NARDI FUNDULEA, ROMANIA ROMANIAN AGRICULTURAL RESEARCH, NO. 43, 2026
www.incda-fundulea.ro; doi.org/10.59665/rar4365 First Online: May, 2026. DII 2067-5720 RAR 2026-145

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

**Taimoor Abdul Sattar[1], Sohail Ahmad Jan[1*], Jameel M. Al-Khayri[2*], Sami Ullah Jan[1], Ayesha Bibi[3], Othman Al-Dossary[2], Bader Alsubaie[2], Mohammed I. Aldaej[2]**

[1]Department of Bioinformatics and Biosciences, Capital University of Science and Technology, Islamabad, Pakistan
[2]Department of Agricultural Biotechnology, College of Agriculture and Food Sciences, King Faisal University, Al-Ahsa, Saudi Arabia
[3]Department of Human Nutrition and Dietetics, Women University Mardan, Pakistan
[*]Corresponding authors. E-mail: sjan.parc@gmail.com; jkhayri@kfu.edu.sa

## 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. Although *AAPs* have been extensively studied in vascular plants, they have not yet been explicitly reported in non-vascular bryophytes. In the present study, 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 napu,* and *Arabidopsis thaliana*. Results proposed that the *PpAAPs* indeed belong to the Amino Acid Permease (*AAP*) gene family and shared significant structural and functional homology with *AtAAPs*.**

**Keywords:** Amino Acid Permeases (*AAPs*), transporter protein, *Physcomitrella patens*, *Arabidopsis thaliana*, nitrogen use efficiency.

## 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 membrane, play an important role in the ingestion of these amino acids. They are specialized proteins that assist transportation across cellular membrane. They are group of proteins that recognizes and binds to certain amino acids and helps them move into or out of the cell based on physiological demands and concentration gradients (Sekito et al., 2008). Amino acid transporters have been reported in several vascular plants and many important biomolecules such as Nitrogen (N), Potassium (K) and Sodium (Na) are in direct interaction with these transporter families which are responsible for permeability (Jan et al., 2023). Research suggests that these nutrients play an essential role in reproduction, development and growth of plants (Yao et al., 2020).

Amino acid transporters (*AATs*) can be further divided into several divisions

Received 25 March 2026; accepted 19 May 2026.

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*) (Yao et al., 2020). Amino Acid Permeases (*AAPs*) are a sub-family within the Amino Acid/Auxin Permease (*AAAP*) family. Generally, *AAPs* play important roles in uptake of amino acids from soil, long distance transport of amino acids from soil to other parts of plant through 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 phloem. *AtAAP2* is known to be expressed in the vascular tissue and is involved in long-distance transport of amino acids, ensuring a steady supply of nutrients. *AtAAP3* is involved in 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 uptake of amino acids from the soil while *AtAAP6* plays a vital role in redistribution of amino acids during stress conditions. Some *AAPs* have fewer known functions as in the case of *AtAAP7* but is known to play a role in nutrients and ion transport since it belongs to the *AAP* family. *AtAAP8* has been reported to supply amino acids in developing embryos. In a similar study on rice (*Oryza sativa*), researchers focused on the function of four *AAPs* of *O. sativa*, 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 that studied earlier in *A. thaliana* (Taylor et al., 2015).

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. Therefore, the current study is conducted 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*, and to study the evolutionary relationship of *AAP* gene family between vascular and non-vascular plants.

## MATERIAL AND METHODS

### Screening of Genome Databases

The full-length gene, protein, and coding sequences of all members of *Arabidopsis thaliana AAP* (*AtAAP*) gene family were retrieved from *Arabidopsis* genome database (TAIR: http://arabidopsis.org/). Two genome databases i.e. NCBI (http://ncbi.nlm.nig.gov/) and Phytozome (v13: http://phytozome-next.jgi.doe.gov) were screened for putative *PpAAP* sequences. All of the retrieved sequences were aligned to eliminate alternative spliced variants as well as redundant sequences (Jan et al., 2023).

### Removal of Redundant Sequences

All of the potential *PpAAP* sequences obtained, were aligned to remove any repetitive or redundancy sequences. 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. All of the 16 selected sequences were assigned new lab IDs (*PpAAP1-PpAAP2* and so on).

### 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 quantitative position of gene, chromosome number and gene length.

### Physicochemical Properties and Localization of *PpAAPs*

The sequences selected on the basis of conserved domains (*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) (Jan et al., 2023). 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 catalogue 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). Both of these analyses were performed for scale sequences (*AtAAPs*) as well as sample sequences (*PpAAPs*), for comparison.

