J. Sylhet Agril. Univ. 12(1): 16-37, 2025 *https://doi.org/10.3329/jsau.v12i1.85894*

ISSN: 2308-1597 (Print) 3105-4080 (Online)

Journal of the Sylhet Agricultural University

Journal home page: http://www.jsau.sau.ac.bd

Research Article

EVOLUTIONARY HISTORY AND FUNCTIONALITY ANALYSIS OF AROGENATE DEHYDRATASES OF *ARABIDOPSIS THALIANA*: A BIOINFORMATIC APPROACH

Mst Rubaiat Nazneen Akhand[1,2,*], Hafsa Akter[2], Jamil Ahmed[1,2] and Nusrat Jahan[2]

[1]*Department of Biochemistry and Chemistry, Sylhet Agricultural University, Sylhet-3100*

[2]*Faculty of Biotechnology and Genetic Engineering, Sylhet Agricultural University, Sylhet-3100*

Article info

Article history

Received : 18.11.2024

Accepted : 15.05.2025

Published : 30.06.2025

Keywords

ADT, Arabidopsis, Phenylalanine, Aromatic amino acid, Evolution

Corresponding author

Mst Rubaiat Nazneen Akhand

E-mail: rubaiatmna.biochem@sau.ac.bd

Abstract

Phenylalanine is an important plant derived aromatic amino acid that acts as an integral precursor of various specialized metabolites. Shikimate pathway is considered as the fundamental route for phenylalanine biosynthesis in plants. The crucial last step of this pathway is catalyzed by arogenate dehydratase (ADT) to synthesize phenylalanine from arogenate. *Arabidopsis thaliana* has six copies of ADT that catalyze phenylalanine production and initiate the downstream pathway to produce a large number of secondary metabolites. However, how do the copies of ADT have been evolved in different trajectories of land plant is relatively unknown. It is also unclear whether all the copies of *Arabidopsis* ADT have evolved at the same rate during the course of evolution. In addition, it is unknown how the multiple copies of *Arabidopsis* ADT contribute towards functional complexity. Our phylogenetic study revealed that all the domains of ADT had evolved before the land plant evolution as it was found in red algae, green algae and in fungus, indicating the presence of functional ADTs in all the spectrum of plants, algae and fungi. ADTs of *Arabidopsis* formed four different clusters which might suggest to their functional differences. Expression analysis revealed the dominative role of ADT2 in seed development, while the rest of the ADTs showed spatiotemporal expression and contributed to other developmental functions of *Arabidopsis*.

Introduction

Arogenate dehydratase (ADT) catalyzes the production of phenylalanine from arogenate. Among 20 standard amino acids phenylalanine, tryptophan and tyrosine are the aromatic amino acids. These aromatic amino acids are regarded as essential amino acids for humans and most of the animals in the animal kingdom as these cannot be synthesized by them. These amino acids are the integral components of structural and functional proteins in plants. In addition to being essential components of various plant protein, phenylalanine, tyrosine and tryptophane are the source of numerous aromatic secondary metabolites such as phenylpropanoids (Dixon & Paiva, 1995), lignin (Liu et al., 2018) and indole alkaloids (Kutchan, 1995). They, therefore, play significant roles in both development and defense systems of land plants.

Cite This Articale

Akhand MRN, Akter H, Ahmed J and Jahan N. 2025. Evolutionary History and Functionality Analysis of Arogenate Dehydratases of *Arabidopsis Thaliana*: A Bioinformatic Approach. J. Sylhet Agril. Univ. 12(1): 16-37, 2025. https://doi.org/10.3329/jsau.v12i1.85894

The biochemistry for the biosynthesis of these essential amino acids is only limited to bacteria, fungi and plants (Maeda et al., 2010). Among these essential aromatic amino acids, phenylalanine acts as a major precursor of the phenylpropanoid pathway and generates enormous array of secondary metabolites. Thus, the biosynthesis of phenylalanine is an exceptionally important pathway in plants. The absence of a phenylalanine biosynthetic pathway in animals makes this pathway an attractive target for herbicide design (El-Azaz et al., 2016).

Knowledge of the biosynthetic pathway for phenylalanine provides a basis for manipulation of their levels or those of aromatic secondary metabolites in land plants. The immediate precursor of phenylalanine is arogenate that comes from the shikimic acid pathway. The conversion of arogenate to phenylalanine is catalyzed by the ADT that belongs to the lyases enzyme family, specifically the hydro-lyases, which cleave carbon-oxygen bonds. L-arogenate eliminates carboxyl and hydroxide groups attached to the 2, 5-cyclohexene ring and leaves as carbon dioxide and water. As a result, 2, 5-cyclohexene ring is converted into a phenyl ring and L-phenylalanine amino acid is formed (Corea et al., 2012).

Being the part of shikimate acid pathway, ADT biosynthesis occurs in plastid (Bross et al., 2017). The shikimate pathway in plastids links central carbon metabolism to the biosynthesis of aromatic amino acids phenylalanine, tyrosine and tryptophan. Phosphoenolpyruvate and erythrose 4-phosphate are the two intermediary products of glycolysis and hexose monophosphate shunt. Those two products are converted into chorismate that act as a direct precursor for the biosynthesis of aromatic amino acids and other compounds, such as vitamins K1 and B9 (Tzin & Galili, 2010). Alternatively, chorismate is used by anthranilate synthase, the first enzyme in the biosynthetic pathway of Trp, or by chorismate mutase to produce prephenate for phenylalanine and Tyr biosynthesis (Maeda et al., 2010).

Two alternative routes of phenylalanine biosynthesis have been proposed: (I) In arogenate pathway, prephenate is trans-aminated by prephenate aminotransferase (PAT) enzyme to produce arogenate which is then decarboxylated and dehydrated via ADT to produce phenylalanine (Bonner & Jensen, 1987) and (II) in phenylpyruvate pathway, the decarboxylation and dehydration of prephenate by prephenate dehydratase (PDT) enzyme results in the production of phenyl pyruvate. Subsequently, phenyl pyruvate aminotransferase enzyme converts phenylpyruvate to phenylalanine (El-Azaz et al., 2016).

The arogenate pathway has been proposed to be the main phenylalanine biosynthesis route in plants (Maeda et al., 2011). However, the observed PDT activity in etiolated *Arabidopsis thaliana* seedlings (El-Azaz et al., 2016) and recently in *Petunia hybrida* petals (Yoo et al., 2013) specifies the presence of a functional phenylpyruvate pathway in plants. Unlike ADT, PDT is not ubiquitously present in all the plant kingdoms (Maeda et al., 2010,Maeda et al., 2011). In the conventional pathway of phenylalanine biosynthesis, ADT catalyzes the crucial last step of shikimate pathway and thereby initiating the phenylpropanoid pathway. Therefore, ADTs are the enzyme of interest in this study.

