Amino Acid Transporters in *Arabidopsis thaliana*

Numerous amino acid transporters in the *Arabidopsis* genome have specialized functions in plant development and food intake. Because of the way they function in stems, flowers, and siliques, transporters *AtAAP1, AtAAP2, AtAAP4*, and *AtAAP5* likely assist in loading phloem from source tissues and may also carry amino acids to developing embryos. Other transporters, such as *AtAAP3*, which is mainly expressed in the roots, may aid in the uptake of amino acids from the phloem or topsoil. Furthermore, *AtAAP6*, which is present primarily in roots and leaves, appears to be critical for the uptake of amino acids from the xylem, owing to its high affinity. Other amino acid transporters, such as *AtLHT1* and *AtLHT2*, appear to be required for the uptake of amino acids from soil and specialized cellular roles. *Arabidopsis* also has 3 amino acid proline transporters, *AtProT1, AtProT2*, and *AtProT3*, all of which have different expression patterns and roles. *AtANT1* is known for its broad expression, whereas *AtCAT1* is known among the *APC* superfamily for its possibility of versatility in phloem physiology. Other *AtCATs*, on the other hand, have unique roles and patterns of expression across tissues. Surprisingly, the genes that are orthologous to amino acid transporters

might have different expression patterns in various plant species, implying different evolutionary paths. Although tomatoes have proline transporters similar to *Arabidopsis*, their activities and expression patterns differ. The many roles of these transporters in plants highlight their relevance in the development of plants, nutrition, and adaptation [20].

2.5 Amino Acid Permeases

AAPs are specialized proteins that play an important role in cell physiology. They act as gatekeepers in the cell's membrane, directing the passage of amino acid molecules through and out of a cell [21]. Integral membrane proteins suggest *AAPs* remain profoundly buried within cellular membrane lipid bilayers. In contrast to peripheral membrane proteins, which adhere to the membrane's surface, integral proteins are those that span the whole length of the membrane, typically with domains on both the outside and inside of the cell. As a result, they can form conduits or pores across which amino acids can flow, ensuring the selective transit of these important molecules. One of the key tasks of these *AAPs* is to mediate amino acid transport. Because of this mediation, *AAPs* can specifically allow specific amino acids to travel through while prohibiting others. They accomplish this selectivity by combining the structure and charge of the protein, as well as the existence of specialized binding sites designed to recognize only particular amino acids or classes of amino acids. Amino acids are required by all cells for a variety of reasons.

Cells in many organisms may be unable to synthesize all of the amino acids required. As a result, they have to acquire them from their surroundings. *AAPs* are critical players in this process. *AAPs* may assist the transport of amino acids into a cell when extracellular amino acid levels exceed within the cell, thereby guaranteeing that the cell has a consistent supply of these important molecules [21].

2.5.1 Function as Integral Membrane Proteins

Since *AAPs* are integral membrane proteins, they are intimately entwined inside the cell's lipid bilayer. They do not merely live on the surface, but also pierce it, with protein sections projecting on both the interior and exterior sides of the membranes. This transmembrane configuration is not random. As the permease moves across the lipid environment, it creates complicated pathways. These channels play an important role in the passage of the amino acids throughout the cell membrane. These channels, which operate as selective gates, are precisely designed to allow just certain amino acids through. Notwithstanding the ever-changing external environment, this differentiation ensures that the interior cell environment remains regulated and in harmony. Cells can also adhere to one another, thanks to cell-to-cell adhesion, which is mediated by specific integral membrane proteins. In multi-cellular organisms where cell coordination is crucial, this is especially significant. Integral proteins give the cell membrane its structural integrity by anchoring it to the cell's internal or external structures. Integral membrane proteins are also found in complexes of proteins like the chain of electron transportation in mitochondria, which are essential for the respiration of cells and ATP synthesis [21].

2.5.2 Specificity of Amino Acid Permeases

Every protein has a role to play in the great orchestra of cellular activity, and *AAPs* are not an exemption. The substrate selectivity of such proteins is one of their most fascinating characteristics. *AAPs* have developed to recognize and attach to a single amino acid, facilitating its transport while blocking the transport of others. Because of this fine-tuned specificity, the cell can finely regulate the entry and outflow of specific amino acids based on its metabolic needs. Not all permeases, however, are restricted to one particular amino acid 'key'. Some evolved to be adaptable enough to be able to accommodate a variety of structurally connected amino acids. This group specificity frequently reflects structural similarities between different amino acids, enabling the *AAPs* to identify a shared

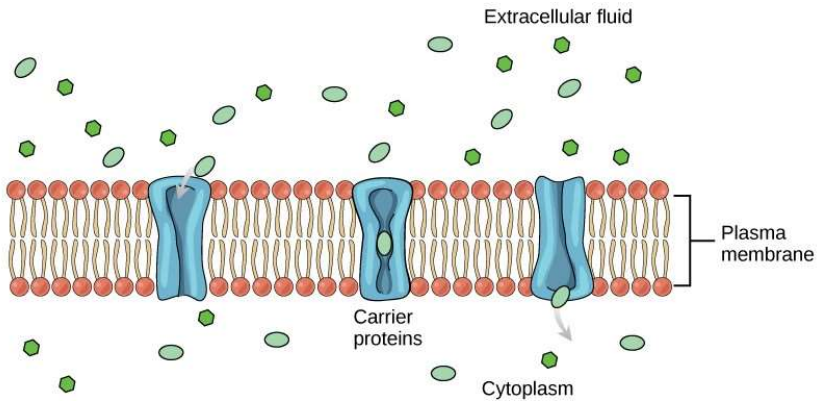


FIGURE 2.4: Selective permeability and specificity of phospholipid bilayer [23]

feature or motif. This adaptability guarantees that cells can effectively adapt to a variety of situations where a variety of amino acids may be accessible [22]. Figure 2.4 shows the selective permeability mechanism of *AAPs* within the phospholipid bilayer.

2.5.3 Transport Mechanism and Regulation

The interior of the cell is a center of activity, and the correct elements, particularly amino acids, need to be present for appropriate reactions to occur. As transporters, *AAPs* achieve this by utilizing specific processes for transporting amino acids over the membrane. Facilitated diffusion is a passive process in which amino acids travel spontaneously from an area where they are abundant to another where they are scarce. The *AAPs* function as a specialized door, increasing the movement's efficiency. This technique uses no energy; instead, the inherent capacity of molecules to expand out is used. Active transportation delivers amino acids from low-concentration locations to high-concentration areas, going against the natural flow. Cells must invest energy because this is energetically unfavorable. Cells accomplish this ingeniously by linking the transportation of peptides to the movement of ions, often protons (H^+). Both the amino acid as well as the proton are channeled in the same path in a symport process, whereas they travel in opposite directions in an antiport system. This deliberate linkage ensures that

the energy-intensive process is viable [21]. *AAPs* expression and activity are frequently altered according to the cellular surroundings:

- Nutrient Availability: The presence or lack of various amino acids in the surroundings can control the expression of associated permeases in many organisms. For example, if an amino acid is scarce, cells may stimulate the expression that encodes the permease essential for its uptake.
- Specific amino acid requirements for the synthesis of proteins or other metabolic activities may additionally affect permease expression and functionality.
- Post-translational Modifications: Post-translational modifications such as phosphorylation can influence the action of *AAPs* [22].

2.5.4 Genome-Wide Analysis of Amino Acid Permeases

Genome-wide studies are an important and interesting way of identifying and analyzing important gene families in plants or animals. Various such studies have been conducted in the past, that serve as a foundation for our theoretical understanding of these gene families, as well as pivotal building blocks for various genetic engineering techniques. *AAPs* have been studied in the past for their molecular characterization, in hopes of improving crop yield and nitrogen preservation. In a relevant study by [24], genome-wide identification of *AAPs* was conducted for molecular characterization of their transcriptional responses to nutrient stresses in *Brassica napus*. Three primary goals outlined in the study were to identify all of the *AAP* genes in *B. napus*, to characterize the genomic features and transcriptional responses of *AAP* gene members to nitrogen stresses, such as ammonium toxicity and nitrate limitation, and to look into the transcriptional responses of *AAPs* to other nutrient stresses, such as phosphate constraints, boron deficiency, cadmium toxicity, and salt stress. The molecular characterization and genome-wide identification of *AAP* members point to both functional divergence and evolutionary conservation between *Arabidopsis* and *B. napus*. The results were intended to offer a thorough understanding of amino acid importation and transport

Gene ID	Gene name	Block	CDS (bp)	Exon/ intron	Amino acid (aa)	Ka	Ks	Ka/Ks	Divergent time (Mya)
At1g58360	*AtAAP1*	D	1458	6/5	485				
BnaA01g21750D	*BnaA1.AAP1*	D	1524	6/5	507	0.0394	0.3931	0.1002	13.10
BnaA03g59400D	*BnaA3.AAP1*	D	1623	8/7	540	0.0565	0.3987	0.1417	13.29
BnaA09g14700D	*BnaA9.AAP1*	D	1440	6/5	479	0.0466	0.4043	0.1153	13.48
BnaC01g42990D	*BnaC1.AAP1*	D	1524	6/5	507	0.0405	0.3747	0.1081	12.49
BnaC04g18440D	*BnaC4.AAP1*	D	1455	6/5	484	0.0489	0.3854	0.1269	12.85
BnaCnng25620D	*BnaCn.AAP1*	D	1440	6/5	479	0.0489	0.3965	0.1233	13.22
At5g09220	*AtAAP2*	R	1482	6/5	493				
BnaA03g02650D	*BnaA3.AAP2*	R	1464	7/6	487	0.0293	0.4707	0.0622	15.69
BnaA10g22670D	*BnaA10.AAP2*	R	1458	6/5	485	0.0366	0.5027	0.0728	16.76
BnaC03g03750D	*BnaC3.AAP2*	R	1464	6/5	487	0.0313	0.4676	0.0669	15.59
BnaC09g47230D	*BnaC9.AAP2*	R	1458	6/5	485	0.0437	0.4787	0.0913	15.96
At1g77380	*AtAAP3*	E	1431	7/6	476				
BnaA07g33510D	*BnaA7.AAP3*	E	1431	7/6	476	0.0374	0.3945	0.0948	13.15
BnaC06g38080D	*BnaC6.AAP3a*	E	1431	7/6	476	0.0386	0.3865	0.0999	12.88
BnaC06g38090D	*BnaC6.AAP3b*	E	1431	7/6	476	0.0408	0.3847	0.1061	12.82
At5g63850	*AtAAP4*	X	1401	6/5	466				
BnaA02g33930D	*BnaA2.AAP4*	X	1401	6/5	466	0.0304	0.5479	0.0555	18.26
BnaA06g22970D	*BnaA6.AAP4*	X	1401	5/4	466	0.0392	0.4567	0.0858	15.22
BnaC02g42740D	*BnaC2.AAP4*	X	1401	5/4	466	0.0306	0.5398	0.0567	17.99
BnaC03g50500D	*BnaC3.AAP4*	X	1401	5/4	466	0.0423	0.4367	0.0969	14.56
At1g44100	*AtAAP5*	C	1443	5/4	480				
BnaAnng17090D	*BnaAn.AAP5*	C	1248	4/3	415	0.0804	0.4964	0.1620	16.55
BnaA08g04440D	*BnaA8.AAP5*	C	1446	5/4	481	0.0840	0.5449	0.1542	18.16
BnaA05g18660D	*BnaA5.AAP5*	C	1464	5/4	487	0.0801	0.5204	0.1539	17.35
BnaA10g08840D	*BnaA10.AAP5*	C	1431	5/4	476	0.1206	0.5015	0.2405	16.72
BnaC06g00580D	*BnaC6.AAP5*	C	1431	5/4	476	0.1149	0.4909	0.2341	16.36
At5g49630	*AtAAP6*	W	1446	6/5	481				
BnaCnng14480D	*BnaCn.AAP6*	W	1440	6/5	479	0.0560	0.4756	0.1177	15.85
At5g23810	*AtAAP7*	Q	1404	6/5	467				
BnaA09g05130D	*BnaA9.AAP7*	Q	1413	7/6	470	0.1092	0.3543	0.3082	11.81
BnaC09g04700D	*BnaC9.AAP7*	Q	1059	6/5	352	0.1169	0.3813	0.3066	12.71
At1g10010	*AtAAP8*	A	1428	6/5	475				
BnaA06g38000D	*BnaA6.AAP8a*	A	1401	6/5	466	0.1387	0.4977	0.2787	16.59
BnaA06g38010D	*BnaA6.AAP8b*	A	1389	6/5	462	0.1385	0.5880	0.2355	19.60
BnaA09g57230D	*BnaA9.AAP8a*	A	1446	6/5	481	0.1387	0.4977	0.2787	16.59
BnaA09g57240D	*BnaA9.AAP8b*	A	1584	6/5	527	0.1385	0.5880	0.2355	19.60
BnaC05g49200D	*BnaC5.AAP8a*	A	1410	7/6	469	0.1361	0.5426	0.2508	18.09
BnaC05g49210D	*BnaC5.AAP8b*	A	1449	6/5	482	0.1385	0.5880	0.2355	19.60
BnaC08g42410D	*BnaC8.AAP8a*	A	1446	6/5	481	0.1508	0.2216	0.6805	7.39
BnaC08g42420D	*BnaC8.AAP8b*	A	1515	7/6	504	0.1463	0.5326	0.2747	17.75
BnaC08g42430D	*BnaC8.AAP8c*	A	1446	6/5	481	0.1321	0.5257	0.2513	17.52

FIGURE 2.5: Molecular characterization of Amino Acid Permeases (*AAPs*) in *Brassica napus* and *Arabidopsis thaliana*. Source: Adapted from [24]

in *B. napus* amid various nutritional stressors. The molecular characterization of *AAPs* in *Brassica napus* and *Arabidopsis thaliana* is presented in Figure 2.5.

The study presents the role amino acids play in this process and the significance of nitrogen in seed yield and the levels of protein in plants. For the best possible growth and development of plant organs, nitrogen is essential. Plants take up inorganic nitrogen nutrients in the form of ammonium and nitrate. These nutrients

provide amino acids, which are essential for the growth and metabolism of the reproductive and vegetative organs. More than 100 potential amino acid transporter genes (*AATs*) are highlighted in the study, including the Usually Multiple Acids Move In and Out Transporters (*UMAMIT*) family, which is characterized in the model *Arabidopsis*, and the Amino acid–Polyamine–Choline (*APC*) transporter superfamily. Among them, *AAPs* are implicated in several physiological processes, including amino acid absorption, phloem loading, seed loading, and grain yield. They are presented to be a moderate-affinity system with broad substrate specificity. The reason for reduced nitrogen efficiency, according to this study, is that rapeseed senescent leaves separate before nitrogen nutrients and are completely remobilized to sink organs. The goal of the study was to improve *B. napus's* nitrogen remobilization efficiency by molecularly modifying amino acid transporters, specifically *AAPs* [24].

2.5.5 Evolutionary Consideration

The widespread presence of *AAPs* throughout all life domains emphasizes their vital evolutionary importance. Their prevalence from simplest microbes to sophisticated eukaryotes indicates an ancient origin and key function in coordinating amino acid transport, which is critical for cell development and survival. Evolution, on the other hand, is not a one-size-fits-all process. Although several permeases are conserved throughout species, emphasizing their critical roles, the particular types and amounts of these kinds of proteins in different creatures vary significantly. This variation is a result of individual evolutionary journeys in which animals or plants tailored their permease variety to their specific environmental difficulties and ecological niches. It is therefore understandable that all vascular and non-vascular plants have originated from a common ancestor, but have gone through extensive evolutionary change over generations to suit the needs of the environment, and to ensure their survival, growth, and development. Various former studies have compared the evolutionary divergence and presence of variation among the *AAPs* across different species. For instance, the study by [2], reveals the phylogenetic relationship between *NLP* gene family of different species of plants,

and confirms the ancestral lineage of *NLPs* among bryophytes and vascular plants to be common. In summary, while the fundamental role of *AAPs* stays constant, their evolutionary subtleties reflect the complex tapestry of life's flexibility [25].