### Motif Composition in *PpAAP* Gene Family

The presence of consensus motifs was determined using an online tool called Multiple Expectation Maximizations for Motif Elicitation or MEME v5.5.4 (https://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 an exception of motif-finding threshold, which was kept at 20 to ensure specificity and precision (Jan et al., 2023).

### 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*, *Brassica napus*, *Vicia faba*, *Brassica rapa*, *Zea mays*, *Glycine max*, *Raphanus sativus*, *Brassica oleracea*, *Cannabis sativa*, *Eucalyptus grandis* and *Cocus nucifera*. The *AAP* gene sequences for all plants were retrieved through specific literature review and by using specific keywords in NCBI. Multiple alignment was initially performed through Clustal Omega tool (https://www.ebi.ac.uk/Tools/msa/clustalo/). The guide/newick tree generated from Clustal Omega tool was used as query for visualization of rooted phylogenetic tree in another online tool called Interactive Tree of Life v6 (https://itol.embl.de/) (Jan et al., 2023).

### Gene Structure Determination in *PpAAP* Family

The full-length and coding sequences of *PpAAPs* and *AtAAPs* were retrieved and used to examine 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 un-translated regions (UTRs) present in the sequences (Jan et al., 2023). The tool also helped in comparing gene lengths between the two families.

## RESULTS AND DISCUSSION

### Genome-Wide Identification and Analysis of *AAP* Gene in *Physcomitrella patens*

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 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*", "*Amino Acid Permeases*" and through the accession numbers isolated from NCBI. A total of 23 sequences were extracted and were 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 *PpAAPs* was generated as shown in Table 1.

*Table 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]* |

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) (Zhou et al., 2020). 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 2 confirming that the common domain of Amino Acid Permeases as present in *A. thaliana* protein sequences is present in *P. patens* sequences.

*Table 2.* Features of conserved domains identified in *AtAAPs* and *PpAAPs*

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

**Determination of Physicochemical Properties and Localization in *PpAAPs***

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 of 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, 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 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 3).

**Identification of Consensus Motifs**

Composition of motif regions were 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, higher similarity between motif regions point to functional homology itself. Figure 1 shows the presence of conserved motif regions with 8 motifs (or 40%) present in all *PpAAPs*. 8 motifs (or 40%) were present in 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 majority of motifs in all *PpAAPs* indicate that they indeed belong to the same family and clued to be similar in function. Figure 2 shows all of the 20 motifs and their sequences, present in the *PpAAPs*, as identified by the MEME tool.