ADTs have been identified in every higher plant analyzed to date and most of them encode multiple versions of ADT isoenzymes. In *Arabidopsis,* six ADTs have been found including ADT1-6 (Cho et al., 2007). All six ADTs in *Arabidopsis* possess identical protein structures and amino acid composition (Cho et al., 2007). All *Arabidopsis* ADT proteins are localized within or around chloroplasts where phenylalanine is synthesized (Bross et al., 2017, Rippert et al., 2009). Different ADTs of *Arabidopsis* get activated at different developmental stages at a specific time. Following their activation, they target specific anatomical parts within the plant to exert their function.

The anatomical parts of stems, leaves, roots, flowers, siliques, and seeds have been found to express *Arabidopsis* ADT1 and ADT2 (Cho et al., 2007). Of these, ADT2 is most widely expressed in leaves and seeds and a "housekeeping" role of ADT2 in phenylalanine biosynthesis has also been suggested earlier (Rippert et al., 2009). Compared to other anatomical parts, *Arabidopsis* ADT4 and ADT5 are strongly expressed in stems and roots, while the remaining isoenzymes usually have much lower levels of expression (Corea et al., 2012).

Arabidopsis ADT1 and ADT2 expressions are predominantly associated with genes involved in fundamental cellular functions, such as transcription, translation, cell division, and biosynthesis of nucleic or amino acids. Numerous shikimates, phenylpropanoid, and aromatic amino acid biosynthesis gene are provisionally co-regulated with *Arabidopsis* ADT3, ADT4, ADT5, and ADT6 (Corea et al., 2012). However, it is relatively unknown whether the evolutionary divergence of these enzymes has directed the multiple functionalities toward different biosynthetic pathways. Again, no developmental stage-based expression patterns have been found for *Arabidopsis* ADTs in literature. As like spatiotemporal expression of *Arabidopsis* ADTs, there is evidence of more specialized function of each ADT in *Arabidopsis*, even though *Arabidopsis* ADTs are similar in sequence and all six of them can decarboxylate and dehydrate arogenate.

To understand how different ADTs of *Arabidopsis* have evolved throughout the land plant evolution, we first focused on the water-to-land plant evolution. The water-to-land transition (plant terrestrialisation) that took place millions of years back (Yoon et al., 2004), is one of the most remarkable challenges mastered by plants. This habitat changes necessitated molecular adaptations to address a wide range of new stresses. Plant terrestrialisation has triggered drastic change as the Earth's atmosphere and soil cover has been transformed and brought the earth into existence. Eukaryotic photosynthetic species has evolved more than 1.5 billion years ago (Yoon et al., 2004) and divided into three groups: Glaucophyta, Rhodophyta and Viridiplantae. The occurrence of gene duplication event in land plants results in multiple copy number of genes. In every higher plant, ADT encodes multiple ADT isoenzymes. For example, *Lotus japonicus* has two, *Petunia hybrida* (Maeda et al., 2010), *Theobroma cacao*, *Arabidopsis lyrate*, *Populus trichocarpa*, *Fragaria vesca* and *Vitis vinifera* has three, *Pinus taeda* contains nine (El-Azaz et al., 2016) ADT isoforms. Increased functional complexity is associated with the presence of several copy numbers of protein. Therefore, the copy number variation (CNV) might also render functional complexity of the *Arabidopsis* ADTs, which are distributed on a different pattern across specific tissue and organ of different cell types (Corea et al., 2012).

An overall outline regarding the functional complexity of *Arabidopsis* ADTs is absent to date. Any specific knowledge about where and when *Arabidopsis* ADTs show spatiotemporal expression are relatively unknown. Without complete evolutionary study, it is not possible to extrapolate when this tissue dependent, growth phase dependent functional expression of different *Arabidopsis* ADTs outsets. Through our study, we aimed to explore if all the domains architecture and conserved regions of different isoforms of *Arabidopsis* ADT remain same, to understand whether all the copies of *Arabidopsis* ADT have evolved at the same rate throughout the evolution and to characterize the expression pattern of each ADT in some specific tissues at different development stages of *Arabidopsis thaliana.* To cope with the objectives of the study, we opted for the use of the model organism *Arabidopsis*. *Arabidopsis* ADT protein family comprises six members: ADT1 to ADT6 and are involved in phenylalanine biosynthesis.

Two approaches were used to accomplish these objectives. In the first approach, physicochemical properties, domains, motifs and structural variation were analyzed to have a deep understanding about *Arabidopsis* ADTs. Multiple sequence alignment was carried out and a phylogenetic tree of selected species was constructed to explore any pattern of evolution. In the second approach, expression variation of *Arabidopsis* ADTs was analyzed to investigate which ADT expresses most at which specific stage and tissue. Expression variation was analyzed in terms of the stages of development, the anatomy of specific tissues and co-expression.

Material and methods

Sequence retrieval

All the six ADTs of *Arabidopsis thaliana* were retrieved from The Arabidopsis Information Resource (TAIR; https://www.arabidopsis.org/) (Berardini et al., 2015). In the next step, Phytozome Version 12 (Goodstein et al., 2012) was used. In phytozome, some species from different classes including

Angiosperms, Gymnosperms, Pteridophytes (lycophytes), Bryophytes (mosses and liverworts), Chlorophytes and Rhodophytes were used as target and retrieved *Arabidopsis* ADT sequences were used as query sequence. Basic Local Alignment Search Tool (BLAST) was performed in phytozome (https://phytozome-next.jgi.doe.gov/). However, some important clades of plants were not found in phytozome and those species belonged to the clades Fern, Hornworts and Charophytes. ADT sequence from these clades were attempted to retrieve from UniProt Knowledge (https://www.uniprot.org/help/uniprotkb) (Morgat et al., 2020).

Physicochemical properties, domain and motif analysis

To calculate the various physical and chemical properties, ProtParam (http://www.expasy.org/tools/protparam.html) was opened from ExPASY server. To analyze the domain of the retrieved fasta sequences InterPro (https://www.ebi.ac.uk/interpro/) was used. Motif analysis was done by ScanProsite (https://prosite.expasy.org/scanprosite/).

Multiple sequence alignment and visualization

Multiple sequence Alignment (MSA) using Fast Fourier Transform version 7.0 is one of the most accurate and rapid programs (MAFFT; https://mafft.cbrc.jp/alignment/software/) (Katoh & Standley, 2013). This program was run in the Linux platform using red hat Linux four core servers. The E-INS-i (-- genafpair) algorithm was used for MSA. In this algorithm, all pairwise alignments were computed with a local algorithm with the generalized affine gap cost. For visualization, alignment file was opened in Jalview version 2.10.5 (https://www.jalview.org/) (Waterhouse et al., 2009).