Chapter 3

Methodology

The overview of methodology is shown in Figure 3.1

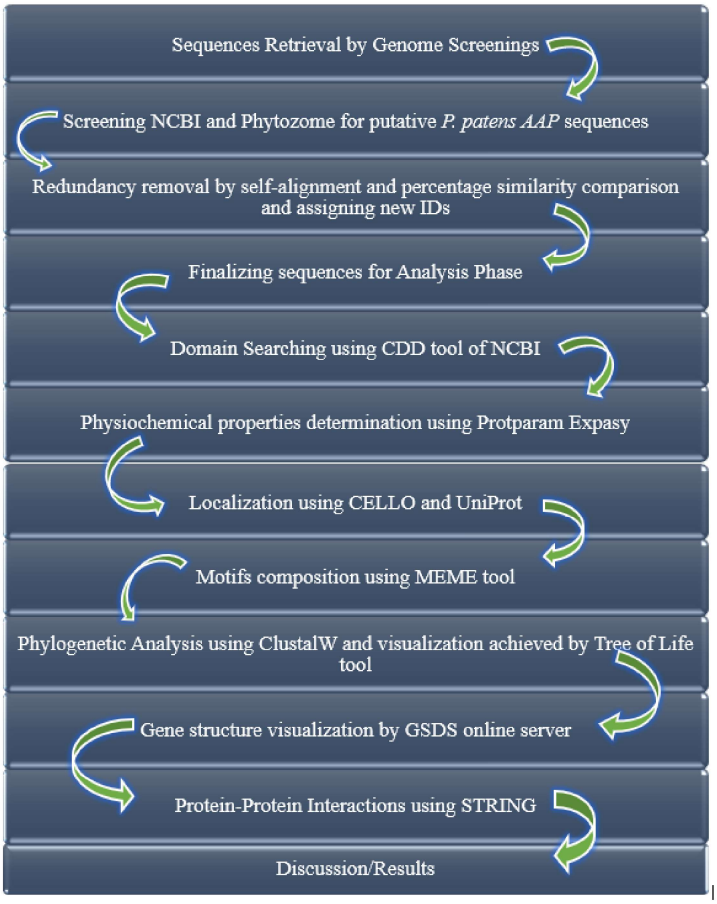


FIGURE 3.1: Overview of Methodology

3.1 Sequence Retrieval Phase

The sequence retrieval phase, which comprises screening for, removing redundancy, and finalizing samples was completed in three phases which are listed as follows:

3.1.1 Genome Screening

Since *Arabidopsis thaliana* was selected as a scale organism in this study, the full-length, protein, and coding sequences of *A. thaliana* amino acid permeases (*AtAAPs*) were downloaded from *Arabidopsis* genome database (TAIR: http://arabidopsis.org/). These sequences were retrieved in FASTA format and stored to be used in all sampling and analysis as scale. Two genome databases were screened for identification and characterization of any putative *Physcomitrella patens* amino acid permeases (*PpAAPs*). Firstly, the *AtAAP* protein sequences were used as BLAST-query in screening the NCBI database (https://www.ncbi.nlm.nih.gov/). Second, a database called Phytozome v13 (https://phytozome-next.jgi.doe.gov/) was screened using keywords such as "*AAPs*" and "Amino Acid Permeases" while having *P. patens* selected as the plant of interest [2]. Sequences from both of these screenings were downloaded, and their accession numbers were saved for future reference.

3.1.2 Removal of Redundant Sequences

All of the potential *PpAAP* sequences obtained, were aligned to remove any repetitive or redundant sequences [26]. This was achieved, firstly, by comparing accession numbers saved through BLAST-query and secondly, by self-aligning all sequences in the NCBI's BLAST tool (https://blast.ncbi.nlm.nih.gov/). The resultant sequences were arranged in an order for the next phase of sampling. All potential *PpAAP* sequences, whether from NCBI or Phytozome were rearranged and matched with their identical ones from the other database to ensure no sequence is being repeated and samples are accurate and specific.

3.1.3 Finalizing Sample Sequences

Once repetitive sequences had been removed, and the sequences had been matched from both of the databases according to their accession numbers, they were finalized for the analysis phase.

All of the 16 selected sequences were assigned new lab IDs (*PpAAP1-PpAAP2* and so on). The analysis results, for all *PpAAPs*, were maintained in an Excel sheet along with their respective accession numbers for future reference.

3.2 Analysis Phase

The finalized sequences were then subjected to analysis, which included several computational tools and softwares which were used to achieve genome-wide identification and analysis of Amino Acid Permeases in *P. patens*.

3.2.1 Domain Searching for Putative *Physcomitrella patens* Amino Acid Permeases (*PpAAPs*)

The scale sequences of *Arabidopsis thaliana* (*AtAAPs*) and the selected sample sequences of *P. patens* (*PpAAPs*) were uploaded for domain identification. This was achieved by using an online database of NCBI called Conserved Domain Database or CDD (https://www.ncbi.nlm.nih.gov/cdd).

The sequences were uploaded into Bulk CD-search, and the results described the common domain between the sequences indicating that the sequences are from the same family and allowed for the calculation of the quantitative position of gene, chromosome number, and gene length.

3.2.2 Physicochemical Properties & Localization of *PpAAPs*

The sequences selected based on conserved domains of (*PpAAPs*) were further analyzed for their physicochemical parameters such as Protein length (aa), Molecular weight (MW), Theoretical isoelectric point (pI) and Grand Average of Hydropathicity (GRAVY). This was achieved by an online tool called Protparam Expasy (https://web.expasy.org/protparam). It is an extendable and integrative portal that provides a catalog of over 160 software programs and database tools while additionally promoting a variety of biological science and medical research areas. Subcellular localization was predicted using another online tool called CELLO (http://cello.life.nctu.edu.tw) [2]. Both of these analyses were performed for scale sequences (*AtAAPs*) as well as sample sequences (*PpAAPs*), for comparison.

3.2.3 Motif Composition in *PpAAP* Gene Family

The presence of consensus motifs were determined using an online tool called Multiple Expectation Maximizations for Motif Elicitation or MEME v5.5.4 (meme-suite.org/meme/tools/meme). It is an online tool considered one of the most accurate for motif elicitation. All parameter settings were kept at default settings with the exception of the motif-finding threshold, which was kept at 20 to ensure specificity and precision [2].

3.2.4 Phylogenetic Analysis of *PpAAPs*

PpAAP sequences of non-vascular *P. patens* were aligned along with the *AAP* gene families of several other vascular plants. The selected organisms were *Arabidopsis thaliana* or thale cress [27], *Brassica napus* or canola [28], *Vicia faba* or broad bean [5], *Brassica rapa* or wild turnip [29], *Zea mays* or maize [30], *Glycine max* or soybean [31], *Raphanus sativus* or radish, *Brassica oleracea* or wild cabbage, *Cannabis sativa* or marijuana, *Eucalyptus grandis* or rose gum and

Cocus nucifera or coconut [32]. The *AAP* gene sequences for all plants were retrieved through a specific literature review and by using specific keywords in NCBI [33]. Multiple alignment was initially performed through the Clustal Omega tool (https://www.ebi.ac.uk/Tools/msa/clustalo/). The guide/Newick tree generated from the Clustal Omega tool was used as a query for visualization of the rooted phylogenetic tree in another online tool called Interactive Tree of Life v6 (https://itol.embl.de/) [2].

3.2.5 Protein-Protein Interactions in *PpAAP* Family

The cellular proteins in interaction with *PpAAPs* and *AtAAPs* were predicted using an online tool called STRING (https://string-db.org/), and their figures along with the description, were exported from the tool for comparative study and future reference. The interacting proteins of *PpAAPs* were compared with those of *AtAAPs* to identify any functional homology [34].

3.2.6 Gene Structure Determination in *PpAAP* Family

The full-length and coding sequences of *PpAAPs* and *AtAAPs* were retrieved and used to examine the structural components of these sequences using an online server called Gene Structure Display Server (GSDS) (http://gsds.gao-lab.org/). This tool assisted in determining exons, introns and untranslated regions (UTRs) present in the sequences [2]. The tool also helped in comparing gene lengths between the two families.

Chapter 4

Results

4.1 Genome Screening and Finalizing Samples

In the current study, two genome databases (NCBI: https://www.ncbi.nlm.nih.gov/ and Phytozome v.13: https://phytozome-next.jgi.doe.gov/.) were screened to identify putative *AAPs* in *P. patens* genome (Taxonomic ID: 3218) using the *AAP* protein sequences of *A. thaliana* retrieved from TAIR: https://www.arabidopsis.org/. The screening was done by using *AAPs* of *A. thaliana* as query sequences while selecting *P. patens* as the organism of interest. A total of 24 sequences were obtained from NCBI. The results were stored in an Excel sheet, along with their accession numbers and percentage similarities for future use. Next, the genome database Phytozome was screened using specific keywords "*AAPs*", and "Amino Acid Permeases" and through the accession numbers isolated from NCBI. A total of 23 sequences were extracted and stored alongside the previous sequences along with their accession numbers and percentage similarities. A total of 47 sequences, both from NCBI and Phytozome were stored in the first phase. The redundant, incomplete, or splice variant sequences were removed from the list and the accession numbers of sequences from both genome databases were matched to form a final list. The sequences were compared with each other and with *A. thaliana AAP* scale sequences to remove identical or highly similar sequences with the similarity threshold kept at 80% to ensure uniqueness. As a result, a final list of 16 putative

TABLE 4.1: List of finalized samples of *PpAAPs*

Name	Phytozome	NCBI	Full Name
PpAAP1	Pp3c24_6070V3.1	XP_024363601.1	amino acid permease 3-like [Physcomitrium patens]
PpAAP2	Pp3c13_3320V3.1	XP_024392330.1	amino acid permease 3-like isoform X1 [Physcomitrium patens]
PpAAP3	Pp3c14_9480V3.1	XP_024395019.1	lysine histidine transporter 1-like [Physcomitrium patens]
PpAAP4	Pp3c13_12390V3.1	XP_024393185.1	lysine histidine transporter-like 2 [Physcomitrium patens]
PpAAP5	Pp3c3_11320V3.1	XP_024372214.1	GABA transporter 1-like isoform X2 [Physcomitrium patens]
PpAAP6	Pp3c11_19940V3.1	XP_024388170.1	GABA transporter 1-like [Physcomitrium patens]
PpAAP7	Pp3c23_12700V3.1	XP_024361745.1	GABA transporter 1-like isoform X1 [Physcomitrium patens]
PpAAP8	Pp3c8_19000V3.1	XP_024381744.1	GABA transporter 1-like [Physcomitrium patens]
PpAAP9	Pp3c6_21750V3.1	XP_024379398.1	amino acid transporter AVT3B-like [Physcomitrium patens]
PpAAP10	Pp3c12_5490V3.1	XP_024390273.1	auxin transporter protein 1-like [Physcomitrium patens]
PpAAP11	Pp3c9_4450V3.1	XP_024384797.1	proline transporter 2-like [Physcomitrium patens]
PpAAP12	Pp3c6_1540V3.1	XP_024377316.1	lysine histidine transporter-like 8 [Physcomitrium patens]
PpAAP13	Pp3c21_14080V3.1	XP_024358675.1	LOW QUALITY PROTEIN: proline transporter 3-like [Physcomitrium patens]
PpAAP14	Pp3c9_20170V3.1	XP_024384627.1	amino acid transporter AVT1B-like [Physcomitrium patens]
PpAAP15	Pp3c10_6930V3.1	XP_024386910.1	amino acid transporter AVT1B-like [Physcomitrium patens]
PpAAP16	Pp3c15_890V3.1	XP_024396912.1	amino acid transporter ANT1-like [Physcomitrium patens]

PpAAPs was generated which are shown in Table 4.1 showing their accession numbers as present in Phytozome and NCBI respectively. The percentage similarity between *PpAAPs* is given in Table 4.2 while an overall percentage similarity index between *PpAAPs* and *AtAAPs* sequences is presented in Table 4.3(a).

TABLE 4.2: Percentage similarities between *PpAAPs* through self-alignment

Name	*PpAAP1*	*PpAAP2*	*PpAAP3*	*PpAAP4*	*PpAAP5*	*PpAAP6*	*PpAAP7*	*PpAAP8*	*PpAAP9*	*PpAAP10*	*PpAAP11*	*PpAAP12*	*PpAAP13*	*PpAAP14*	*PpAAP15*	*PpAAP16*
PpAAP1	100%	61.36%	29.59%	27.92%	29.04%	28.63%	28.14%	25.92%	21.56%	21.84%	22.85%	24.71%	23.14%	20.00%	18.73%	23.66%
PpAAP2	61.36%	100%	30.12%	29.76%	27.15%	28.64%	26.71%	26.75%	22.99%	22.78%	24.34%	23.86%	23.39%	19.95%	18.72%	25.62%
PpAAP3	29.59%	30.12%	100%	63.15%	26.55%	27.62%	28.08%	28.91%	27.14%	21.66%	24.51%	34.93%	25.17%	24.40%	19.90%	20.46%
PpAAP4	27.92%	29.76%	63.15%	100%	29.37%	27.42%	28.21%	30.34%	24.20%	28.14%	27.79%	38.33%	27.35%	21.77%	23.27%	23.34%
PpAAP5	29.04%	27.15%	26.55%	29.37%	100%	36.10%	63.43%	51.96%	24.56%	22.38%	27.43%	23.65%	25.00%	21.37%	22.94%	24.25%
PpAAP6	28.63%	28.64%	27.62%	27.42%	36.10%	100%	37.19%	31.35%	23.19%	25.20%	32.46%	25.18%	30.97%	20.91%	20.57%	22.52%
PpAAP7	28.14%	26.71%	28.08%	28.21%	63.43%	37.19%	100%	51.41%	20.93%	23.22%	28.57%	25.94%	27.08%	23.34%	25.24%	22.65%
PpAAP8	25.92%	26.75%	28.91%	30.34%	51.96%	31.35%	51.41%	100%	24.03%	25.40%	27.02%	26.55%	26.84%	22.22%	22.87%	25.43%
PpAAP9	21.56%	22.99%	27.14%	24.20%	24.56%	23.19%	20.93%	24.03%	100%	23.66%	23.68%	22.27%	0.00%	27.48%	26.14%	40.81%
PpAAP10	21.84%	22.78%	21.66%	28.14%	22.38%	25.20%	23.22%	25.40%	23.66%	100%	21.43%	21.75%	0.00%	29.55%	30.34%	29.82%
PpAAP11	22.85%	24.34%	24.51%	27.79%	27.43%	32.46%	28.57%	27.02%	23.68%	21.43%	100%	24.38%	59.85%	19.14%	19.02%	22.13%
PpAAP12	24.71%	24.28%	34.93%	38.33%	23.65%	25.18%	25.94%	26.55%	22.27%	21.75%	24.38%	100%	24.35%	20.86%	21.12%	20.30%
PpAAP13	23.14%	22.62%	25.17%	27.35%	25.00%	30.97%	27.08%	26.84%	0.00%	0.00%	59.85%	24.35%	100%	0.00%	0.00%	26.25%
PpAAP14	20.00%	19.95%	24.40%	21.77%	21.37%	20.91%	23.34%	22.48%	27.48%	29.55%	19.14%	20.86%	0.00%	100%	73.74%	25.00%
PpAAP15	18.73%	18.20%	20.75%	23.27%	22.94%	20.57%	25.24%	22.44%	26.14%	30.34%	19.02%	21.12%	0.00%	74.46%	100%	24.48%
PpAAP16	23.66%	25.62%	20.46%	23.34%	24.25%	22.52%	22.65%	25.43%	40.81%	0.00%	22.13%	20.30%	26.25%	25.00%	24.48%	100%
AtAAP1	100%	55.99%	57.23%	57.95%	55.51%	73.56%	48.71%	76.21%	57.58%	50.22%	29.59%	28.84%	27.59%			
AtAAP2	58.17%	100%	74.62%	89.20%	66.95%	58.61%	49.21%	55.99%	57.94%	55.71%	27.59%	25.88%	25.49%			
AtAAP3	57.23%	74.62%	100%	73.66%	71.07%	60.13%	51.62%	58.86%	61.51%	58.23%	28.04%	26.17%	27.15%			
AtAAP4	57.95%	89.20%	73.66%	100%	65.61%	58.52%	48.76%	55.56%	58.95%	54.76%	25.27%	25.33%	24.84%			
AtAAP5	55.51%	66.95%	71.07%	65.61%	100%	56.03%	48.12%	55.31%	60.65%	58.22%	25.96%	25.48%	26.95%			