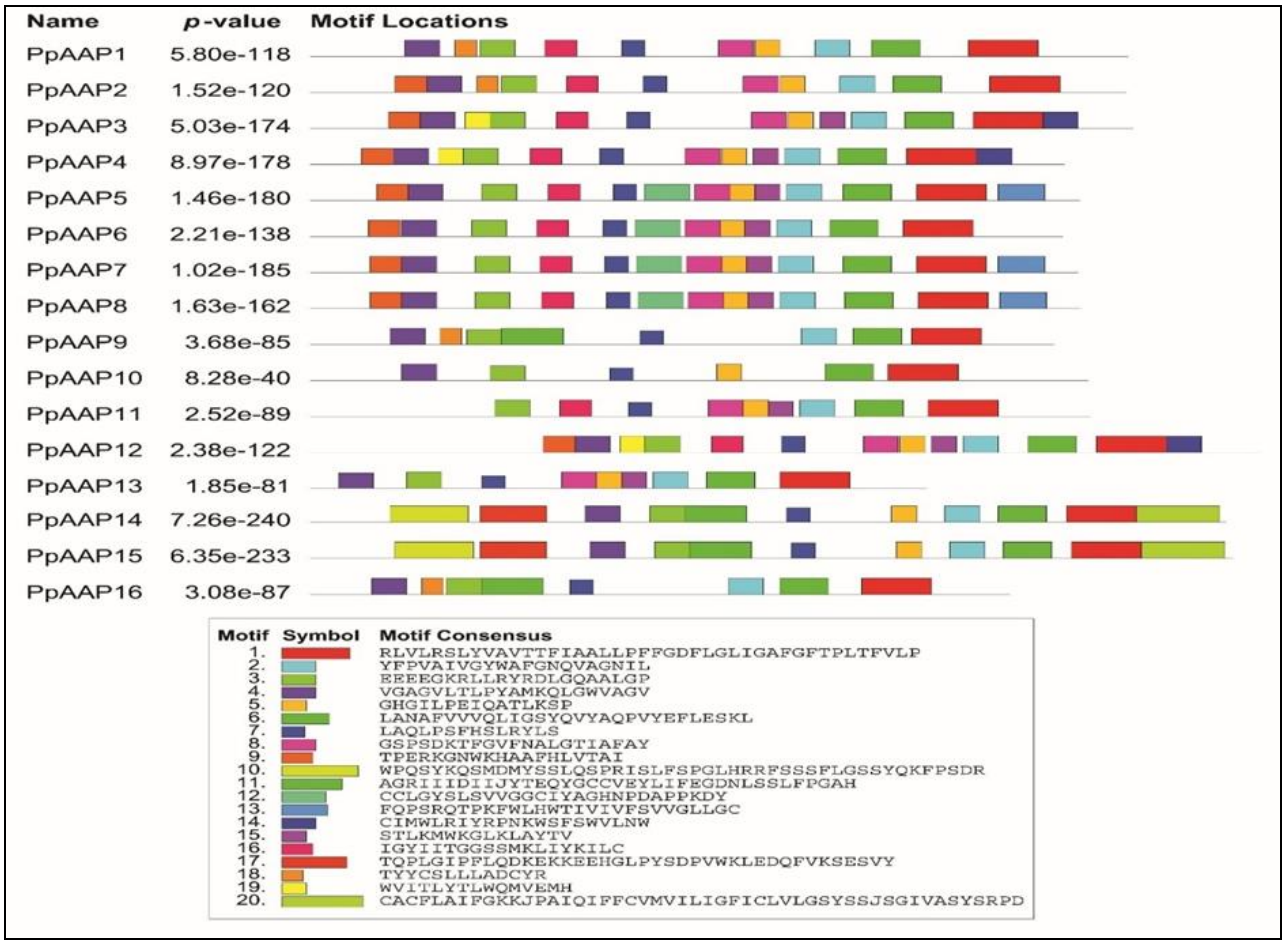



*Figure 1.* Identification of Conserved Motifs in *PpAAPs*

*Table 3*. Physicochemical properties and Localization of *AtAAPs* and *PpAAAPs*

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

*Table 3*. Continued

| 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 |

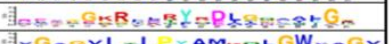
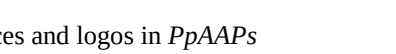
| Motif | Sequence | Logo |
|---|---|---|
| Motif 1 | RLVLRSLYVAVTTFIAALLPFFGDFLGLIGAFGFTPLTFVLP |  |
| Motif 2 | YFPVAIVGYWAFGNQVAGNIL |  |
| Motif 3 | EEEEGKRLLRYRDLGQAALGP |  |
| Motif 4 | VGAGVLTLPYAMKQLGWVAGV |  |
| Motif 5 | GHGILPEIQATLKSP |  |
| Motif 6 | LANAFVVVQLIGSYQVYAQPVYEFLESKL |  |
|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

*Figure 2*. List of Consensus Motifs with their respective sequences and logos in *PpAAPs*

**Sequence Alignment and Phylogenetic Relationship of the *AAP* Gene Family**

The *PpAAPs* and the *AtAAPs* were compared against each other, prior to analysis, to confirm 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 spliced variants. This allowed for evolutionary diversity to be identified among all involved members of *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 Clustal Omega tool was then used as query for visualization of rooted phylogenetic tree in another online tool called Interactive Tree of Life v6. Figure 3 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 in 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 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*. This distribution of *AAP* gene family represented the presence of imperative evolutionary relationship between the *AAP* family across species and established significant evolutionary divergence in between non-vascular bryophytes and vascular tracheophytes.

**Determination of Gene Structure**

The gene and coding sequences of *AtAAPs* and *PpAAPs* were used to analyze their structural features which included the identification of exons, introns and un-translated regions (UTRs). This was done using an online server called GSDS or Gene Structure Display Server (http://gsds.gao-lab.org/). Figure 4 displays the result of gene structure determination. It was observed that the lowest number of exon present was in *PpAAP9*, which was 1, followed by the presence of 2 exons in *PpAAP16*. The number of exons for the rest of *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 were observed in *PpAAP14* and *PpAAP15*, which were 10 and 11, respectively.