Secondary and tertiary structure prediction

Easy Sequencing in PostScript (ESPript; http://espript.ibcp.fr/ESPript/ESPript/) version 3.0 (Robert & Gouet, 2014) was used for the generation of the secondary structures of the ADT sequences. Iterative Threading ASSEmbly Refinement (I-TASSER; https://zhanglab.ccmb.med.umich.edu/I-TASSER/) version 5.1 (Yang & Zhang, 2015) was used for predicting three-dimensional structure models of protein molecules from amino acid sequences. It involves the detection of the structural templates from protein data bank by using a technique called fold recognition (or threading). For better visualization, these models were uploaded in PyMOL version 2.4 (https://pymol.org/2/) and modified by highlighting the secondary structures.

Model assessment

RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) is an offshoot of RAPPER which generates a Ramachandran plot using data derived by the Richardsons and coworkers (Lovel et al 2022). For model assessment purposes, PDB (Protein databank) file formats were used in RAMPAGE to get the best models for all six *Arabidopsis* ADTs.

Phylogenetic tree construction and visualization

Before constructing the phylogenetic tree, Phyutility program version 2.7.1 (https://github.com/blackrim/phyutility) was used to trim the unaligned residues (Smith & Dunn, 2008). Alignment position with more than 50% gaps were pruned. After that, phylogenetic tree was drawn using the RAxML software (https://raxml-ng.vital-it.ch/#/) version 1.0.0 (Stamatakis, 2014) using -auto option and 100 bootstrap values. Following tree construction, visualization was done by Interactive Tree of Life (iTOL; https://itol.embl.de/login.cgi) version 5.7 (Letunic & Bork, 2019).

Expression, development and anatomy analysis of arabidopsis ADTs

GENEVESTIGATOR (https://genevestigator.com/) utilizes user-friendly visualization tools to facilitate the analysis of deeply curated bulk tissue and single-cell transcriptomic data retrieved from public repositories

(Hruz et al., 2008). For analysis of the expressions of *Arabidopsis* ADTs gene at various development stages (germinated seed, seedling, young rosette, developed rosette, bolting, young flower, developed flower, flowers and siliques, mature siliques and senescence), Affymetrix Arabidopsis ATH1 Genome array was selected which includes 10562 samples information of *Arabidopsis thaliana*. From there 127 anatomical parts, targeted 18 anatomical parts (seedling, cotyledon, hypocotyl, flower, carpel, petal, sepal, seed, embryo, replum, shoot, stem, leaf, adult leaf, senescent leaf, roots, root tip and lateral root) were nominated. Finally, co-expression analysis using STRING (https://string-db.org/cgi/about) was performed.

Results

Acquisition of sequences to sketch evolutionary history of ADT

Twenty-two, five, one, four, seventeen and three sequences were retrieved from phytozome for the clades of angiosperms, gymnosperms, pteridophytes, bryophytes, chlorophytes and rhodophytes, respectively. Two charophytes and one fungal ADT sequence were also retrieved from UniProtKB. A total of sixty-one sequences (Supplementary Table 1) of forty-five species from different clades have been selected to carry out evolutionary study. No hits were found to be present for ADT of fern and hornworts in UniProtKB. For further confirmation, those classes were again searched in NCBI. ADT was observed for rest of plant groups with an exception in fern and hornworts ADT as no hits were detected for them.

Table 1. Domain information including the InterPro IDs and the position of the six ADT present in *Arabidopsis thaliana*

Protein	Length of Protein in Amino Acid	Prephenate dehydratase	ACT domain	Prephenate dehydratase, conserved site
		IPR001086	**IPR002912**	**IPR018528**
ADT1	392	107-284	296-387	258-329
ADT2	381	100-277	289-375	251-322
ADT3	424	122-301	313-404	275-346
ADT4	424	126-305	319-410	279-352
ADT5	425	127-306	320-411	280-353
ADT6	413	117-296	308-399	270-341

Physicochemical properties of Arabidopsis ADTs

The physicochemical characteristics has been shown in Figure 1. For *Arabidopsis* ADT1, molecular weight was recorded to be 43604.91 Da with a theoretical pI value of 6.83. Leu was present at a highest percentage of 10.7% in *Arabidopsis* ADT1. This protein had an instability index of 48.3, indicating the unstable nature of this protein. The estimated aliphatic index and GRAVY values for *Arabidopsis* ADT1 were 93.83 and -0.122 respectively.

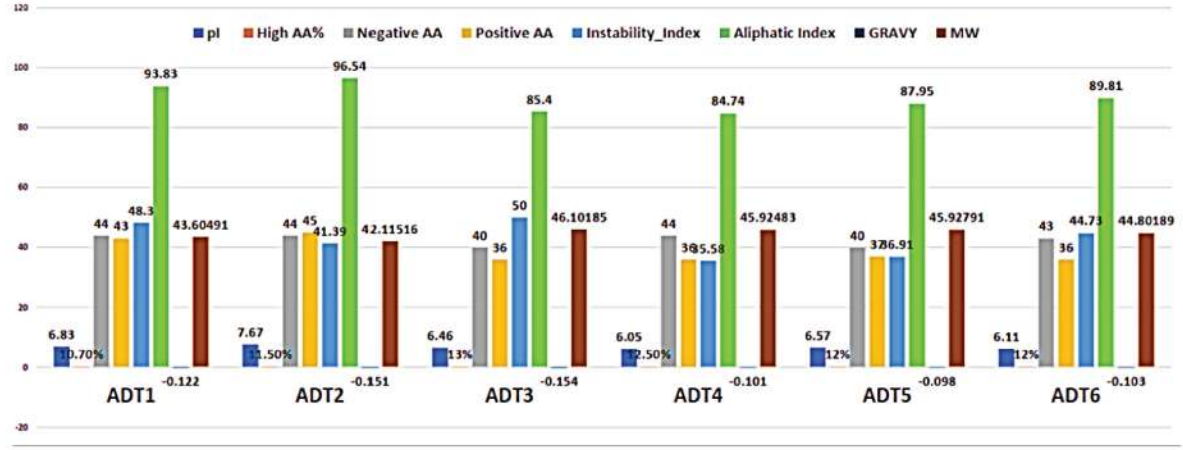


Figure 1. Physicochemical properties of *Arabidopsis* ADTs

It was found that *Arabidopsis* ADT2 had the highest pI value and aliphatic index of 7.67 and 96.54 respectively where Leu was present at a highest percentage than other amino acids. The highest molecular weight of 46101.85 Da was observed in *Arabidopsis* ADT3 among all other *Arabidopsis* ADTs and Ser was found to be present at highest percentage. Lowest GRAVY was observed for *Arabidopsis* ADT3 which indicates the possibility of better interaction with water. An instability index of 35.58 and 36.91, indicating *Arabidopsis* ADT4 and *Arabidopsis* ADT5 to be stable among all other ADTs. In *Arabidopsis* ADT6, Ala was found to be present at the highest percentage that is similar to *Arabidopsis* ADT5 and ADT4.