TABLE 4.3(A): Percentage similarity between *AtAAPs* and *PpAAPs*

Name	***AtAAP1***	***AtAAP2***	***AtAAP3***	***AtAAP4***	***AtAAP5***	***AtAAP6***	***AtAAP7***	***AtAAP8***	***PpAAP1***	***PpAAP2***	***PpAAP3***	***PpAAP4***	***PpAAP5***
AtAAP1	100%	55.99%	57.23%	57.95%	55.51%	73.56%	48.71%	76.21%	57.58%	50.22%	29.59%	28.84%	27.59%
AtAAP2	58.17%	100%	74.62%	89.20%	66.95%	58.61%	49.21%	55.99%	57.94%	55.71%	27.59%	25.88%	25.49%
AtAAP3	57.23%	74.62%	100%	73.66%	71.07%	60.13%	51.62%	58.86%	61.51%	58.23%	28.04%	26.17%	27.15%
AtAAP4	57.95%	89.20%	73.66%	100%	65.61%	58.52%	48.76%	55.56%	58.95%	54.76%	25.27%	25.33%	24.84%
AtAAP5	55.51%	66.95%	71.07%	65.61%	100%	56.03%	48.12%	55.31%	60.65%	58.22%	25.96%	25.48%	26.95%
AtAAP6	73.56%	58.61%	60.13%	58.52%	56.03%	100%	48.88%	68.10%	59.91%	52.36%	29.68%	29.28%	26.48%
AtAAP7	48.71%	49.32%	51.01%	48.76%	48.12%	48.88%	100%	52.34%	49.41%	47.49%	26.72%	25.83%	27.91%
AtAAP8	76.21%	55.99%	58.86%	55.56%	55.31%	68.10%	52.34%	100%	59.65%	52.74%	28.60%	30.49%	26.60%
PpAAP1	57.58%	57.94%	61.51%	58.95%	60.65%	59.78%	48.87%	59.65%	100%	61.36%	29.59%	27.92%	29.04%
PpAAP2	50.22%	52.84%	58.23%	54.76%	58.22%	52.36%	47.49%	52.74%	61.36%	100%	30.12%	29.76%	27.15%
PpAAP3	29.59%	27.59%	26.29%	25.27%	25.96%	29.68%	26.02%	28.60%	29.59%	30.12%	100%	63.15%	26.55%
PpAAP4	28.84%	25.88%	25.26%	25.33%	25.48%	29.28%	25.90%	30.49%	27.92%	29.76%	63.15%	100%	29.37%
PpAAP5	27.59%	25.49%	26.85%	24.84%	26.95%	26.48%	28.25%	26.60%	29.04%	27.15%	26.55%	29.37%	100%
PpAAP6	27.03%	27.58%	28.63%	27.46%	28.26%	26.28%	26.86%	27.82%	28.63%	28.64%	27.62%	27.42%	36.10%
PpAAP7	26.38%	26.27%	28.60%	26.95%	25.82%	27.06%	27.61%	27.33%	28.14%	26.71%	28.08%	28.21%	63.43%
PpAAP8	25.35%	24.20%	25.11%	27.45%	26.79%	26.02%	24.83%	24.17%	25.92%	26.75%	28.91%	30.34%	51.96%
PpAAP9	22.25%	23.28%	21.71%	22.77%	21.07%	21.52%	21.96%	22.08%	21.56%	22.99%	27.14%	24.20%	24.56%
PpAAP10	19.68%	22.22%	21.84%	23.36%	21.96%	0.00%	0.00%	0.00%	21.84%	22.78%	21.66%	28.14%	22.38%
PpAAP11	25.83%	25.40%	25.32%	25.55%	23.85%	24.77%	23.74%	24.33%	22.85%	24.34%	24.51%	27.79%	27.43%
PpAAP12	24.95%	24.63%	23.72%	23.66%	21.98%	25.85%	22.45%	26.39%	24.71%	24.28%	34.93%	38.33%	23.65%
PpAAP13	22.11%	23.51%	24.42%	25.12%	22.79%	22.31%	23.97%	22.20%	23.14%	22.62%	25.17%	27.35%	25.00%
PpAAP14	21.11%	19.44%	21.46%	20.42%	20.80%	21.34%	19.80%	20.82%	20.00%	19.95%	24.40%	21.77%	21.37%
PpAAP15	21.70%	19.08%	20.99%	22.06%	19.82%	18.61%	18.87%	21.38%	18.73%	18.20%	20.75%	23.27%	22.94%
PpAAP16	22.60%	26.76%	23.86%	28.17%	24.51%	22.03%	20.73%	21.35%	23.66%	25.62%	20.46%	23.34%	24.25%

TABLE 4.3(B) : Percentage similarity between *AtAAPs* and *PpAAPs*.

	PpAAP6	***PpAAP7***	***PpAAP8***	***PpAAP9***	***PpAAP10***	***PpAAP11***	***PpAAP12***	***PpAAP13***	***PpAAP14***	***PpAAP15***	***PpAAP16***
AtAAP1	27.03%	26.38%	25.35%	22.25%	19.68%	25.83%	24.95%	22.11%	21.11%	21.70%	22.60%
AtAAP2	27.58%	26.64%	24.20%	23.28%	22.22%	25.40%	24.63%	23.51%	19.44%	20.05%	26.76%
AtAAP3	29.67%	28.57%	25.58%	21.71%	21.32%	25.32%	23.45%	25.00%	21.41%	20.96%	23.86%
AtAAP4	27.46%	26.95%	27.45%	22.77%	23.36%	25.55%	23.66%	25.12%	20.69%	22.06%	28.17%
AtAAP5	28.26%	25.82%	26.79%	21.07%	21.96%	23.85%	21.98%	22.79%	20.80%	19.82%	24.51%
AtAAP6	26.07%	27.06%	26.02%	21.52%	36.36%	24.77%	25.85%	23.12%	24.21%	26.56%	22.03%
AtAAP7	27.02%	28.39%	24.60%	21.96%	19.57%	23.74%	22.45%	23.97%	19.80%	18.29%	20.73%
AtAAP8	27.82%	27.33%	24.17%	22.08%	36.36%	24.33%	26.39%	22.20%	20.82%	21.74%	21.35%
PpAAP1	28.63%	28.14%	25.92%	21.56%	21.84%	22.85%	24.71%	23.14%	20.00%	18.73%	23.66%
PpAAP2	28.64%	26.71%	26.75%	22.99%	22.78%	24.34%	23.86%	23.39%	19.95%	18.72%	25.62%
PpAAP3	27.62%	28.08%	28.91%	27.14%	21.66%	24.51%	34.93%	25.17%	24.40%	19.90%	20.46%
PpAAP4	27.42%	28.21%	30.34%	24.20%	28.14%	27.79%	38.33%	27.35%	21.77%	23.27%	23.34%
PpAAP5	36.10%	63.43%	51.96%	24.56%	22.38%	27.43%	23.65%	25.00%	21.37%	22.94%	24.25%
PpAAP6	100%	37.19%	31.35%	23.19%	25.20%	32.46%	25.18%	30.97%	20.91%	20.57%	22.52%
PpAAP7	37.19%	100%	51.41%	20.93%	23.22%	28.57%	25.94%	27.08%	23.34%	25.24%	22.65%
PpAAP8	31.35%	51.41%	100%	24.03%	25.40%	27.02%	26.55%	26.84%	22.22%	22.87%	25.43%
PpAAP9	23.19%	20.93%	24.03%	100%	23.66%	23.68%	22.27%	0.00%	27.48%	26.14%	40.81%
PpAAP10	25.20%	23.22%	25.40%	23.66%	100%	21.43%	21.75%	0.00%	29.55%	30.34%	29.82%
PpAAP11	32.46%	28.57%	27.02%	23.68%	21.43%	100%	24.38%	59.85%	19.14%	19.02%	22.13%
PpAAP12	25.18%	25.94%	26.55%	22.27%	21.75%	24.38%	100%	24.35%	20.86%	21.12%	20.30%
PpAAP13	30.97%	27.08%	26.84%	0.00%	0.00%	59.85%	24.35%	100%	0.00%	0.00%	26.25%
PpAAP14	20.91%	23.34%	22.48%	27.48%	29.55%	19.14%	20.86%	0.00%	100%	73.74%	25.00%
PpAAP15	20.57%	25.24%	22.44%	26.14%	30.34%	19.02%	21.12%	0.00%	74.46%	100%	24.48%
PpAAP16	22.52%	22.65%	25.43%	40.81%	0.00%	22.13%	20.30%	26.25%	25.00%	24.48%	100%

4.2 Conserved Domain Identification

The common domain between the scale organism *A. thaliana* (Taxonomy ID: 3702) and the study organism *P. patens* (Taxonomy ID: 3218) for this particular gene family was found to be "Aa_trans" (PF01490) and "SLC5-6-like_sbd superfamily" (CL00456) [24].

The CDD tool of NCBI (https://www.ncbi.nlm.nih.gov/cdd) was used for this purpose, where the *AAP* protein sequences of *A. thaliana* in FASTA format were combined with the *AAP* protein sequences of *P. patens* and uploaded in bulk search.

The results show two common domains as shown in Table 4.4 confirming that the common domain of amino acid permeases as present in *A. thaliana* protein sequences, is present in *P. patens* sequences.

TABLE 4.4: Identification of Conserved Domain

Organism	Query	Hit type	PSSM-ID	From	To	E-Value	Bitscore	Accession	Short name
	AtAAP1	specific	279788	37	472	7.70E-132	387.432	pfam01490	Aa_trans
	AtAAP2	specific	279788	46	481	5.78E-127	375.491	pfam01490	Aa_trans
	AtAAP3	specific	279788	30	450	3.86E-119	354.69	pfam01490	Aa_trans
Arabidopsis thaliana (TAIR)	*AtAAP4*	specific	279788	19	454	4.47E-120	357.001	pfam01490	Aa_trans
	AtAAP5	specific	279788	28	468	1.28E-128	379.343	pfam01490	Aa_trans
	AtAAP6	specific	279788	33	470	2.25E-130	383.58	pfam01490	Aa_trans
	AtAAP7	superfamily	382020	26	462	1.11E-77	247.99	cl00456	SLC5-6-like_sbd superfamily
	AtAAP8	specific	279788	28	462	9.97E-121	358.927	pfam01490	Aa_trans
	PpAAP1	specific	279788	42	481	1.08E-113	341.593	pfam01490	Aa_trans
	PpAAP2	specific	279788	55	464	7.98E-98	300.762	pfam01490	Aa_trans
	PpAAP3	specific	279788	51	479	3.03E-90	281.117	pfam01490	Aa_trans
	PpAAP4	specific	279788	35	437	1.63E-91	283.043	pfam01490	Aa_trans
	PpAAP5	superfamily	382020	45	450	4.28E-61	204.462	cl00456	SLC5-6-like_sbd superfamily
	PpAAP6	superfamily	382020	39	437	2.05E-66	218.329	cl00456	SLC5-6-like_sbd superfamily
	PpAAP7	superfamily	382020	41	450	1.02E-64	214.092	cl00456	SLC5-6-like_sbd superfamily
Physcomitrella patens (Phytozome)	*PpAAP8*	superfamily	382020	41	451	3.92E-57	194.062	cl00456	SLC5-6-like_sbd superfamily
	PpAAP9	superfamily	382020	35	442	8.20E-70	226.804	cl00456	SLC5-6-like_sbd superfamily
	PpAAP10	superfamily	382020	1	467	0	697.63	cl00456	SLC5-6-like_sbd superfamily
	PpAAP11	superfamily	382020	57	457	1.31E-51	179.809	cl00456	SLC5-6-like_sbd superfamily
	PpAAP12	superfamily	382020	144	558	7.05E-50	177.498	cl00456	SLC5-6-like_sbd superfamily
	PpAAP13	superfamily	382020	5	318	8.57E-29	115.481	cl00456	SLC5-6-like_sbd superfamily
	PpAAP14	superfamily	382020	151	536	1.60E-58	200.225	cl00456	SLC5-6-like_sbd superfamily
	PpAAP15	superfamily	382020	154	535	2.31E-54	189.054	cl00456	SLC5-6-like_sbd superfamily
	PpAAP16	superfamily	382020	22	412	1.90E-72	232.967	cl00456	SLC5-6-like_sbd superfamily

4.3 Determination of Physicochemical Properties and Localization

Physicochemical properties and localization of *AtAAPs* and *PpAAPs* can be observed in Table 4.5(a) while the chromosomal distribution of *PpAAPs* is presented

in Table 4.6. The physicochemical properties of *PpAAPs* were determined using Protparam Expasy and were compared to that of *AtAAPs*. The average gene lengths of *PpAAPs* were found to be significantly higher than those of *AtAAPs* while the average protein lengths of the two were almost similar with the average of *PpAAPs* slighter higher than the latter.

Similar was the instance for GRAVY (Grand Average of Hydropathicity) values where *PpAAPs* were higher as compared to *AtAAPs*. However, the average molecular weight (MW) and Theoretical Iso-electric point (pI) of *AtAAPs* were found higher than those of *PpAAPs*. The average gene lengths of *AtAAPs* and *PpAAPs* were found 3157 and 4118.25 bp, respectively. A slight difference in the average protein length placed *AtAAPs* and *PpAAPs* at 477.875 and 475.6875, respectively. Likewise, the average molecular weight of *AtAAPs* and *PpAAPs* were found to be 52450.2375 and 52038.40063 Kilo Daltons (kDa), respectively. Similar was the instance for GRAVY (Grand Average of Hydropathicity) values where *PpAAPs* were higher as compared to *AtAAPs*. However, the average molecular weight (MW) and Theoretical Iso-electric point (pI) of *AtAAPs* were found higher than those of *PpAAPs*. The average gene lengths of *AtAAPs* and *PpAAPs* were found 3157 and 4118.25 bp, respectively. A slight difference in the average protein length placed *AtAAPs* and *PpAAPs* at 477.875 and 475.6875, respectively. Likewise, the average molecular weight of *AtAAPs* and *PpAAPs* were found to be 52450.2375 and 52038.40063 Kilo Daltons (kDa), respectively.

All of the *AtAAPs* and *PpAAPs* (except *PpAAP3*, *PpAAP14*, *PpAAP15* and *PpAAP16*) had pI values above 7 indicating them basic proteins with an average of *AtAAPs* and *PpAAPs* 8.91625 and 7.965625, respectively. The exception of proteins *PpAAP3*, *PpAAP14*, *PpAAP15* and *PpAAP16* were recorded to be 6.89, 5.54, 5.18, and 6.28, respectively, suggesting them as acidic proteins.

All *AAPs* from both plants showed positive GRAVY values, indicating them as hydrophobic proteins. In addition, the study of sub-cellular localization using CELLO and UniProt of both *AtAAPs* and *PpAAPs* showed them located in the plasma membrane.