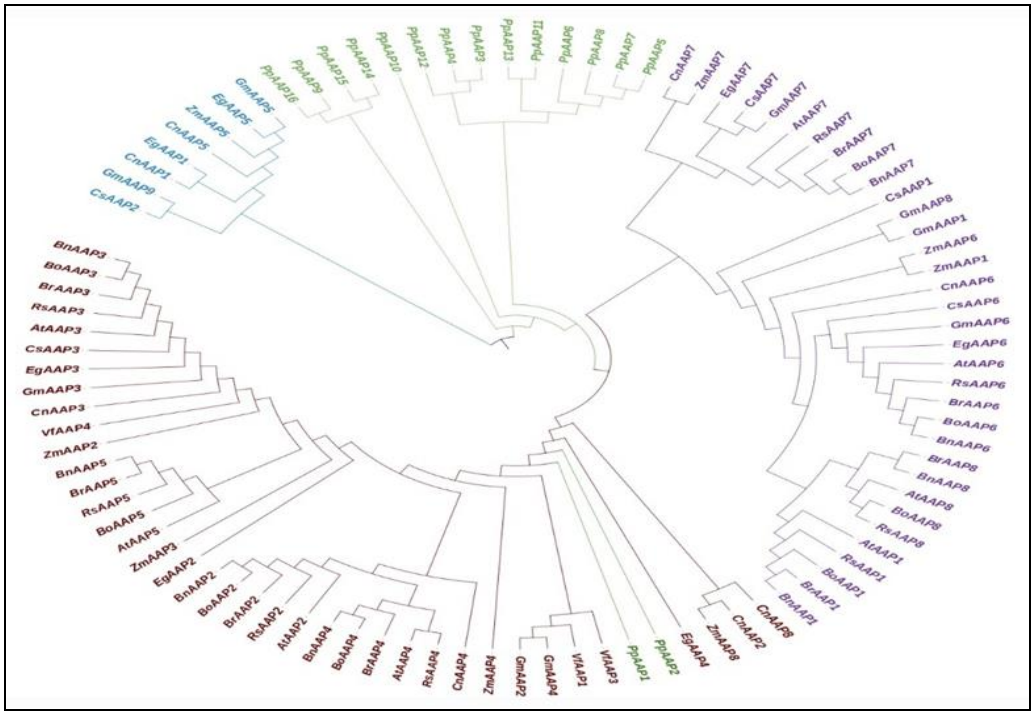



*Figure 3.* Phylogenetic Analysis of *PpAAPs* with different economically important plant species

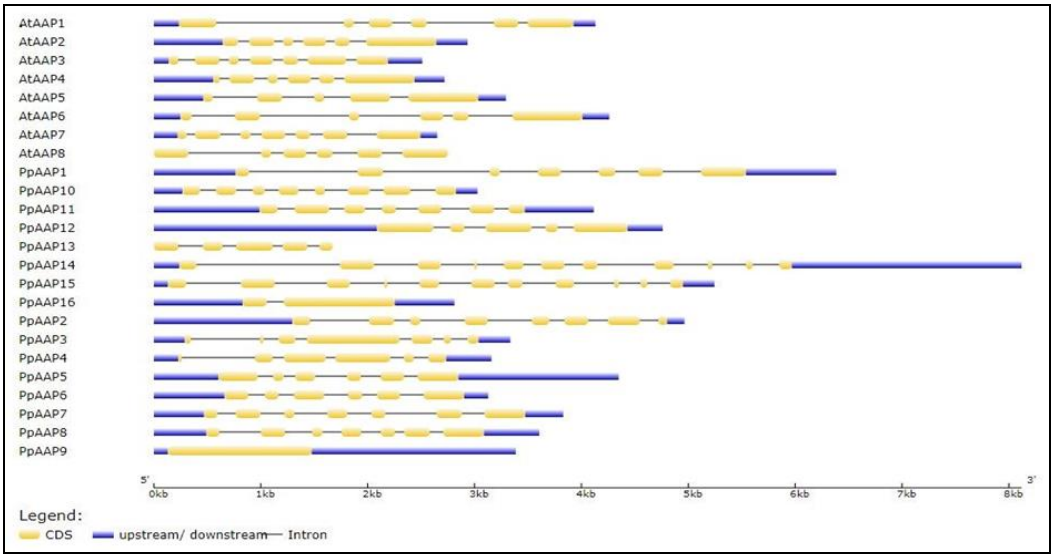



*Figure 4*. Gene structure determination

Amino Acid Permeases (*AAPs*) are specialized proteins that play an integral role in cell's membrane, acting as gatekeepers of the cell, allowing some biomolecules to pass while restricting others (Andre, 1995). 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*) (Yao et al., 2020). In this study, genome-wide identification and analysis of amino acid permease family was conducted over a non-vascular bryophyte moss called *P. 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 Hofmann and Theg (2003) 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 the study of Jan et al. (2023) 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.