Functional domains and motifs of arabidopsis ADTs

Arabidopsis ADTs shared three common domains, Prephenate dehydratase (PDT; InterPro accession ID, IPR001086), ACT domain (InterPro accession ID, IPR002912) and Prephenate dehydratase, conserved site (PDTcs; InterPro accession ID, IPR018528). From InterPro result, these domains and conserved sites were also found to be present in all clades of land plants, green algae, red algae and fungus. But the domains (Figure 2) and conserved site differed in their sequence length in *Arabidopsis* ADTs. The length of the proteins and starting and ending position of each domains of *Arabidopsis* ADTs has been represented in Table 1. In our study, two conserved motifs -ESRP and TRF were found.

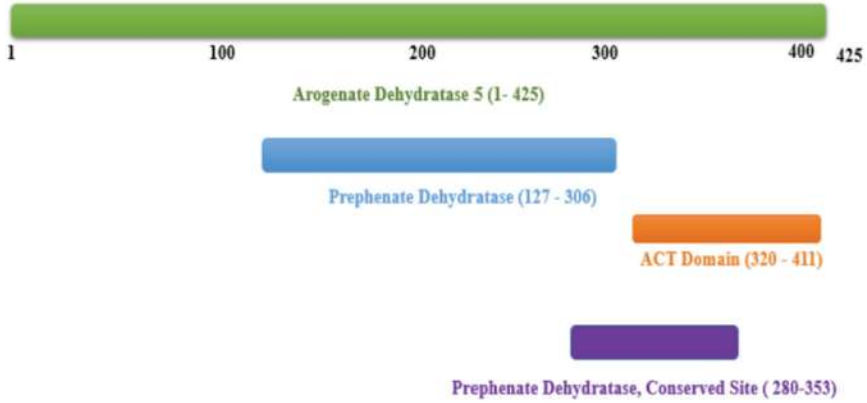


Figure 2. Domain structure of *Arabidopsis* ADTs. Green color represents the length of Arogenate dehydratase 5 protein (ADT5) that spans 1st to 425th amino acid (aa). Blue domain represents prephenate dehydratase (127aa-306aa), orange color represents ACT domain (320aa-411aa) and purple domain represent the conserved site of prephenate dehydratase (280aa-353aa).

Tertiary structure of arabidopsis ADTs

In I-TASSER, each model's confidence was quantitatively evaluated by C-score which was calculated based on the significance of threading template alignments and the convergence parameters of the simulations of the structure assembly. Table-2 represents the quality of the tertiary structures in forms of C-score, TM-score and RMSD. The tertiary structure reveals that Arabidopsis ADTs consist of total fourteen β strands and nine or ten α helices of varying lengths. The PDT domain of Arabidopsis ADTs was composed of ten β strands, seven α helices with one or two strict β turns. Similarly, ACT domain predicted to be made up of four β strands and two α helices with an exception in Arabidopsis ADT2 which had three α helices. Strict β-turns in ACT domain were found to be present only in Arabidopsis ADT1, ADT4 and ADT5. In a nutshell, all six Arabidopsis ADT resulted with a less variation in secondary structural conformation (Figure 3).

Table 2. Overview of tertiary models of *Arabidopsis* ADTs (The C-score is a confidence score used by I-TASSER to estimate the quality of predicted models. RMSD is the average distance of all residue pairs in two structures. The RMSD problem, which is sensitive to local error, is resolved using the TM-score.)

Models	C-Score	Expected TM Score	Expected RMSD	Custer Density
ADT1	-2.36	0.44 ± 0.14	12.4 ± 4.3	0.0251
ADT2	-2.57	0.44 ± 0.14	12.4 ± 4.3	0.0240
ADT3	-2.34	0.42 ± 0.14	12.2 ± 4.1	0.0205
ADT4	-2.48	0.43 ± 0.14	12.9 ± 4.2	0.0230
ADT5	-2.61	0.41 ± 0.14	13.3 ± 4.1	0.0202
ADT6	-2.24	0.45 ± 0.15	12.2 ± 4.4	0.0287

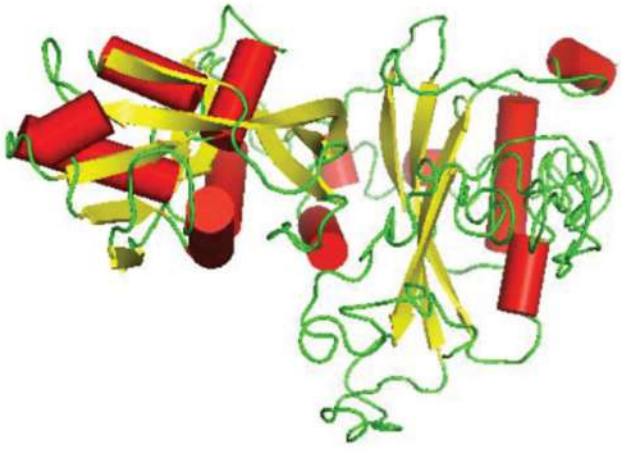


Figure 3. Tertiary Structure of *Arabidopsis* ADT [Red color represents for α helices, Yellow color for β strands and Green color for loop]

Validation of Arabidopsis ADT models

In Ramachandran plot, the horizontal axis displays φ values, while the vertical one shows ψ values. The counting starts from -180 in the left-hand corner and extend to +180 for both the vertical and horizontal axes. The angle for an amino acid is represented by each dot on the plot. The regions in Ramachandran Plot are:

- Quadrant I indicate the region where some of the conformations are permitted. Rare left-handed alpha helices lie in this region.
- Quadrant II implies the biggest area in the graph. The most favorable conformations of atoms are found in this region. For beta strands, the sterically permitted conformations are displayed in this region.
- Quadrant III shows the next biggest region in the graph where right-handed alpha helices lie.
- Quadrant IV has almost no defined region. Due to steric clash, this conformation (ψ around -180 to 0 degrees, φ around 0-180 degrees) is disfavored (Pal & Chakrabarti, 2002).

In case of *Arabidopsis* ADT3, no amino acids were observed in disfavored region (0.0%) and number of residues presented in favored region tended to be high. So, this was regarded as a good quality model. On the other hand, *Arabidopsis* ADT4 was considered as a poor model since having a relatively lower number of residues in favored region and a higher proportion of amino acids in disfavored region compared to other models (Table 3).