TABLE 4.5(A): Physicochemical properties and Localization of *AtAAPs* and *PpAAAPs*

Gene	Gene ID	Chr.	Location	Gene Length (bp)	Protein Length (aa)	Organism
AtAAP1	AT1G58360.1	1	21676388 - 21680519	4132	485	*Arabidopsis thaliana* (TAIR)
AtAAP2	AT5G09220.1	5	2866222 - 2869156	2935	493	*Arabidopsis thaliana* (TAIR)
AtAAP3	AT1G77380.1	1	29074879 - 29077390	2512	476	*Arabidopsis thaliana* (TAIR)
AtAAP4	AT5G63850.1	5	25550937 - 25553656	2720	466	*Arabidopsis thaliana* (TAIR)
AtAAP5	AT1G44100.1	1	16764392 - 16767685	3294	480	*Arabidopsis thaliana* (TAIR)
AtAAP6	AT5G49630.1	5	20142430 - 20146690	4261	481	*Arabidopsis thaliana* (TAIR)
AtAAP7	AT5G23810.1	5	8028238 - 8030888	2651	467	*Arabidopsis thaliana* (TAIR)
AtAAP8	AT1G10010.1	1	3265976 - 3268726	2751	475	*Arabidopsis thaliana* (TAIR)
PpAAP1	Pp3c24_6070V3.1	24	4129527 - 4135911	6385	491	*Physcomitrella patens* (Phytozome)
PpAAP2	Pp3c13_3320V3.1	13	1907780 - 1912743	4964	490	*Physcomitrella patens* (Phytozome)
PpAAP3	Pp3c14_9480V3.1	14	6102358 - 6105692	3335	494	*Physcomitrella patens* (Phytozome)
PpAAP4	Pp3c13_12390V3.1	13	9192066 - 9195223	3158	453	*Physcomitrella patens* (Phytozome)
PpAAP5	Pp3c3_11320V3.1	3	8012546 - 8016894	4349	462	*Physcomitrella patens* (Phytozome)
PpAAP6	Pp3c11_19940V3.1	11	13163762 - 13166889	3128	452	*Physcomitrella patens* (Phytozome)
PpAAP7	Pp3c23_12700V3.1	23	8553314 - 8557143	3830	461	*Physcomitrella patens* (Phytozome)
PpAAP8	Pp3c8_19000V3.1	8	12576936 - 12580541	3606	462	*Physcomitrella patens* (Phytozome)
PpAAP9	Pp3c6_21750V3.1	6	13894096 - 13897481	3386	447	*Physcomitrella patens* (Phytozome)
PpAAP10	Pp3c12_5490V3.1	12	3858122 - 3861149	3028	467	*Physcomitrella patens* (Phytozome)
PpAAP11	Pp3c9_4450V3.1	9	2562076 - 2566190	4115	468	*Physcomitrella patens* (Phytozome)
PpAAP12	Pp3c6_1540V3.1	6	808186 - 812946	4761	570	*Physcomitrella patens* (Phytozome)
PpAAP13	Pp3c21_14080V3.1	21	8981031 - 8982702	1672	370	*Physcomitrella patens* (Phytozome)
PpAAP14	Pp3c9_20170V3.1	9	13642115 - 13650232	8118	550	*Physcomitrella patens* (Phytozome)
PpAAP15	Pp3c10_6930V3.1	10	4814746 - 4819990	5245	554	*Physcomitrella patens* (Phytozome)
PpAAP16	Pp3c15_890V3.1	15	500557 - 503368	2812	420	*Physcomitrella patens* (Phytozome)

TABLE 4.5(B): Physicochemical properties and Localization of *AtAAPs* and *PpAAAPs* (II)

	Gene	MW (KDa)	pI	GRAVY	Localization
Arabidopsis thaliana (TAIR)	*AtAAP1*	52894.90	8.96	0.422	Plasma Membrane
	AtAAP2	54146.90	9.16	0.437	Plasma Membrane
	AtAAP3	52036.60	9.06	0.506	Plasma Membrane
	AtAAP4	51428.10	9.29	0.491	Plasma Membrane
	AtAAP5	52537.90	8.38	0.485	Plasma Membrane
	AtAAP6	53020.30	8.62	0.372	Plasma Membrane
	AtAAP7	51722.30	8.7	0.503	Plasma Membrane
	AtAAP8	51814.90	9.16	0.477	Plasma Membrane
Physcomitrella patens (Phytozome)	*PpAAP1*	54352.11	8.51	0.326	Plasma Membrane
	PpAAP2	54161.35	8.38	0.199	Plasma Membrane
	PpAAP3	55087.50	6.89	0.485	Plasma Membrane
	PpAAP4	50277.23	9.12	0.568	Plasma Membrane
	PpAAP5	49864.80	9.14	0.647	Plasma Membrane
	PpAAP6	49292.30	8.06	0.485	Plasma Membrane
	PpAAP7	50203.47	7.05	0.565	Plasma Membrane
	PpAAP8	50407.64	8.84	0.726	Plasma Membrane
	PpAAP9	48714.41	8.29	0.663	Plasma Membrane
	PpAAP10	50589.75	9.06	0.775	Plasma Membrane
	PpAAP11	51399.94	8.74	0.522	Plasma Membrane
	PpAAP12	62730.44	9.18	0.406	Plasma Membrane
	PpAAP13	40494.55	9.19	0.644	Plasma Membrane
	PpAAP14	59768.19	5.54	0.395	Plasma Membrane
	PpAAP15	60268.53	5.18	0.369	Plasma Membrane
	PpAAP16	45002.20	6.28	0.842	Plasma Membrane

TABLE 4.6: Chromosomal distribution of *PpAAPs*

Gene Locus	Chromosome	Position	Gene Description
Pp3c24_6070V3.1	24	4129527 - 4135911	Amino Acid Permease 1 (*PpAAP1*)
Pp3c13_3320V3.1	13	1907780 - 1912743	Amino Acid Permease 2 (*PpAAP2*)
Pp3c14_9480V3.1	14	6102358 - 6105692	Amino Acid Permease 3 (*PpAAP3*)
Pp3c13_12390V3.1	13	9192066 - 9195223	Amino Acid Permease 4 (*PpAAP4*)
Pp3c3_11320V3.1	3	8012546 - 8016894	Amino Acid Permease 5 (*PpAAP5*)
Pp3c11_19940V3.1	11	13163762 - 13166889	Amino Acid Permease 6 (*PpAAP6*)
Pp3c23_12700V3.1	23	8553314 - 8557143	Amino Acid Permease 7 (*PpAAP7*)
Pp3c8_19000V3.1	8	12576936 - 12580541	Amino Acid Permease 8 (*PpAAP8*)
Pp3c6_21750V3.1	6	13894096 - 13897481	Amino Acid Permease 9 (*PpAAP9*)
Pp3c12_5490V3.1	12	3858122 - 3861149	Amino Acid Permease 10 (*PpAAP10*)
Pp3c9_4450V3.1	9	2562076 - 2566190	Amino Acid Permease 11 (*PpAAP11*)
Pp3c6_1540V3.1	6	808186 - 812946	Amino Acid Permease 12 (*PpAAP12*)
Pp3c21_14080V3.1	21	8981031 - 8982702	Amino Acid Permease 13 (*PpAAP13*)
Pp3c9_20170V3.1	9	13642115 - 13650232	Amino Acid Permease 14 (*PpAAP14*)
Pp3c10_6930V3.1	10	4814746 - 4819990	Amino Acid Permease 15 (*PpAAP15*)
Pp3c15_890V3.1	15	500557 - 503368	Amino Acid Permease 16 (*PpAAP16*)

4.4 Identification of Consensus Motifs

Composition of motif regions was achieved using MEME v5.5.4, by uploading the protein sequences *PpAAPs* in a FASTA file. A total of 20 consensus motifs were figured out in *PpAAP* proteins in self-comparison, since as the literature suggests, the higher similarity between motif regions points to functional homology itself [35]. Figure 4.1 shows the presence of conserved motif regions with 8 motifs (or 40%) present in all *PpAAPs*. 8 motifs (or 40%) were present in the majority of

PpAAPs and 4 (or 20%) were rare motifs, present in only some *PpAAPs*. The least number of motifs was observed in *PpAAP10*, which was 6, followed by *PpAAP9* and *PpAAP16*, which were recorded to have 8 motif regions each. *PpAAP11* and *PpAAP13* had 9 motif regions each. No protein had all 20 motifs present in them, indicating their uniqueness and specificity while the presence of the majority of motifs in all *PpAAPs* indicates that they indeed belong to the same family and are clued to be similar in function. Table 4.7(a) onwards shows all of the 20 motifs and their sequences, present in the *PpAAPs*, as identified by the MEME tool.

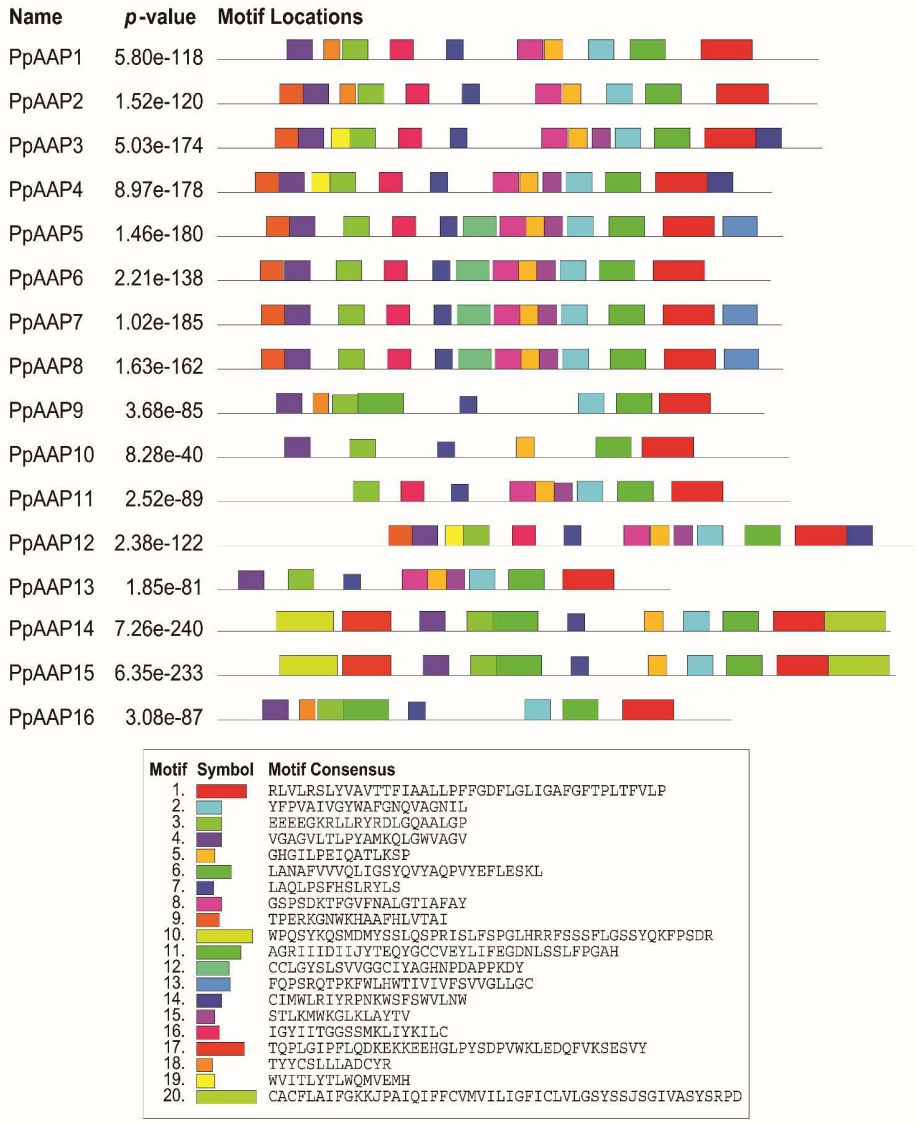


FIGURE 4.1: Identification of Conserved Motifs in *PpAAPs*.

TABLE 4.7(A): List of Consensus Motifs along with their sequences and logos in *PpAAPs*

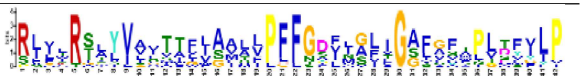
Motif	Sequence	Logo
Motif 1	RLVLRSLYVAVTTFIAALLPFFGDFLGLIGAFGFTPLTFVLP	
Motif 2	YFPVAIVGYWAFGNQVAGNIL	
Motif 3	EEEEGKRLLRYRDLGQAALGP	
Motif 4	VGAGVLTLPYAMKQLGWVAGV	
Motif 5	GHGILPEIQATLKSP	
Motif 6	LANAFVVVQLIGSYQVYAQPVYEFLESKL	
Motif 7	LAQLPSFHSLRYLS	
Motif 8	GSPSDKTFGVFNALGTIAFAY	
Motif 9	TPERKGNWKHAAFHLVTAI	
Motif 10	WPQSYKQSMDMYSSLQSPRISLFSPGLHRRFSSSFLGSSYQKFPSDR	

TABLE 4.7(B): List of Consensus Motifs along with their sequences and logos in *PpAAPs* (II)

Motif	Sequence	Logo
Motif 11	AGRIIIDIIJYTEQYGCCVEYLIFEGDNLSSLFPGAH	
Motif 12	CCLGYSLSVVGGCIYAGHNPDAPPKDY	
Motif 13	FQPSRQTPKFWLHWTIVIVFSVVGLLGC	
Motif 14	CIMWLRIYRPNKWSFSWVLNW	
Motif 15	STLKMWKGLKLAYTV	
Motif 16	IGYIITGGSSMKLIYKILC	
Motif 17	TQPLGIPFLQDKEKKEEHGLPYSDPVWKLEDQFVKSESVY	
Motif 18	TYYCSLLLADCYR	
Motif 19	WVITLYTLWQMVEMH	
Motif 20	CACFLAIFGKKJPAIQIFFCVMVILIGFICLVLGSYSSJSGIVASYSRPD	

4.5 Sequence Alignment and Phylogenetic Analysis

The *PpAAPs* and the *AtAAPs* were compared against each other, prior to analysis, to confirm the appropriate selection and singularity of every *PpAAP* gene for its uniqueness. For this reason, the similarity threshold was kept at 80% and ensured that all *PpAAPs* and *AtAAPs* shared less than 80% of similarity in their protein sequences. A similar process was also done across *PpAAPs* through self-alignment and removal of redundant, repeat, or splice variants. This allowed for evolutionary diversity to be identified among all involved members of the *PpAAP* gene family.

The resultant *PpAAP* gene family of this alignment used for analysis in this study, was compared with the *AAP* gene families of *Arabidopsis thaliana* (*AtAAPs*), *Brassica napus* (*BnAAPs*), *Vicia faba* (*VfAAPs*), *Brassica rapa* (*BrAAPs*), *Zea mays* (*ZmAAPs*), *Glycine max* (*GmAAPs*), *Raphanus sativus* (*RsAAPs*), *Brassica oleracea* (*BoAAPs*), *Cannabis sativa* (*CsAAPs*), *Eucalyptus grandis* (*EgAAPs*) and *Cocos nucifera* (*CnAAPs*). These *AAP* gene families, along with *PpAAPs*, were used as query to perform multiple alignment using the Clustal Omega tool. The guide tree generated from the Clustal Omega tool was then used as a query for visualization of rooted phylogenetic tree in another online tool called Interactive Tree of Life v6. Figure 4.2 shows the phylogenetic relationship between *P. patens* and these plants. The phylogenetic tree shows the evolutionary relationship between these plants in 6 clades with several further divisions. The *AAP* gene family of non-vascular *P. patens* show evolutionary divergence from the other 11 vascular plants. Variation also exists between *PpAAPs* as *PpAAP1* and *PpAAP2* reside far away from the other members. This variation may also point to structural and functional diversity. Furthermore, it was observed that the majority of *PpAAP* members shared similarity and were conserved with each other, with the exception of *PpAAP1* and *PpAAP2*, and may also share structural and functional similarity. Closest neighbors of *PpAAP1* and *PpAAP2* were observed to be *CnAAP2*, *CnAAP8*, *VfAAP1*, *VfAAP3*, *GmAAP2*, *GmAAP4*, *EgAAP4* and *ZmAAP8* while, in the clade where