Abiotic factors affect plant growth and development. Proper identification and characterization of abiotic stress related genes play key role against abiotic stresses (Shinwari et al., 2020; Khan et al., 2023; Farooq et al., 2025; Saleem et al., 2025). The whole genome sequencing of *P. patens* was done by Rensing et al. (2008) 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 identification and analysis of a particular gene family with low economical cost, while quite precisely predicting the associations of gene families with certain concepts that might prove vital for academic researches of the future, and industrial and commercial applications. For instance, 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 (Zhang et al., 2015).

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) (Zhou et al., 2020) 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 of it (Xia et al., 2022). 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 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 (Zhou et al., 2020), 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 neighbor joining method to study the evolutionary relationship between plants. For instance, in a former study of comparison between rice and *Arabidopsis* done by (Vij and Tyagi, 2006) 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 introns.

## CONCLUSIONS

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 new way for academic discoveries, improvement in industrial and commercial projects and genetic engineering processes. The aim of this study was 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. 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 *Arabidopsis 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 *Physcomitrella patens* and predicts their structural and functional conservation in *Physcomitrella patens* compared to the model plant *Arabidopsis 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 future.

## FUNDING

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. KFU262309].

## ACKNOWLEDGEMENTS

The authors extend their appreciation for the support of the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. KFU262309].

## REFERENCES

Andre, B., 1995. *An overview of membrane transport proteins in Saccharomyces cerevisiae*. Yeast, 11(16): 1575-1611.

Farooq, H., Mazhar, R., Jan, S.A., 2025. *Mitigating climate change impacts on agriculture through AI-driven crop improvement*. MOJ Biol. Med., 10(1): 29-34.

Hofmann, N.R., and Theg, S.M., 2003. *Physcomitrella patens as a model for the study of chloroplast protein transport: conserved machineries between vascular and non-vascular plants*. Plant Mol. Biol., 53(5): 621-632.

Jan, S.U., Yousaf, M., Zafar, N., 2023. *Genome-wide study and expression analysis of Nodule-inception-like protein (NLP) gene family in Physcomitrella patens reveal its role in nitrogen response*. Pak. J. Bot., 55(1): 113-128.

Khan, A., Bibi, F., Jan, S.A., Shinwari, Z.K., 2023. *Role of Dof Transcription Factors under Abiotic Stresses.* Proc. Pak. Acad. Sci. (B): Life Environ. Sci., 60(03): 367-376.

Rensing, S.A., Lang, D., Zimmer, A.D., 2008. *The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants*. Sci., 319(5859): 64-69.

Sekito, T., Fujiki, Y., Ohsumi, Y., Kakinuma, Y., 2008. *Novel families of vacuolar amino acid transporters*. IUBMB Life, 60(8): 519-525.

Saleem, H., Jan, S.A., Shinwari, Z.K., Shinwari, A.K., Israr, A., Iqbal, A., Hussain, Z., 2025. *Adapting agriculture: strategic crop responses and mitigation to climate shifts: an updated review*. Pak. J. Bot., 57(4): 1581-1585.

Shinwari, Z.K., Jan, S.A., Nakashima, K., Yamaguchi-Shinozaki, K., 2020. *Genetic engineering approaches to understanding drought tolerance in plants*. Plant Biotechnol. Rep., 14(2): 151-162.

Taylor, M.R., Reinders, A., Ward, J.M., 2015. *Transport function of rice amino acid permeases (AAPs)*. Plant Cell Physiol., 56(7): 1355-1363.

Vij, S., and Tyagi, A.K., 2006. *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*. Mol. Genet. Genomics, 276(6): 565-575.

Xia, L., Xu, Z., Wang, Y., 2022. *Genome-wide identification and expression analysis of the 14-3-3 gene family in mango (Mangifera indica L.).* Int. J. Mol. Sci., 23(3): 1593.

Yao, X., Nie, J., Bai, R., Sui, X., 2020. *Amino acid transporters in plants: identification and function*. Plants, 9(8): 972.

Zhang, J., Wang, L., Liu, Y., 2015. *Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (Glycine max) germplasm.* BMC Genomics, 16(1): 9.

Zhou, T., Zhao, Y., Zhang, Y., 2020. *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 Biol., 20(1): 10.