Table 3. Overview of the quality of arabidopsis ADTs tertiary models

Models	No. of residues of favored region (~98% expected)	No. of residues in allowed region (~2% expected)	No. of residues of outlier region	Quality of Model
ADT1	525 (92.8%)	23 (4.1%)	18 (3.2%)	MODERATE
ADT2	525 (92.8%)	23 (4.1%)	18 (3.2%)	MODERATE
ADT3	276 (92.6%)	22 (7.4%)	0 (0.0%)	GOOD
ADT4	269 (88.5%)	28 (9.2%)	7 (2.3%)	POOR
ADT5	140 (92.1%)	8 (5.3%)	4 (3.6%)	MODERATE
ADT6	279 (91.8%)	13 (4.3%)	12 (3.9%)	MODERATE

Identification of conserved regions and active sites

The conserved regions (Figure 4) were positioned 162-172, 174, 176-179, 181,182, 189, 194, 195, 204, 206, 207, 209, 230, 231, 232 and 233. The amino acids that were localized in these positions were Gly, Ser, Iso, His, Arg, Asn, Tyr, Asp, Leu, Leu, Leu, His, Lys, His, Iso, Val, Asp, Val, Cys, Pro, Gly, Val, Ser, His, Gln, Glu, Thr, Ala and Ala respectively. The highly conserved regions were found to be fully covered by PDT domain.

The area where particular substrates bind to the enzyme and catalyze the chemical reaction is known as the active site of an enzyme. The active enzyme site is formed by the substrate binding site and the catalytic site. The catalytic site is placed next to the binding site and catalysis is carrying out. Catalytic sites consist of approximately two to four amino acids that are involved in the catalysis. The amino acids that are catalytically important are: Cys, His, Asp, Lys, Ser, Thr and Tyr (Holliday et al., 2009). In *Arabidopsis*, the positions of active sites in all six ADTs were obtained from I-TASSER (Figure 5).

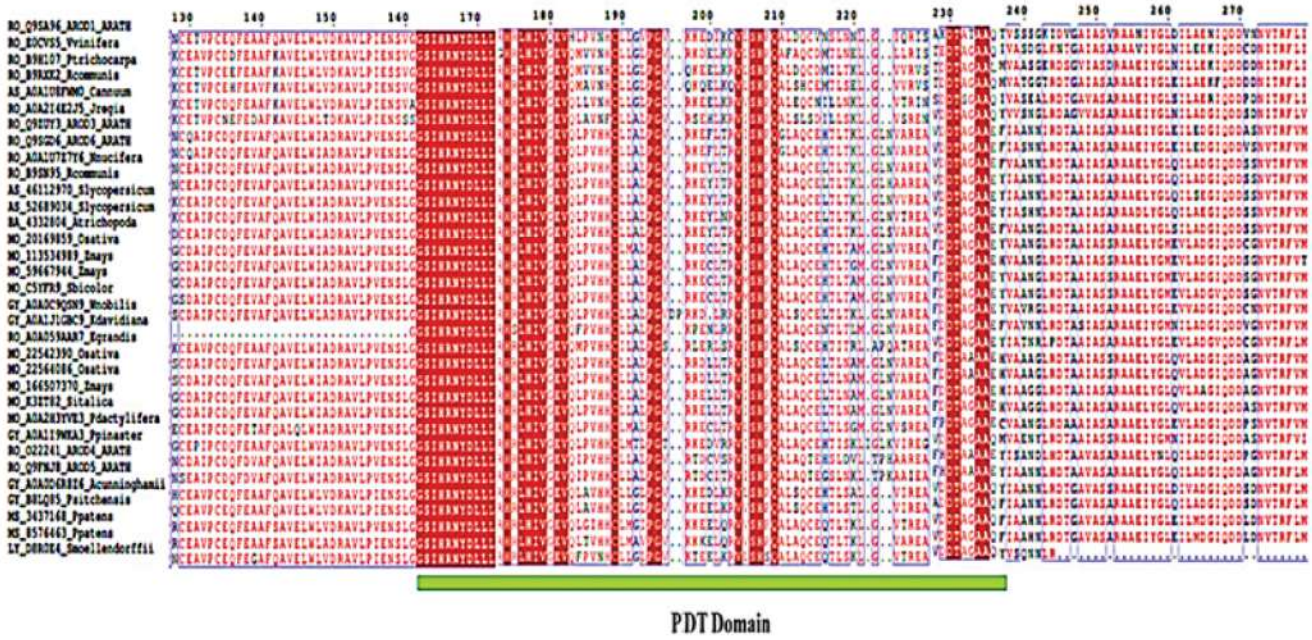


Figure 4. Multiple sequence alignment showing the highly conserved region of PDT domain of ADTs of different clades of land plants including model plant *Arabidospsis.*