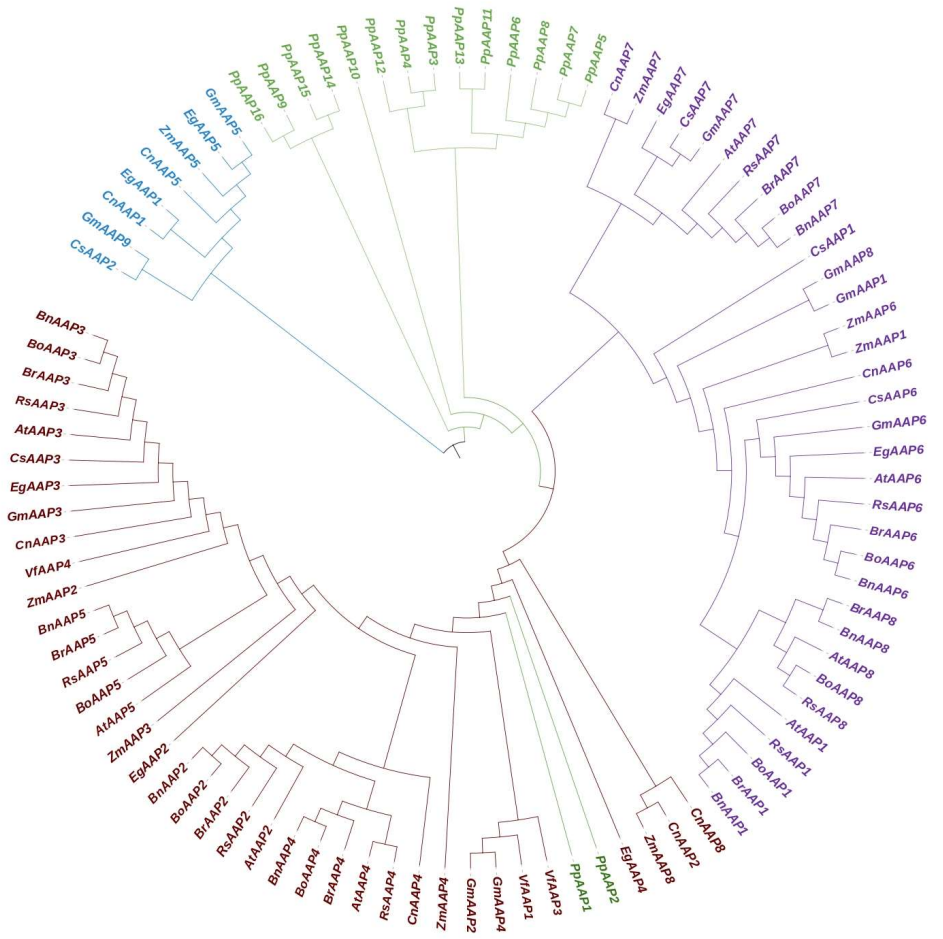


FIGURE 4.2: Phylogenetic Analysis of *PpAAPs* with different economically important plant species

majority of *PpAAPs* (*PpAAP3*-*PpAAP16*) lie, the closest members were observed to be *CnAAP1*, *CnAAP5*, *CnAAP7*, *EgAAP1*, *EgAAP5*, *EgAAP7*, *GmAAP5*, *GmAAP7*, *GmAAP9*, *ZmAAP5*, *ZmAAP7*, *CsAAP2*, *CsAAP7*, *AtAAP7*, *RsAAP7*, *BrAAP7*, *BoAAP7* and *BnAAP7*. It was also noted that *PpAAP1* and *PpAAP2* could prove to be quite economically important since these proteins may share structural and functional attributes with economically important plants such as *Glycine max* and *Zea mays*. This distribution of the *AAP* gene family represented the presence of an imperative evolutionary relationship between the *AAP* family across species and established significant evolutionary divergence between non-vascular bryophytes and vascular tracheophytes.

4.6 Gene Structure Determination

The gene and coding sequences of *AtAAPs* and *PpAAPs* were used to analyze their structural features which included the identification of exons, introns, and untranslated regions (UTRs). This was done using an online server called GSDS or Gene Structure Display Server (http://gsds.gao-lab.org/). Figure 4.3 displays the result of gene structure determination.

It was observed that the lowest number of exons present was in *PpAAP9*, which was 1, followed by the presence of 2 exons in *PpAAP16*. The number of exons for the rest of the *PpAAPs* range from 5 to 11 with *PpAAP12* and *PpAAP13* containing 5 exons each, *PpAAP4*, *PpAAP5* and *PpAAP6* having 6 exons, *PpAAP1*, *PpAAP3*, *PpAAP7*, *PpAAP8* and *PpAAP11* had 7 exons each, while 8 exons were observed in *PpAAP2* and *PpAAP10*. The highest number of exons was observed in *PpAAP14* and *PpAAP15*, which were 10 and 11, respectively.

The range of exons in *AtAAPs* was observed to be 5-7, with *AtAAP5* having 5 exons, while *AtAAP1*, *AtAAP2*, *AtAAP4*, *AtAAP6* and *AtAAP8* were observed to have 6 exons each. The highest number of exons were present in *AtAAP3* and *AtAAP7* which was 7.

4.7 Protein-Protein Interactions

Tables 4.8(a) onward show the interacting proteins of *PpAAPs* along with their nodes and annotations. Protein-protein interactions were studied using STRING (https://string-db.org/) and the *AAP* interacting protein network was predicted using it.

The majority of *PpAAPs* were interacting with proteins having Aa_trans domain which meant that they were involved in amino acid transportation and other related roles since interacting proteins are predicted to be involved in similar functions [36].

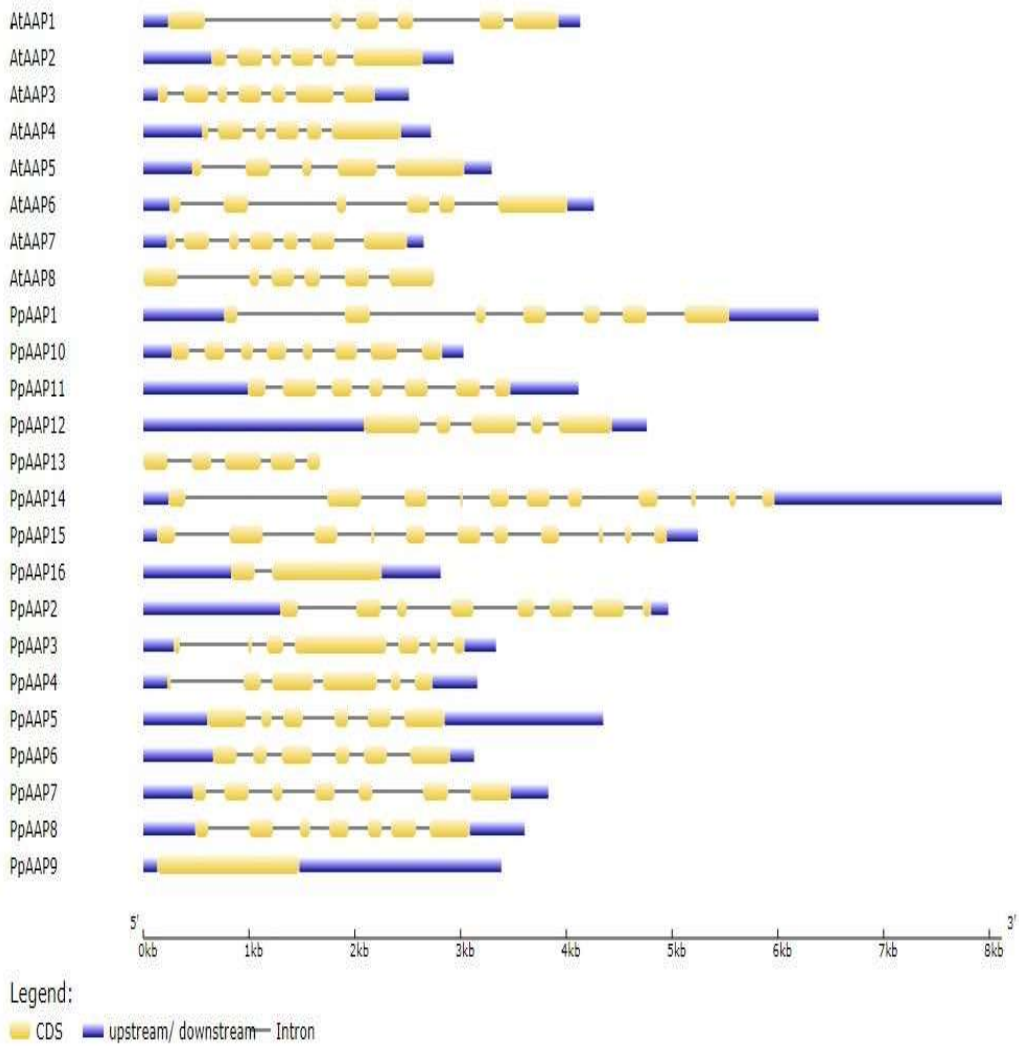


FIGURE 4.3: Gene Structure Determination of *PpAAPs* and *AtAAPs*

The majority of *PpAAPs* were interacting with proteins having Aa_trans domain which meant that they were involved in amino acid transportation and other related roles since interacting proteins are predicted to be involved in similar functions [36].

These interactions demonstrate and confirm the role of *PpAAPs* in amino acids and other biomolecule transportation in and out of the cell since both *PpAAPs* and their interacting proteins contain the conserved domain however, significant variation was observed between the interacting proteins of *PpAAPs* in totality.

Many interacting proteins were predicted or uncharacterized proteins present in

many *PpAAPs*. Apart from Aa_trans-domain containing proteins, *PpAAP1* also showed interactions with *chk3*, *CRE2*, *CRE1* which were CHASE domain-containing proteins and cytokinin receptor protein, respectively. *PpAAP2*, *PpAAP3*, *PpAAP4*, *PpAAP7*, *PpAAP10* and *PpAAP11* contained several Aa_trans-domain containing proteins and predicted or uncharacterized proteins.

Interacting proteins of *PpAAP5* were observed to be *PlsC* domain-containing proteins, various uncharacterized proteins, ATPase proteins and *glutamate 5-kinase P5CS* proteins. *PpAAP6* had two interacting proteins, both of which were *PlsC* domain-containing proteins. Similar to *PpAAP5*, *PpAAP8* interacting proteins showed variation, with the addition of *IGPS* domain-containing and aminotransferase proteins. *PpAAP9* showed interactions with *PRK* domain-containing protein, ATPase alpha/beta chains family proteins, *PCRF* domain-containing protein, and *RF_PROK_I* domain-containing protein.

PpAAP9 showed interactions with *PRK* domain-containing protein, ATPase alpha/beta chains family proteins, *PCRF* domain-containing protein, and *RF_PROK_I* domain-containing protein.

Most variation was observed in *PpAAP12*, *PpAAP14* and *PpAAP15* with interacting proteins such as Amine oxidase, *F-box* domain-containing proteins, *ZnMc* domain-containing protein, glycosyltransferase family proteins, L-aspartate oxidase, V-type proton ATPase subunit, *2-Hacid_dh* domain-containing protein, Formate dehydrogenase and Vacuolar proton pump subunit B.

The lowest interactions were observed in *PpAAP7* and *PpAAP16* with just one interacting protein, followed by *PpAAP6* having two interacting proteins. *PpAAPs* having higher interactions were observed at lower stringency settings, therefore many interacting proteins were observed to be quite functionally different and diverse. Still, the majority of *PpAAP* interactions were found to be quite relevant and parallel with earlier analysis results.

Still, the majority of *PpAAP* interactions were found to be quite relevant and parallel with earlier analysis results.

TABLE 4.8(A): Protein-Protein interactions of *PpAAP1*

***PpAAP1* (PHYPA_028700)**	
Node	**Annotation**
PHYPA_009556	Aa_trans domain-containing protein.
PHYPA_019123	Aa_trans domain-containing protein.
PHYPA_012079	Aa_trans domain-containing protein.
chk3	CHASE-domain containing histidine kinase 3.
CRE2	CHASE-domain containing histidine kinase 2.
CRE1	Cytokinin receptor 1.
PHYPA_012224	Aa_trans domain-containing protein.
PHYPA_002927	Uncharacterized protein.
PHYPA_007291	Uncharacterized protein.
PHYPA_000447	Uncharacterized protein.

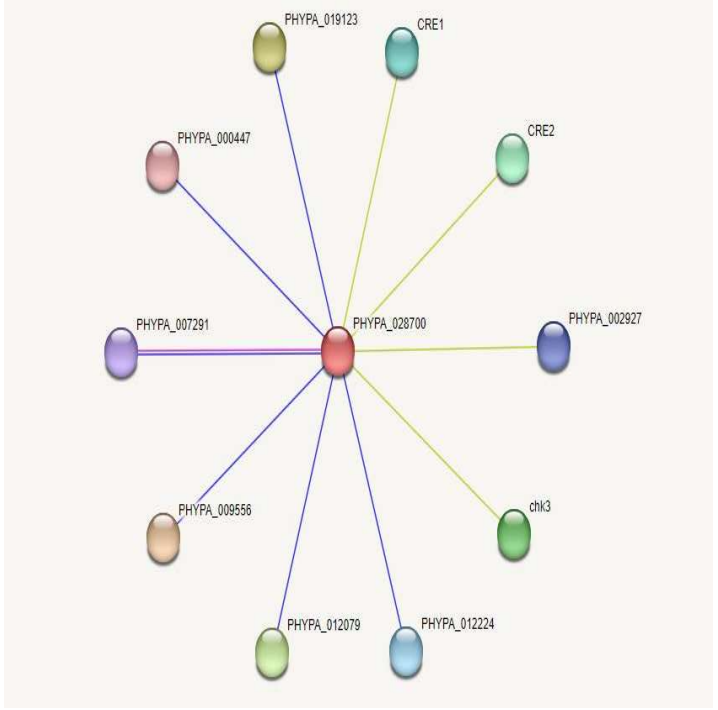


TABLE 4.8(B): Protein-Protein interactions of *PpAAP2*

***PpAAP2* (PHYPA_016919)**	
Node	**Annotation**
PHYPA_012079	Aa_trans domain-containing protein.
PHYPA_019123	Aa_trans domain-containing protein.
PHYPA_012224	Aa_trans domain-containing protein.
PHYPA_009556	Aa_trans domain-containing protein.
PHYPA_015856	Predicted protein.
PHYPA_004194	Aa_trans domain-containing protein.

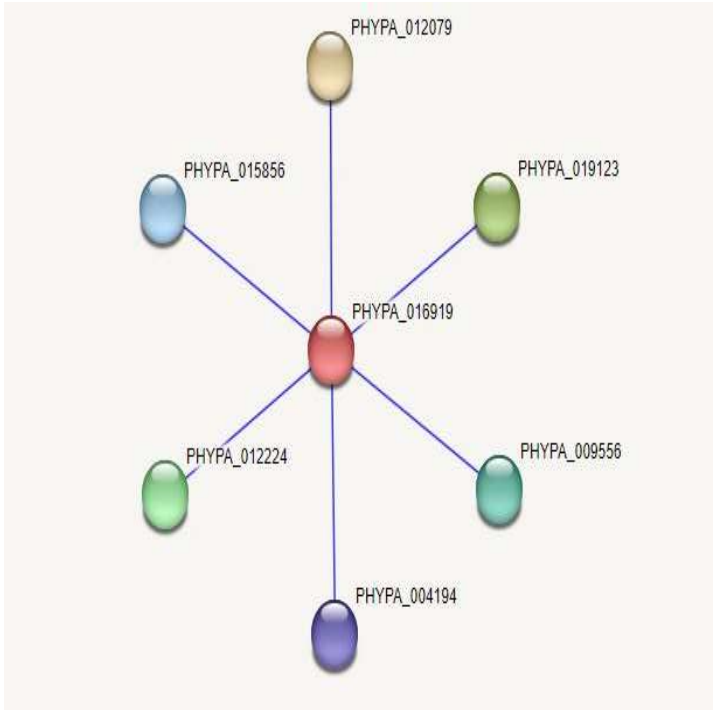


TABLE 4.8(C): Protein-Protein interactions of *PpAAP3*

***PpAAP3* (PHYPA_018257)**	
Node	**Annotation**
PHYPA_012079	Aa_trans domain-containing protein.
PHYPA_012224	Aa_trans domain-containing protein.
PHYPA_015856	Predicted protein.
pplax5	Aa_trans domain-containing protein.
PHYPA_009556	Aa_trans domain-containing protein.
PHYPA_004194	Aa_trans domain-containing protein.
PHYPA_026160	Uncharacterized protein.
PHYPA_008026	Predicted protein.
PHYPA_020722	Predicted protein.

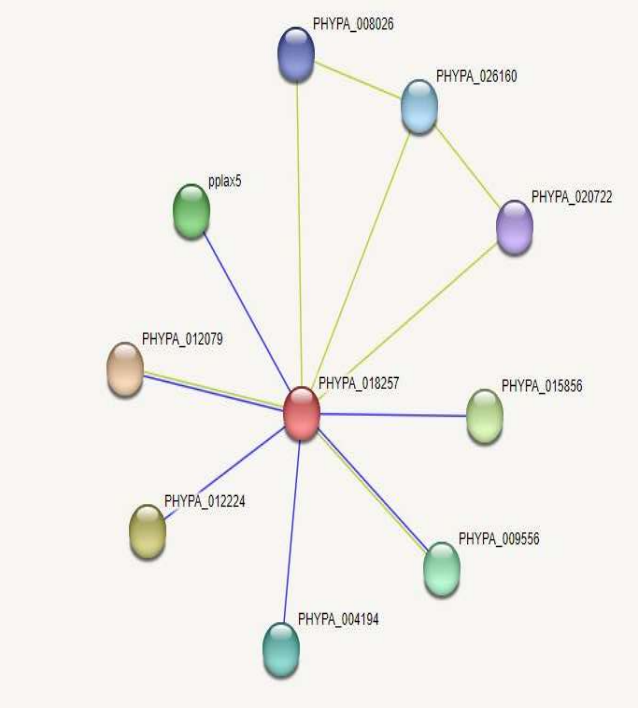


TABLE 4.8(D): Protein-Protein interactions of *PpAAP4*

***PpAAP4* (PHYPA_017275)**	
Node	**Annotation**
PHYPA_012079	Aa_trans domain-containing protein.
PHYPA_019123	Aa_trans domain-containing protein.
PHYPA_015856	Predicted protein.
PHYPA_009556	Aa_trans domain-containing protein.
pplax5	Aa_trans domain-containing protein.
PHYPA_004194	Aa_trans domain-containing protein.
PHYPA_026160	Uncharacterized protein.
PHYPA_008026	Predicted protein.
PHYPA_020722	Predicted protein.

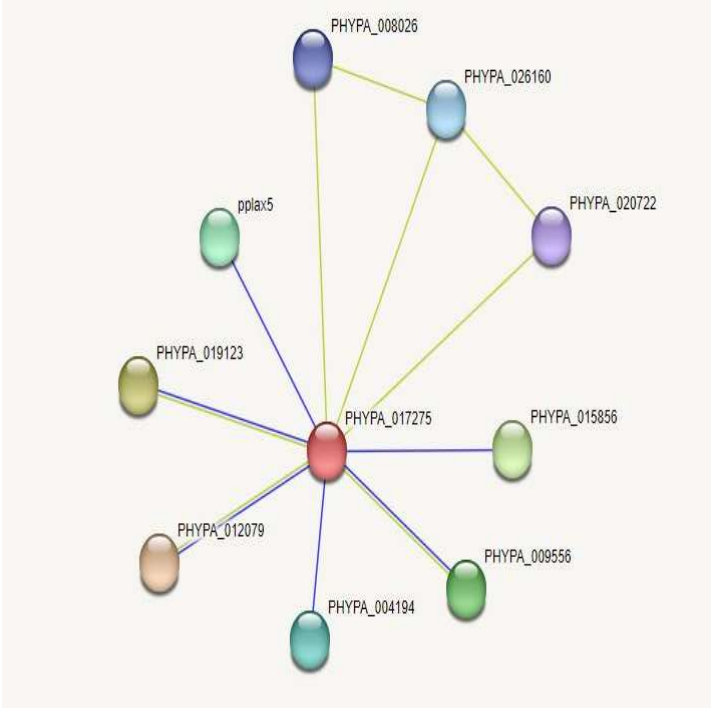


TABLE 4.8(E): Protein-Protein interactions of *PpAAP5*

***PpAAP5* (PHYPA_004284)**	
Node	**Annotation**
PHYPA_003424	Delta-1-pyrroline-5-carboxylate synthase; P5CS plays a key role in proline biosynthesis, leading to osmoregulation in plants; In the N-terminal section; belongs to the glutamate 5- kinase family.
PHYPA_000447	Uncharacterized protein.
PHYPA_011198	Uncharacterized protein.
PHYPA_025321	V-type proton ATPase subunit; Subunit of the integral membrane V0 complex of vacuolar ATPase. Vacuolar ATPase is responsible for acidifying a variety of intracellular compartments in eukaryotic cells, thus providing most of the energy required for transport processes in the vacuolar system. Belongs to the V-ATPase V0D/AC39 subunit family.
PHYPA_010630	Uncharacterized protein.
PHYPA_022765	Vacuolar proton pump subunit B; Non-catalytic subunit of the peripheral V1 complex of vacuolar ATPase; Belongs to the ATPase alpha/beta chains family.
PHYPA_008865	ATP-synt_ab domain-containing protein.
PHYPA_009121	PlsC domain-containing protein.
PHYPA_007152	PlsC domain-containing protein.
PHYPA_020835	Uncharacterized protein.

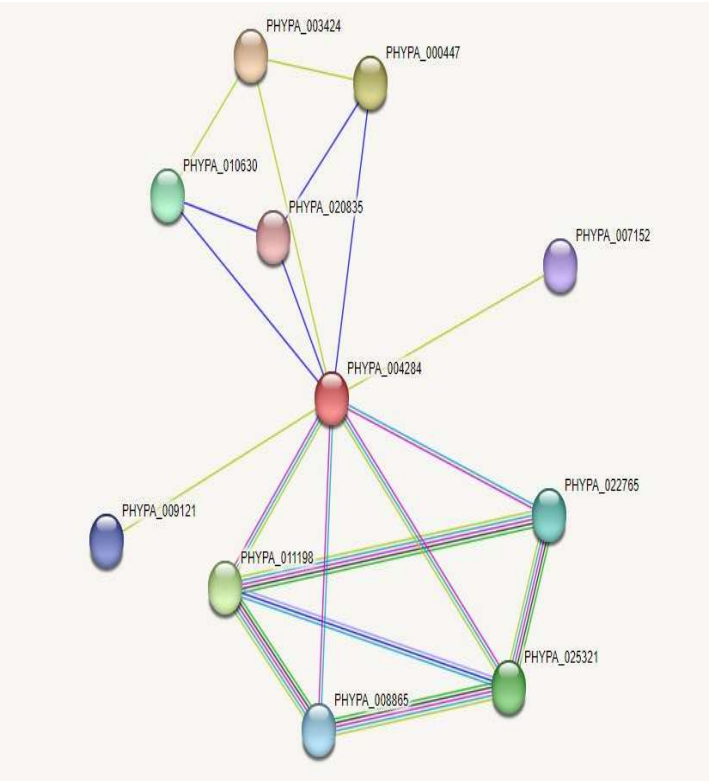


TABLE 4.8(F): Protein-Protein Interactions of *PpAAP6*

***PpAAP6* (PHYPA_015274)**	
Node	**Annotation**
PHYPA_009121	PlsC domain-containing protein.
PHYPA_007152	PlsC domain-containing protein.

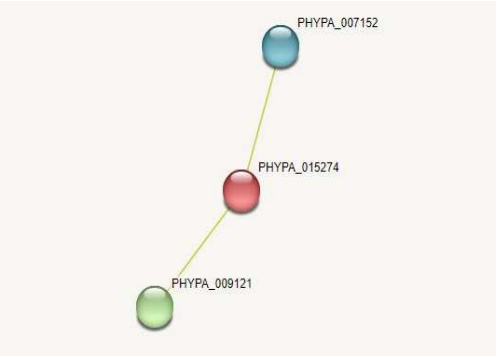


TABLE 4.8(G): Protein-Protein Interactions of *PpAAP7*

***PpAAP7* (PHYPA_027986)**	
Node	**Annotation**
pplax5	Aa_trans domain-containing protein.

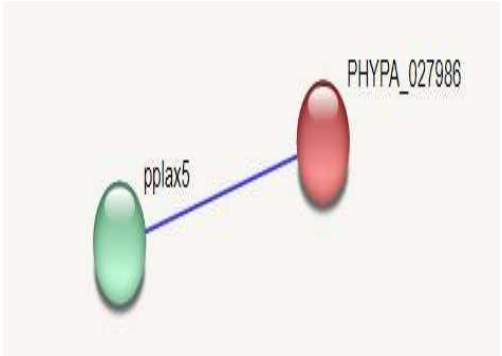


TABLE 4.8(H): Protein-Protein Interactions of *PpAAP8*

***PpAAP8* (PHYPA_011783)**	
Node	**Annotation**
PHYPA_009467	Uncharacterized protein; Belongs to the class-III pyridoxal-phosphate-dependent aminotransferase family.
PHYPA_001517	Uncharacterized protein; Belongs to the class-III pyridoxal-phosphate-dependent aminotransferase family.
PHYPA_003424	Delta-1-pyrroline-5-carboxylate synthase; P5CS plays a key role in proline biosynthesis, leading to osmoregulation in plants; In the N-terminal section; belongs to the glutamate 5- kinase family.
PHYPA_011198	Uncharacterized protein.
PHYPA_025321	V-type proton ATPase subunit; Subunit of the integral membrane V0 complex of vacuolar ATPase. Vacuolar ATPase is responsible for acidifying a variety of intracellular compartments in eukaryotic cells, thus providing most of the energy required for transport processes in the vacuolar system. Belongs to the V-ATPase V0D/AC39 subunit family.
PHYPA_022765	Vacuolar proton pump subunit B; Non-catalytic subunit of the peripheral V1 complex of vacuolar ATPase; Belongs to the ATPase alpha/beta chains family.
PHYPA_008865	ATP-synt_ab domain-containing protein.
PHYPA_029443	Uncharacterized protein.
PHYPA_011736	Uncharacterized protein.
PHYPA_024983	IGPS domain-containing protein.

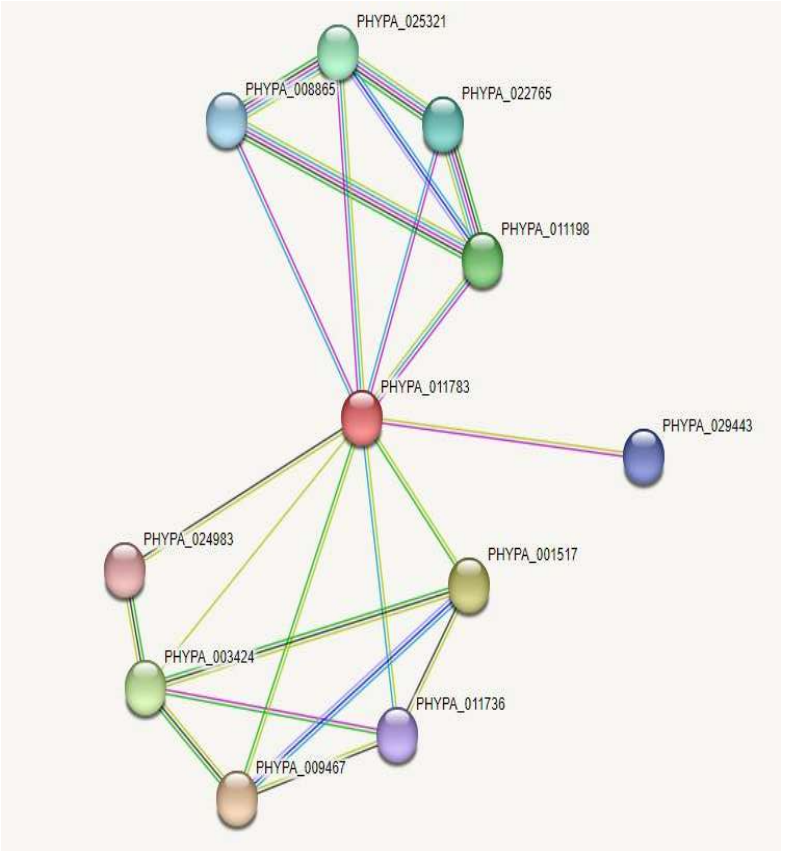


TABLE 4.8(I): Protein-Protein Interactions of *PpAAP9*

***PpAAP9* (PHYPA_009301)**	
Node	**Annotation**
PHYPA_022765	Vacuolar proton pump subunit B; Non-catalytic subunit of the peripheral V1 complex of vacuolar ATPase; Belongs to the ATPase alpha/beta chains family.
PHYPA_008865	ATP-synt_ab domain-containing protein.
PHYPA_022194	PRK domain-containing protein.
PHYPA_026353	Vacuolar proton pump subunit B; Non-catalytic subunit of the peripheral V1 complex of vacuolar ATPase; Belongs to the ATPase alpha/beta chains family.
PHYPA_027294	Vacuolar proton pump subunit B; Non-catalytic subunit of the peripheral V1 complex of vacuolar ATPase; Belongs to the ATPase alpha/beta chains family.
PHYPA_025832	PCRF domain-containing protein.

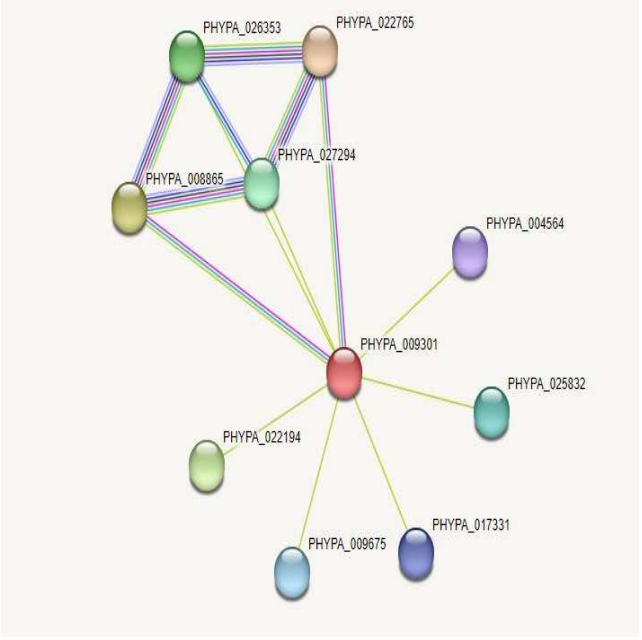


TABLE 4.8(J): Protein-Protein Interactions of *PpAAP10*

***PpAAP10* (PHYPA_015856)**	
Node	**Annotation**
PHYPA_017275	Aa_trans domain-containing protein.
PHYPA_018257	Aa_trans domain-containing protein.
PHYPA_008374	Aa_trans domain-containing protein.
PHYPA_012279	Aa_trans domain-containing protein.
PHYPA_016919	Aa_trans domain-containing protein.
PHYPA_028700	Aa_trans domain-containing protein.
$PHYPA_0$09675	RF_PROK_I domain-containing protein.
$PHYPA_0$17331	Uncharacterized protein.
$PHYPA_0$04564	Uncharacterized protein.

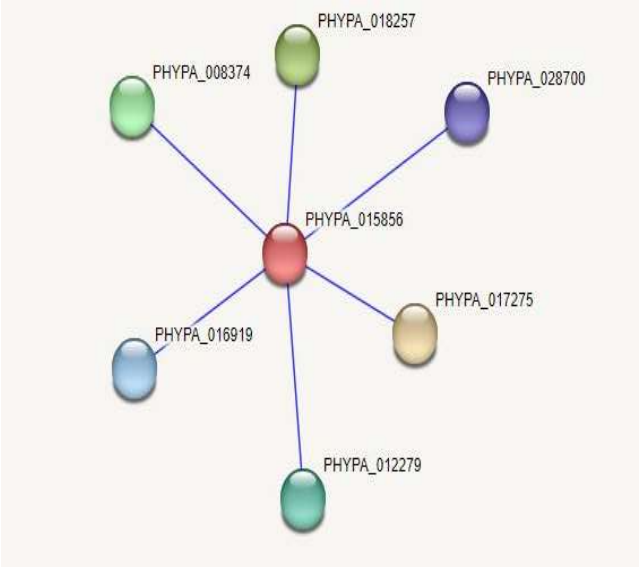


TABLE 4.8(K): Protein-Protein Interactions of *PpAAP11*

***PpAAP11* (PHYPA_012279)**	
Node	**Annotation**
PHYPA_012079	Aa_trans domain-containing protein.
PHYPA_012224	Aa_trans domain-containing protein.
PHYPA_009556	Aa_trans domain-containing protein.
PHYPA_019123	Aa_trans domain-containing protein.
PHYPA_003424	Delta-1-pyrroline-5-carboxylate synthase; P5CS plays a key role in proline biosynthesis, leading to osmoregulation in plants; In the N-terminal section; belongs to the glutamate 5- kinase family.
PHYPA_015856	Predicted protein.
PHYPA_018634	Uncharacterized protein.
PHYPA_013951	Uncharacterized protein.
pplax5	Aa_trans domain-containing protein.