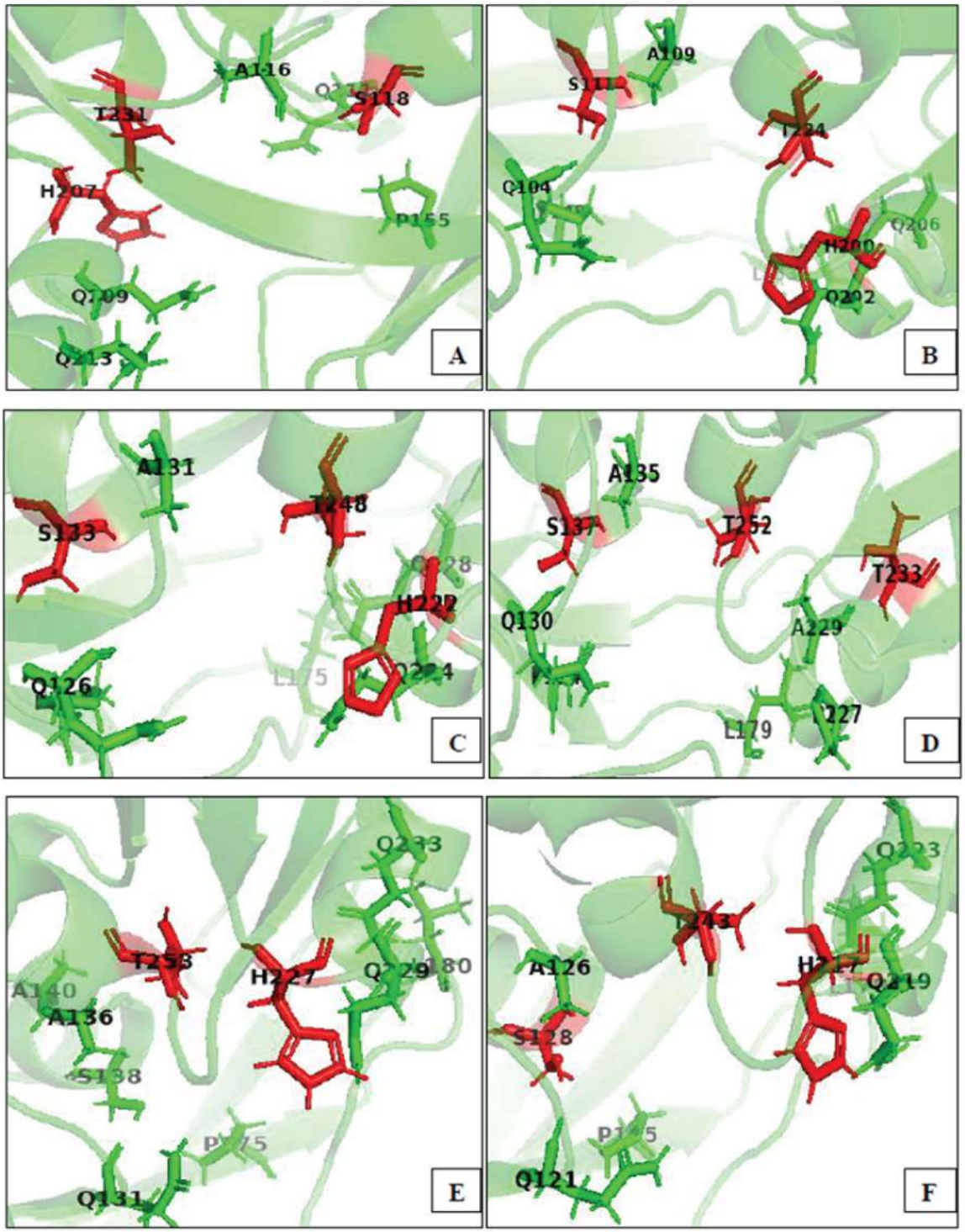
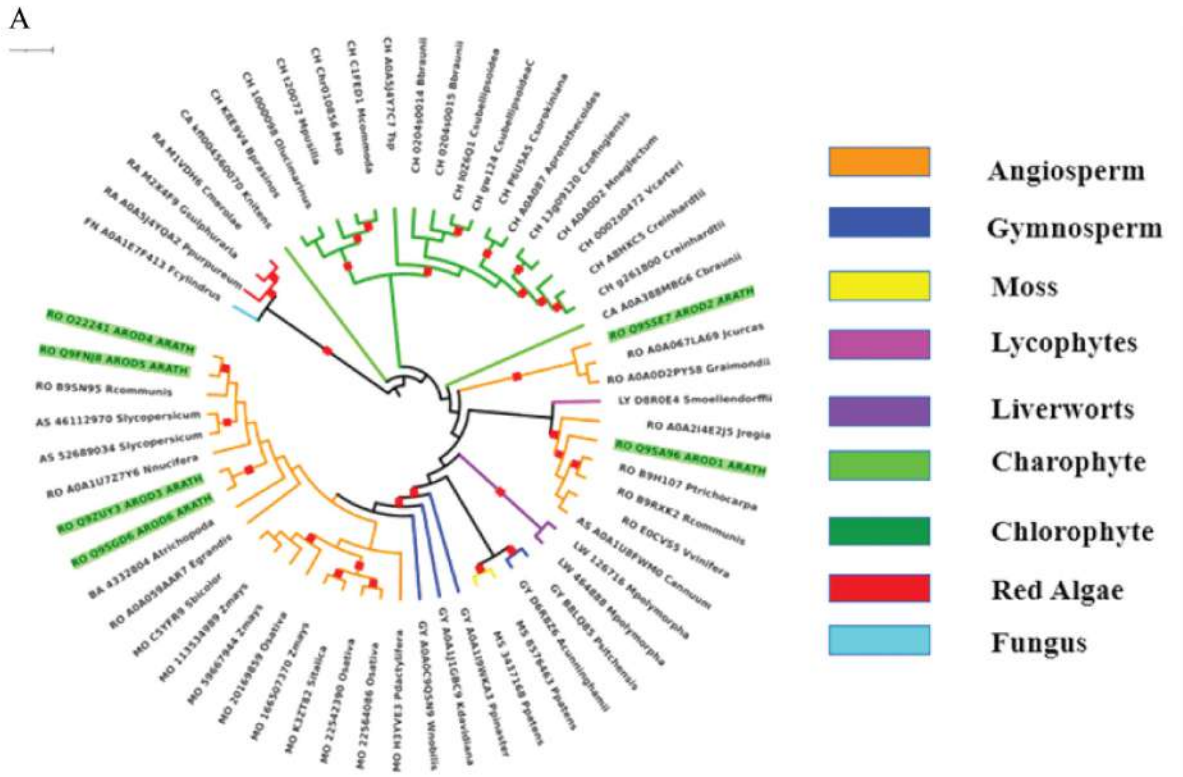


Figure 5. Active Sites of *Arabidopsis* ADTs [A, B, C, D, E and F indicates *Arabidopsis* ADT1, ADT2, ADT3, ADT4, ADT 5 and ADT6 respectively. Catalytically important sites were colored as red with amino acid position]

Evolutionary pattern of Arabidopsis ADTs

Phylogenetic analysis can provide detailed evolutionary insights into the diversification of ADT enzymes in our selected species. Our phylogenetic tree (Figure 6A) was mid-point rooted and constructed with one hundred bootstrap confidences. Within the tree, the fungal sequence was found to make an out group, meaning that all members of the group of interest were more closely related to each other than they were to fungus. Hence fungus was distantly related to the clades of land plants. Following that, ADT enzyme has evolved in red algae and acted as the earliest ancestor of land plants (Figure 6B). As the charophyte sequences were found to split before and after the chlorophyte clade, the charophyte ADT enzyme had been formed before the chlorophytes; however bootstrap value did not support this hypothesis. After red algae, ADT enzyme was evolved in chlorophytes and the chlorophytes were found to form a single cluster in our tree. In many species of chlorophytes, copy number variation (CNV) was observed which hinted the occurrence of gene duplication event in chlorophytes. At this stage the tree was separated into two groups. In one group, lycophytes and angiosperms were found to be together in a clade. While liverworts, moss, gymnosperm and angiosperm were together in another clade. CNV was found in moss liverworts and angiosperms, where the highest level of CNV was observed in angiosperms. Two important clades of hornworts and fern were missing in our constructed tree.