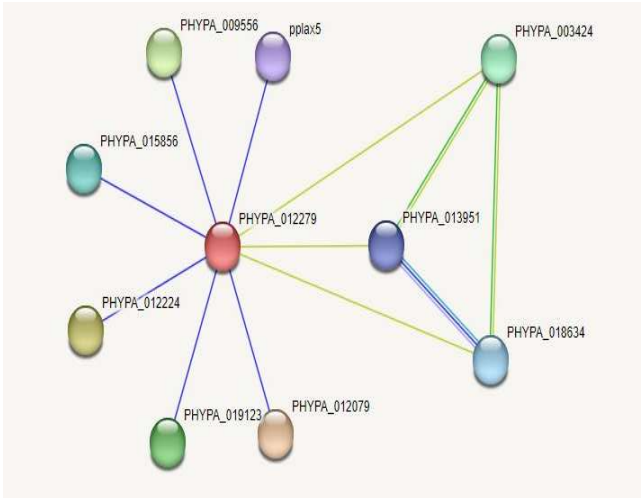


TABLE 4.8(L): Protein-Protein Interactions of *PpAAP12*

***PpAAP12* (PHYPA_008374)**	
Node	**Annotation**
PHYPA_009467	Uncharacterized protein; Belongs to the class-III pyridoxal-phosphate-dependent aminotransferase family.
PHYPA_001517	Uncharacterized protein; Belongs to the class-III pyridoxal-phosphate-dependent aminotransferase family.
PHYPA_003424	Delta-1-pyrroline-5-carboxylate synthase; P5CS plays a key role in proline biosynthesis, leading to osmoregulation in plants; In the N-terminal section; belongs to the glutamate 5- kinase family.
PHYPA_011198	Uncharacterized protein.
PHYPA_025321	V-type proton ATPase subunit; Subunit of the integral membrane V0 complex of vacuolar ATPase. Vacuolar ATPase is responsible for acidifying a variety of intracellular compartments in eukaryotic cells, thus providing most of the energy required for transport processes in the vacuolar system. Belongs to the V-ATPase V0D/AC39 subunit family.
PHYPA_022765	Vacuolar proton pump subunit B; Non-catalytic subunit of the peripheral V1 complex of vacuolar ATPase; Belongs to the ATPase alpha/beta chains family.
PHYPA_008865	ATP-synt_ab domain-containing protein.
PHYPA_029443	Uncharacterized protein.
PHYPA_011736	Uncharacterized protein.
PHYPA_024983	IGPS domain-containing protein.

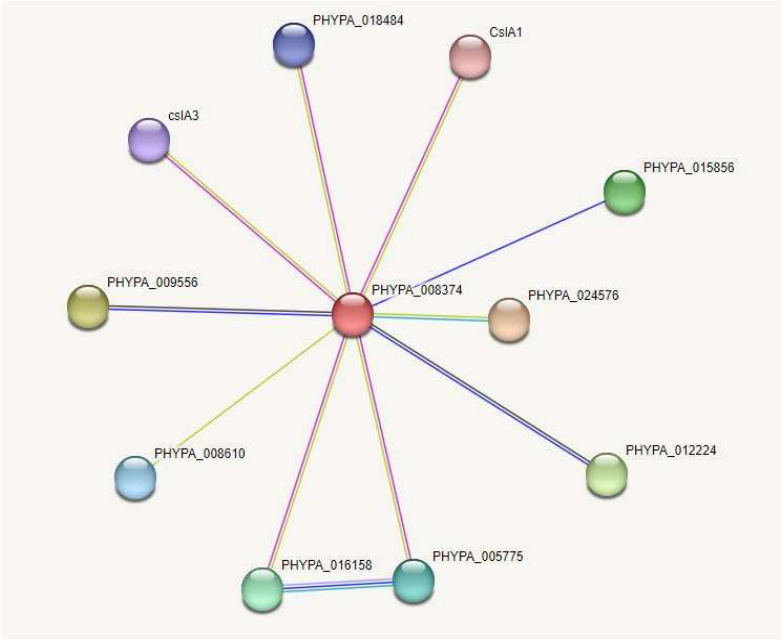


TABLE 4.8(M): Protein-Protein Interactions of *PpAAP13*

***PpAAP13* (PHYPA_026139)**	
Node	**Annotation**
PHYPA_003424	Delta-1-pyrroline-5-carboxylate synthase; P5CS plays a key role in proline biosynthesis, leading to osmoregulation in plants; In the N-terminal section; belongs to the glutamate 5- kinase family.
PHYPA_018634	Uncharacterized protein.
PHYPA_013951	Uncharacterized protein.

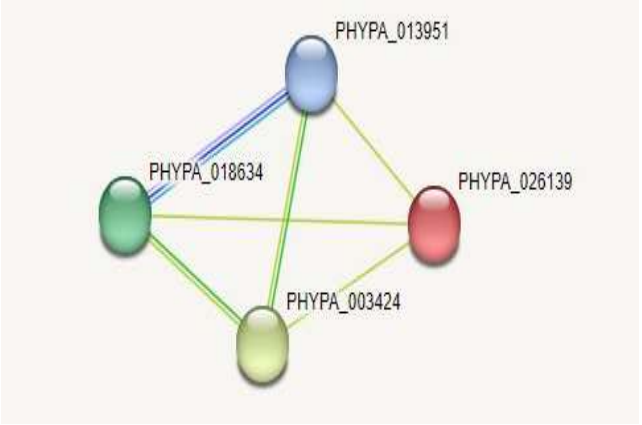


TABLE 4.8(N): Protein-Protein interactions of *PpAAP14*

***PpAAP14* (PHYPA_012961)**	
Node	**Annotation**
PHYPA_014274	Predicted protein; Belongs to the class-III pyridoxal-phosphate-dependent aminotransferase family.
PHYPA_012731	L-aspartate oxidase; Catalyzes the oxidation of L-aspartate to imino aspartate.
PHYPA_003424	Delta-1-pyrroline-5-carboxylate synthase; P5CS play a key role in proline biosynthesis, leading to osmoregulation in plants; In the N-terminal section; belongs to the glutamate 5- kinase family.
PHYPA_015550	PlsC domain-containing protein
PHYPA_011198	Uncharacterized protein
PHYPA_025321	V-type proton ATPase subunit; Subunit of the integral membrane V0 complex of vacuolar ATPase. Vacuolar ATPase is responsible for acidifying a variety of intracellular compartments in eukaryotic cells, thus providing most of the energy required for transport processes in the vacuolar system. Belongs to the V-ATPase V0D/AC39 subunit family.
PHYPA_023631	Predicted protein
PHYPA_020017	2-Hacid_dh domain-containing protein
PHYPA_020821	Formate dehydrogenase, mitochondrial; Catalyzes the NAD(+)-dependent oxidation of formate to carbon dioxide. Involved in the cell stress response; Belongs to the D-isomer specific 2-hydroxyacid dehydrogenase family. FDH subfamily.
PHYPA_022765	Vacuolar proton pump subunit B; Non-catalytic subunit of the peripheral V1 complex of vacuolar ATPase.

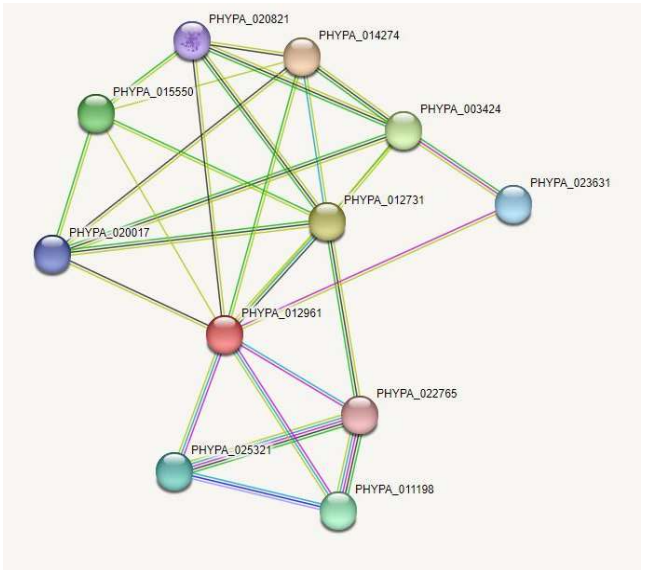


TABLE 4.8(O): Protein-Protein interactions of *PpAAP15*

***PpAAP15* (PHYPA_013530)**	
Node	**Annotation**
PHYPA_014274	Predicted protein; Belongs to the class-III pyridoxal-phosphate-dependent aminotransferase family.
PHYPA_012731	L-aspartate oxidase; Catalyzes the oxidation of L-aspartate to imino aspartate.
PHYPA_003424	Delta-1-pyrroline-5-carboxylate synthase; P5CS plays a key role in proline biosynthesis, leading to osmoregulation in plants; In the N-terminal section; belongs to the glutamate 5- kinase family.
PHYPA_015550	PlsC domain-containing protein
PHYPA_011198	Uncharacterized protein
PHYPA_025321	V-type proton ATPase subunit; Subunit of the integral membrane V0 complex of vacuolar ATPase. Vacuolar ATPase is responsible for acidifying a variety of intracellular compartments in eukaryotic cells, thus providing most of the energy required for transport processes in the vacuolar system. Belongs to the V-ATPase V0D/AC39 subunit family.
PHYPA_023631	Predicted protein
PHYPA_020017	2-Hacid_dh domain-containing protein
PHYPA_020821	Formate dehydrogenase, mitochondrial; Catalyzes the NAD(+)-dependent oxidation of formate to carbon dioxide. Involved in the cell stress response; Belongs to the D-isomer specific 2-hydroxy acid dehydrogenase family. FDH subfamily.
PHYPA_022765	Vacuolar proton pump subunit B; Non-catalytic subunit of the peripheral V1 complex of vacuolar ATPase; Belongs to the.

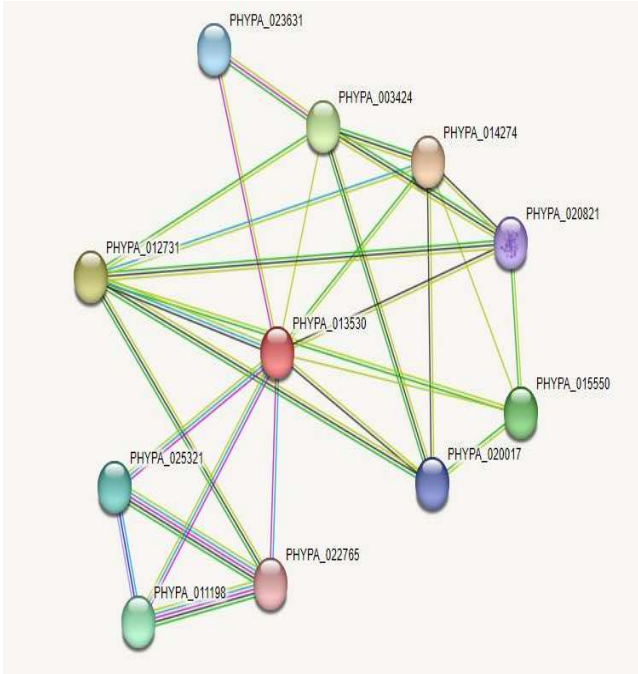
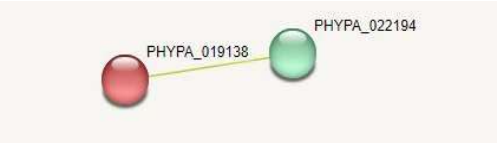


TABLE 4.8(P): Protein-Protein Interactions of *PpAAP16*

***PpAAP16* (PHYPA_019138)**	
Node	**Annotation**
PHYPA_022194	PRK domain-containing protein



Chapter 5

Discussion

Amino acid Permeases (*AAPs*) are specialized proteins that play an integral role in a cell's membrane, acting as gatekeepers of the cell, allowing some biomolecules to pass while restricting others [21]. Present in the lipid bilayer, they are also sometimes referred to as the "integral membrane proteins" and are one of the members of a larger family of transporter proteins called amino acid transporters. Within the amino acid transporter family, they lie under the class of Amino Acid/Auxin Permease Family (*AAAPs*) [3]. In this study, genome-wide identification and analysis of the amino acid permease family was conducted over a non-vascular bryophyte moss called *Physcomitrella patens*. Former studies have revealed much about the presence and role of various amino acid transporters in vascular plants but there have been few studies conducted on the identification and analysis of permeases in non-vascular plants. Our study presents the identification of *PpAAPs* along with their physicochemical properties, phylogenetic relationships as well as structural description, with comparison to *AtAAPs*. Former studies also suggest functional homology between similar identified gene families across species such as in the case of [37], in which various homologous were identified in an attempt to study the chloroplast transport components between a non-vascular and a vascular plant model. A similar conclusion has been drawn in [2], where the *NLP* gene family was comparatively studied between a vascular and non-vascular plant model,

where the function of *NLP* gene family was found conserved. Similar findings were observed in the present study.

The whole genome sequencing of *P. patens* was done in 2008 by [8] that provided foundational ground for this study to be conducted and the *PpAAP* genes to be studied both structurally and functionally. While it is true that genome-wide study pipelines cannot possibly confirm with accuracy, the existing molecular mechanisms or structures present inside the cell, nevertheless, such studies provide a foundational background for future, more detailed *in-vitro* studies and can perhaps, provide insight of important structural and functional attributes of a particular gene family. These studies are also helpful in the identification and analysis of a particular gene family with low economic cost, while quite precisely predicting the associations of gene families with certain concepts that might prove vital for academic research of the future, and industrial and commercial applications. For instance, in a former study conducted in 2015, which was a genome-wide association study of flowering time, maturity date, and plant height in *Glycine max*, it was concluded that the chromosomal regions and loci identified during the study may serve as promising targets for future studies in molecular mechanisms [38]. In the case of *Physcomitrella*, the preliminary information about *AAPs* may prove significant for various genetic engineering tools and techniques that may be implemented for the improvement of crop yield and production, identification of various amino acid synthesis pathways and improvement in Nitrogen-Use-Efficiency (NUE) which can bring about promising results for both, our environment and the agriculture sector.

In the present study, 16 *AAPs* were identified through *P. patens* genome database using various genome-wide computational tools. These *PpAAPs* were compared for their attributes with the *AAP* gene family of *Arabidopsis thaliana*. Since *in-silico* studies are based on comparison algorithms, therefore, the similarities found as a result of this comparison can be used to predict the function of a gene. It has been established through former studies that the conserved domain for *AAP* gene family is "Aa_trans" (PF01490) and "SLC5-6-like_sbd superfamily" (CL00456)

[24] and the endorsement of presence of these domains points out towards the evolutionary relationship that exists between *AtAAPs* and *PpAAPs*.

While comparing the physicochemical properties of *PpAAPs* and *AtAAPs*, it was found that the gene lengths of *PpAAPs* were significantly higher than those of *AtAAPs* while the protein lengths, molecular weights (MW) and pI (theoretical iso-electric point) values were found quite similar and in close proximity. This similarity indicates potential functional homology that might be present in the two plants. A similar study was conducted on the *14-3-3 gene* family of *Mangifera indica* (mango) and the physicochemical properties of the study organism was compared to that of apple's *14-3-3 gene* family. It was found that since the proteins of both mango and apple were acidic proteins, along with other consistently similar physicochemical parameters, they were stable proteins and the gene family shared functional homology as a result [39]. The sub-cellular localization was also found similar for both *P. patens* and *A. thaliana* i.e. plasma membrane, indicating a putative resemblance in function since they are both located in the same region, thus suggesting a similar role in the transportation of biomolecules and amino acids. The GRAVY (Grand Average of Hydropathicity) values show *PpAAPs* to be basic proteins while *AtAAPs* as acidic proteins, however since they both lie in the plasma membrane, it is likely that their rudimentary roles as transporters are more alike than different, if not the same.