Looking at the phylogenetic tree, ADT protein was evolved in angiosperms before the evolution of angiospermic land plants and this could be concluded with high confidence as high bootstrap value was observed. *Arabidopsis* ADTs were found to form four different clusters. In one cluster, *Arabidopsis* ADT1 was present while in another *Arabidopsis* ADT2 was found. *Arabidopsis* ADT3 and ADT6 were found to form a cluster within the tree. Moreover, *Arabidopsis* ADT4 and ADT5 also formed another cluster.



B

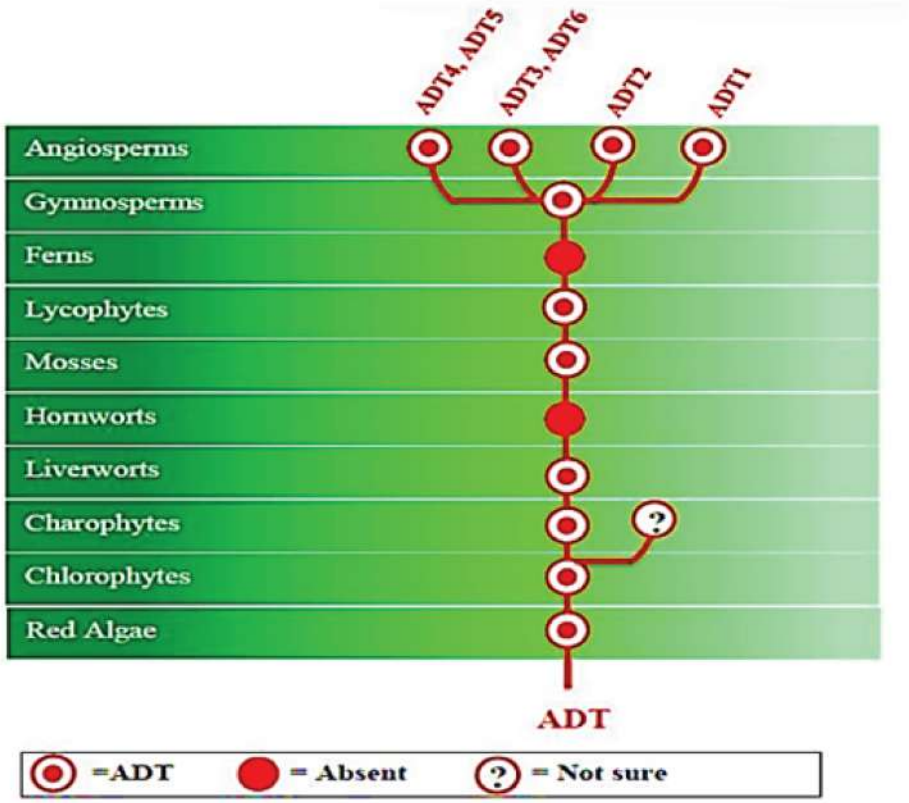


Figure 6. Phylogenetic Tree of ADT Protein Family (A) and Evolutionary Origin of the ADT Protein Family Members (B) [Red squares on the branches of the phylogenetic tree represents bootstrap value (>80), *Arabidopsis* ADTs are highlighted with green background color, Branch color indicates the classes of different species

Developmental expression patterns of Arabidopsis ADTs

In our study, we found that *Arabidopsis* ADT1 was expressed the highest when the rosettes were fully developed and flowers were at younger stage. In other developmental stages, moderate expression potential was found where in seeds it was 0%. In the germinated seed, *Arabidopsis* ADT2 had a higher expression potential of nearly 100%, compared to the rest of other *Arabidopsis* ADTs. Moderate level of expression potential was found in other stages. Whereas, *Arabidopsis* ADT3 had a moderate level of expression potential during all developmental stages except seeds. In seed, it had an expression potential of nearly 0%. During all developmental stages, *Arabidopsis* ADT4 exhibited very low level of expression potential especially in seeds. *Arabidopsis* ADT5 showed moderate level of expression potential when the rosettes were young. Apart from this, low level of expression potential was observed in other developmental stages. In developed rosettes, bolting and seeds *Arabidopsis* ADT6 had a low level of expression potential while in other stages, it was moderate (Figure 7).

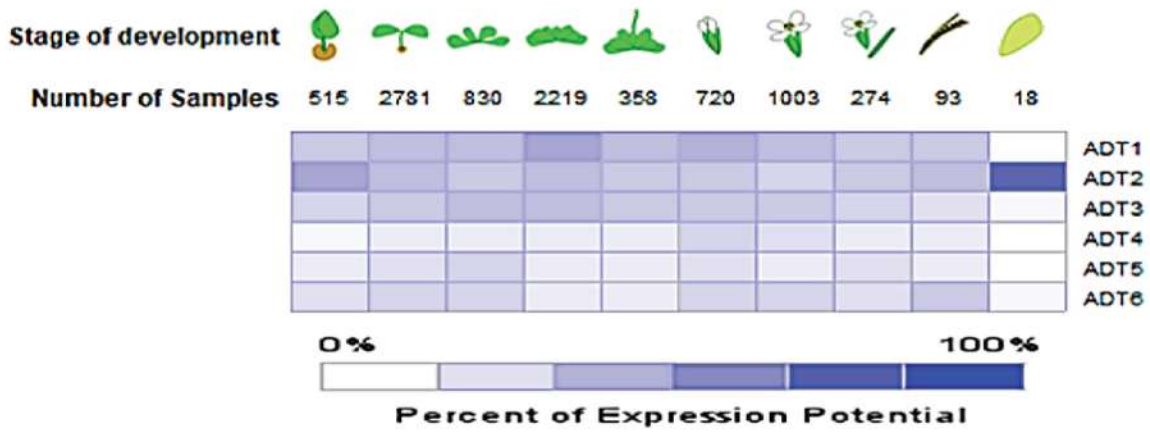


Figure 7. Developmental Expression Patterns of *Arabidopsis* ADTs [Expression potential of 100% to 0% were colored as blue to white]

Anatomical expression patterns of Arabidopsis ADTs

The result of expression potential of *Arabidopsis* ADTs in different anatomical parts of *Arabidopsis* was very interesting (Figure 8). Following the Manhattan distance, *Arabidopsis* ADT2 showed complete distinction than other *Arabidopsis* ADTs. *Arabidopsis* ADT1 had the highest level of expression potential in cotyledon, leaf and adult leaf while lowest level of expression was observed in senescent leaf. In other anatomical parts, *Arabidopsis* ADT1 had moderate expression level. In case of seed and root tip, *Arabidopsis* ADT2 showed expression variation. On the other hand, similar expression potential was observed in the rest of the anatomical parts.