In addition to this, phylogenetic analysis was done to study the evolutionary divergence and variation present within the *PpAAPs* and between the *AAPs* of *Physcomitrella patens* and other vascular plants. Firstly, it was confirmed that both the non-vascular *Physcomitrella* and the vascular plants chosen for observation, did indeed arise from a common ancestor. Former studies support this argument, for instance, a similar conclusion was drawn by [24], while studying the importance of evolution in amino acid permeases in 17 plants, revealing the existence of common ancestry among bryophytes and vascular plants. Moreover, former studies also support using the neighbor-joining method to study the evolutionary relationship between plants. For instance, in a former study of comparison

between rice and *Arabidopsis* done by [40], this method was used to draw an understanding of the phylogenetic relationship of the Stress Associated Protein (*SAP*) gene family. Moreover, it was observed that variation existed between *PpAAPs*, where it is evident that *PpAAP1* and *PpAAP2* lie in close proximity with one other but are diverse when compared to the rest of *PpAAPs*, thus portraying structural and functional diversity that exists between *PpAAPs*. It was also observed that *PpAAP3*-*PpAAP16* were present in close vicinity and would likely be conserved and similar. The presence of such similarity is also observable from comparative analysis such as in the case of *PpAAP9* and *PpAAP16*, which are the closest neighbors to one other, as shown by the phylogenetic tree, share similar consensus motifs and have a similar gene structure. Similar analogies were noticed in between *PpAAP5* and *PpAAP7*, *PpAAP3* and *PpAAP4*, *PpAAP11* and *PpAAP13*, and *PpAAP14* and *PpAAP15*. *PpAAPs* that lie in closer proximity to each other were more likely to have identical motif regions and gene structure and were more likely to be functionally homologous. In addition, the presence of similar conserved domains between *PpAAPs* and *AtAAPs* further elucidates the evolutionary relationship between non-vascular and vascular plants.

The study of consensus motifs between *PpAAPs* presents considerable similarity and thus, sufficiently predicts that all genes under study, indeed belong to the *AAP* gene family of *Physcomitrella patens*. Gene structure study predicts diversity between *PpAAPs* and *AtAAPs*. Apart from the quantitative difference of exons between *PpAAPs* and *AtAAPs*, it was also observed that members of *AtAAPs* had 5-6 introns while for *PpAAPs*, it varied between 4-9 introns.

Protein-Protein interactions of *PpAAPs* when compared to *AtAAPs* also had promising results as the interacting proteins for both *PpAAPs* and *AtAAPs*, had an ample quantity of interacting proteins with Aa_trans domain, indicating their role as transporters. Former studies have shown that proteins interact with each other to perform a particular function [36] hence, knowing a particular function of a group of interacting entities of a protein may help us predict the function of that particular protein. In a related study by [41], the *TOPLESS* gene family

of tomato was focused, specifically the *SlTPL* proteins in their interactive relationship to 17 different *SlIAA* proteins. It was concluded that *SlTPL1*, *SlTPL2*, *SlTPL4*, and *SlTPL5* interacted with most *SlIAAs* proteins however *SlTPL3* and *SlTPL6* showed limited growth when co-expressed with Aux/IAA–AD fusion proteins, thus describing a slight difference in their functional nature and the role of interacting proteins in describing it. In the case of *PpAAPs*, the majority of the interacting proteins are involved in the transportation of biomolecules and amino acids, hence, it is safe to predict that the *PpAAPs* are also involved in the transport of biomolecules. Most notable were *PpAAP1*, *PpAAP2*, *PpAAP3*, *PpAAP4*, *PpAAP7*, *PpAAP10*, *PpAAP11* and *PpAAP12* which had interacting proteins having Aa_trans domain containing interacting proteins indicating to their role as transporters.

Chapter 6

Conclusion and Future Work

Amino acid permeases (*AAPs*) play a central role in cellular metabolism and are, therefore, important proteins in many ways. Our understanding of their composition, their role, and their interactions may shape or pave a new way for academic discoveries, improvement in industrial and commercial projects, and genetic engineering processes. This study aimed to identify and analyze, both structurally and functionally, the amino acid permease gene family in a non-vascular bryophyte in comparison to a vascular plant. For this reason, *Physcomitrella patens*, which is a non-vascular bryophyte moss was chosen for this study. To begin, the sequences of *AAP* genes of *Arabidopsis thaliana* were retrieved from the database, followed by protein-BLAST using NCBI and *Physcomitrella patens* selected as the organism of choice. The resultant sequences, after thorough screenings and removals, were narrowed down to finalized samples, which were then subjected to various computational tools for genome-wide analysis of *AAP* gene family of *P. patens*. Various analyses were performed including conserved domain identification, physicochemical characterization and localization, motif composition, phylogenetic analysis, protein-protein interaction and gene structure determination to structurally and functionally analyze the *AAP* gene family of *P. patens*. Following the successful identification of the *AAP* gene family in *P. patens*, the physicochemical characterization and localization, along with the observation of motifs and the interacting proteins, the amino acid permeases of *P. patens* were found quite similar to those

of *A. thaliana*. The phylogenetic analysis showed evolutionary divergence and variation present between the *AAP* gene family of *P. patens* when compared to other vascular plants. The method used was, therefore, effective and successful in providing in-depth information about the selected gene family. These genes have theoretically been proven to be a part of the *AAP* gene family in *P. patens* and predicts their structural and functional conservation in *P. patens* compared to the model plant *A. thaliana*.

The method used in this study was found successful and quite effective in achieving the aim of this study therefore, this method is proposed to be used in the study of other plants in the future. Other gene families may be identified using the method, having industrial or commercial benefits, with low economic cost and less time consumption. The information revealed from this study provides a solid theoretical background on the subject, however, the study can be proceeded in wet-lab for affirmation of the *AAP* gene family of *P. patens* using molecular laboratory techniques such as the Polymerase Chain Reaction (PCR).

This study paves the way for other gene families to be identified and industrial techniques to be improved which can, in turn, enhance crop production, yield, and efficient use of nitrogen in the soil. Moreover, genetic engineering techniques can be improved which can greatly influence the agriculture sector. These enhancements may hold promising results for our environment while refining various industrial processes at the same time.

Bibliography

[1] Takayuki Sekito, Yuki Fujiki, Yoshinori Ohsumi, and Yoshimi Kakinuma. Novel families of vacuolar amino acid transporters. *IUBMB life*, 60(8):519–525, 2008.

[2] Sami Ullah Jan, Maha Rehman, Tahmina Nazish, Sohail Ahmad Jan, Ayesha Liaqat, Mahmoud Moustafa, Alvina Gul, Noor Ul Huda, Syeda Marriam Bakhtiyar, Sarah Gul, et al. Genome-wide study and expression analysis of nodule-inception-like protein (nlp) gene family in physcomitrella patens reveals its role in nitrogen response. *Pakistan Journal of Botany*, 55(1):113–128, 2023.

[3] Xuehui Yao, Jing Nie, Ruoxue Bai, and Xiaolei Sui. Amino acid transporters in plants: Identification and function. *Plants*, 9(8):972, 2020.

[4] Margaret R Taylor, Anke Reinders, and John M Ward. Transport function of rice amino acid permeases (aaps). *Plant and Cell Physiology*, 56(7):1355–1363, 2015.

[5] Manoela Miranda, Ljudmilla Borisjuk, Annegret Tewes, Ute Heim, Norbert Sauer, Ulrich Wobus, and Hans Weber. Amino acid permeases in developing seeds of vicia faba l.: expression precedes storage protein synthesis and is regulated by amino acid supply. *The Plant Journal*, 28(1):61–71, 2001.

[6] Paulinus P Engel. The induction of biochemical and morphological mutants in the moss physcomitrella patens. *American Journal of Botany*, 55(4):438–446, 1968.

[7] David Cove, Magdalena Bezanilla, Phillip Harries, and Ralph Quatrano. Mosses as model systems for the study of metabolism and development. *Annu. Rev. Plant Biol.*, 57:497–520, 2006.

[8] Stefan A Rensing, Daniel Lang, Andreas D Zimmer, Astrid Terry, Asaf Salamov, Harris Shapiro, Tomoaki Nishiyama, Pierre-François Perroud, Erika A Lindquist, Yasuko Kamisugi, et al. The physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science*, 319(5859):64–69, 2008.

[9] Hansjörg AW Schneider-Poetsch, Stefan Marx, H Üner Kolukisaoglu, Sabine Hanelt, and Birgit Braun. Phytochrome evolution: phytochrome genes in ferns and mosses. *Physiologia Plantarum*, 91(2):241–250, 1994.

[10] Didier G Schaefer and Jean-Pierre Zrÿd. The moss physcomitrella patens, now and then. *Plant physiology*, 127(4):1430–1438, 2001.

[11] N van Gessel, D Lang, and R Reski. Genetics and genomics of physcomitrella patens. *Plant Cell Biology*, 20:1–32, 2017.

[12] David W Meinke, J Michael Cherry, Caroline Dean, Steven D Rounsley, and Maarten Koornneef. Arabidopsis thaliana: a model plant for genome analysis. *Science*, 282(5389):662–682, 1998.

[13] Qiaoping Yuan, Shu Ouyang, Jia Liu, Bernard Suh, Foo Cheung, Razvan Sultana, Dan Lee, John Quackenbush, and C Robin Buell. The tigr rice genome annotation resource: annotating the rice genome and creating resources for plant biologists. *Nucleic acids research*, 31(1):229–233, 2003.

[14] RH Moll, EJ Kamprath, and WA Jackson. Analysis and interpretation of factors which contribute to efficiency of nitrogen utilization 1. *Agronomy journal*, 74(3):562–564, 1982.

[15] Trevor Garnett, Vanessa Conn, and Brent N Kaiser. Root based approaches to improving nitrogen use efficiency in plants. *Plant, cell & environment*, 32(9):1272–1283, 2009.

[16] Mechthild Tegeder and Céline Masclaux-Daubresse. Source and sink mechanisms of nitrogen transport and use. *New phytologist*, 217(1):35–53, 2018.

[17] RK Fagodiya, H Pathak, A Kumar, A Bhatia, and N Jain. Global temperature change potential of nitrogen use in agriculture: A 50-year assessment. *Scientific reports*, 7(1):44928, 2017.

[18] Alamin Alfatih, Jie Wu, Zi-Sheng Zhang, Jin-Qiu Xia, Sami Ullah Jan, Lin-Hui Yu, and Cheng-Bin Xiang. Rice nin-like protein 1 rapidly responds to nitrogen deficiency and improves yield and nitrogen use efficiency. *Journal of Experimental Botany*, 71(19):6032–6042, 2020.

[19] Adriana Ortiz-Lopez, H-C Chang, and DR Bush. Amino acid transporters in plants. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1465(1-2):275–280, 2000.

[20] X Liu and DR Bush. Expression and transcriptional regulation of amino acid transporters in plants. *Amino acids*, 30:113–120, 2006.

[21] Bruno André. An overview of membrane transport proteins in saccharomyces cerevisiae. *Yeast*, 11(16):1575–1611, 1995.

[22] François Verrey, Ellen I Closs, Carsten A Wagner, Manuel Palacin, Hitoshi Endou, and Yoshikatsu Kanai. Cats and hats: the slc7 family of amino acid transporters. *Pflügers Archiv*, 447:532–542, 2004.

[23] LibreTexts. Membrane Transport w/Selective Permeability, url=https://bio.libretexts.org/@go/page/8193., urldate=3 June 2019.

[24] Ting Zhou, Cai-peng Yue, Jin-yong Huang, Jia-qian Cui, Ying Liu, Wen-ming Wang, Chuang Tian, and Ying-peng Hua. Genome-wide identification of the amino acid permease genes and molecular characterization of their transcriptional responses to various nutrient stresses in allotetraploid rapeseed. *BMC plant biology*, 20(1):1–22, 2020.

[25] Chao Zhang, Nana Kong, Minxuan Cao, Dongdong Wang, Yue Chen, and Qin Chen. Evolutionary significance of amino acid permease transporters

in 17 plants from chlorophyta to angiospermae. *BMC genomics*, 21(1):1–11, 2020.

[26] Muhammad Irfan Fareed, Sami Ullah Jan, Tahmina Nazish, Ayesha Liaqat, Ayesha Javaid, Mohsin Ali, Farees Ud Din Mufti, Muhammad Fayyaz, Maha Rehman, Muhammad Jamil, et al. Nodule-inception-like protein (nlp) gene family identified in physcomitrella patens genome responds to variable nitrogen supply. 2021.

[27] Mechthild Tegeder and John M Ward. Molecular evolution of plant aap and lht amino acid transporters. *Frontiers in plant science*, 3:21, 2012.

[28] Jens Tilsner, Nina Kassner, Christine Struck, and Gertrud Lohaus. Amino acid contents and transport in oilseed rape (brassica napus l.) under different nitrogen conditions. *Planta*, 221:328–338, 2005.

[29] Li-Zong Hu, Wei-Bo Yin, Yu-Hong Chen, and Zan-Min Hu. Functional divergence and evolutionary dynamics of the putative aaap gene family in brassica rapa. *Plant molecular biology reporter*, 32:517–530, 2014.

[30] Lei Sheng, Lin Deng, HanWei Yan, Yang Zhao, Qing Dong, Qian Li, XiaoYu Li, BeiJiu Cheng, HaiYang Jiang, et al. A genome-wide analysis of the aaap gene family in maize. *Journal of Proteomics and Bioinformatics*, 7(1):023–033, 2014.

[31] Lin Cheng, Hong-Yu Yuan, Ren Ren, Shi-Qi Zhao, Ya-Peng Han, Qi-Ying Zhou, Dan-Xia Ke, Ying-Xiang Wang, and Lei Wang. Genome-wide identification, classification, and expression analysis of amino acid transporter gene family in glycine max. *Frontiers in Plant Science*, 7:515, 2016.

[32] Jingrong Ma, Chuang Pan, Haiming Chen, Weijun Chen, Wenxue Chen, Ming Zhang, and Qiuping Zhong. Insight of the functional and biological activities of coconut (cocos nucifera l.) protein by proteomics analysis and protein-based bioinformatics. *Molecules*, 27(9):2987, 2022.

[33] Jiujun Du, Changjian Du, Xiaolan Ge, Shuangshuang Wen, Xinglu Zhou, Lei Zhang, and Jianjun Hu. Genome-wide analysis of the aaap gene family in

populus and functional analysis of psaaap21 in root growth and amino acid transport. *International Journal of Molecular Sciences*, 24(1):624, 2022.

[34] Damian Szklarczyk, Annika L Gable, David Lyon, Alexander Junge, Stefan Wyder, Jaime Huerta-Cepas, Milan Simonovic, Nadezhda T Doncheva, John H Morris, Peer Bork, et al. String v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research*, 47(D1):D607–D613, 2019.

[35] Myriam Roussigne, Sophia Kossida, Anne-Claire Lavigne, Thomas Clouaire, Vincent Ecochard, Alexandra Glories, François Amalric, and Jean-Philippe Girard. The thap domain: a novel protein motif with similarity to the dna-binding domain of p element transposase. *Trends in biochemical sciences*, 28(2):66–69, 2003.

[36] Ozlem Keskin, Attila Gursoy, Buyong Ma, and Ruth Nussinov. Principles of protein-protein interactions: what are the preferred ways for proteins to interact? *Chemical reviews*, 108(4):1225–1244, 2008.

[37] Nancy Rosenbaum Hofmann and Steven M Theg. Physcomitrella patens as a model for the study of chloroplast protein transport: conserved machineries between vascular and non-vascular plants. *Plant Molecular Biology*, 53:643–654, 2003.

[38] Jiaoping Zhang, Qijian Song, Perry B Cregan, Randall L Nelson, Xianzhi Wang, Jixiang Wu, and Guo-Liang Jiang. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (glycine max) germplasm. *BMC genomics*, 16:1–11, 2015.

[39] Liming Xia, Xinhua He, Xing Huang, Haixia Yu, Tingting Lu, Xiaojie Xie, Xuemei Zeng, Jiawei Zhu, and Cong Luo. Genome-wide identification and expression analysis of the 14-3-3 gene family in mango (mangifera indica l.). *International Journal of Molecular Sciences*, 23(3):1593, 2022.

[40] Shubha Vij and Akhilesh K Tyagi. Genome-wide analysis of the stress associated protein (sap) gene family containing a20/an1 zinc-finger (s) in rice

and their phylogenetic relationship with arabidopsis. *Molecular Genetics and Genomics*, 276:565–575, 2006.

[41] Yanwei Hao, Xinyu Wang, Xian Li, Carole Bassa, Isabelle Mila, Corinne Audran, Elie Maza, Zhengguo Li, Mondher Bouzayen, Benoît Van Der Rest, et al. Genome-wide identification, phylogenetic analysis, expression profiling, and protein–protein interaction properties of topless gene family members in tomato. *Journal of experimental botany*, 65(4):1013–1023, 2014.