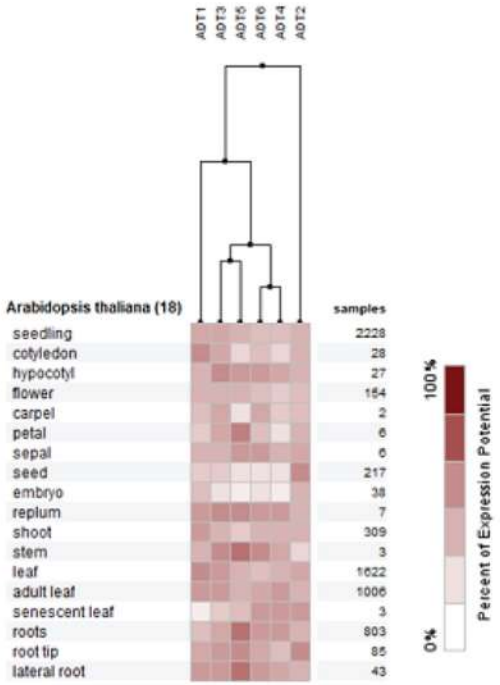


Figure 8. Anatomical Expression Patterns of *Arabidopsis* ADTs [Expression potential of 100% to 0% were colored as brick red to white]

Arabidopsis ADT3 was found to be more expressive in the parts of hypocotyl, leaf, adult leaf, cotyledon, stem and replum rather than other anatomical parts. Lower expression potential was recorded for ADT4 of *Arabidopsis* in cotyledon, petal, seed and embryo, whereas it had more or less similar expression potential in rest of the anatomical parts. Higher expression potential was observed for *Arabidopsis* ADT5 in petals, root, lateral root, and stem. However, it had a lower label of expression potential in seed, embryo, cotyledon and carpel. Expression potential was recorded to be average in other parts of the plant. Lastly, *Arabidopsis* ADT6 displayed lower expression potential in seed and embryo and a similar expression potential was observed in other parts of the plants.

Co-Expression patterns of Arabidopsis ADTs

In our study, all *Arabidopsis* ADTs were found to be co-expressed with nearly similar type of genes (Figure 9). Genes that were found to be uniformly co-expressed in all *Arabidopsis* ADTs includes: Bifunctional aspartate aminotransferase (AAT), Chorismate mutase 1 (CM1), Chorismate mutase 2 (CM2), Chorismate mutase 3 (CM3), Histidinol-phosphate aminotransferase (HPA1), Arogenate dehydrogenase 1 (AT5G34930/ADH1), Arogenate dehydrogenase 2 (AT1G15710/ADH2) and Pyridoxal phosphate dependent transferases superfamily (AAS). Aspartyl protease 5 (ASP5) was found to be co-expressed with all *Arabidopsis* ADTs except ADT1. However, some of the genes were found to be co-expressed only for particular *Arabidopsis* ADTs. For instance, in *Arabidopsis* ADT1, Phenylalanine ammonia lyase 1 (PAL1); in *Arabidopsis* ADT2, Chorismate synthase (EMB1144); in *Arabidopsis* ADT3, Guanine nucleotide-binding protein alpha-1 subunit (GPA1); in *Arabidopsis* ADT4, Arogenate dehydratase 5 (ADT5); in *Arabidopsis* ADT5, Arogenate dehydratase (ADT4); in *Arabidopsis* ADT6 Putative H/ACA ribonucleoprotein complex subunit 1 (AT3G03920) were found to be co-expressed.

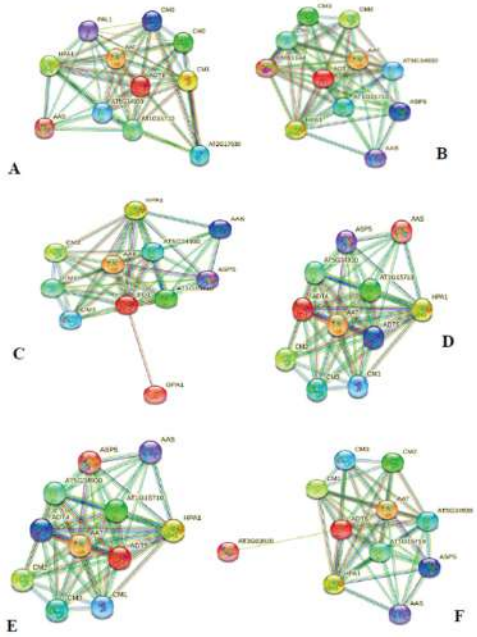


Figure 9. Co-Expression Patterns of Arabidopsis ADTs [A, B, C, D, E and F represents co-expression networks Arabidopsis ADT1, ADT2, ADT3, ADT4, ADT5 and ADT6 respectively. Aspartate aminotransferase; AAT, Chorismate mutase 1; CM1, Chorismate mutase 2; CM2, Chorismate mutase 3; CM3, Histidinol-phosphate aminotransferase; HPA1, Arogenate dehydrogenase 1; AT5G34930, Arogenate dehydrogenase 2; AT1G15710, Pyridoxal phosphate dependent transferases superfamily; AAS, Aspartyl protease 5; ASP5, Phenylalanine ammonia lyase 1; PAL1, Chorismate synthase; EMB1144, Guanine nucleotide-binding protein alpha-1 subunit; GPA1, Arogenate dehydratase 5; ADT5, Arogenate dehydratase; ADT4, Putative H/ACA ribonucleoprotein complex subunit 1; AT3G03920]

The aromatic amino acid phenylalanine, produced in both plants and microorganisms, serves as a building block for proteins and as a pathway intermediate to a wide range of aromatic compounds. In present study, we have reported an outline of the evolutionary history of *Arabidopsis* ADTs and characterized the expression of the members of *Arabidopsis* ADT protein family that are involved in phenylalanine biosynthesis.

4. Discussion

Among all *Arabidopsis* ADTs, ADT2 showed slight variation in some physicochemical properties; molecular weight, pI, and aliphatic index. In case of molecular weight, *Arabidopsis* ADT2 has the lowest value among all other *Arabidopsis* ADTs. Again, *Arabidopsis* ADT2 was found basic in nature where other ADTs were acidic. At pI (7.0), the solubility of any protein is least and mobility in an electro focusing system is zero (Sahay et al., 2020). Therefore, proteins are more stable and compact. The aliphatic index is the relative volume of a protein occupied by aliphatic side chains (Ala, Val, Iso, and Leu) and is regarded as a positive factor for the increase of thermal stability of globular proteins. A higher aliphatic index value indicates the thermal stability of the protein (Sahay et al., 2020). *Arabidopsis* ADT2 with the highest aliphatic index hints more thermal stability than rest of *Arabidopsis* ADTs. These findings indicate possibilities of variation in functionality of *Arabidopsis* ADT2 than rest of the *Arabidopsis* ADTs.

Literature study supported that *Arabidopsis* ADT2 has additional functions in both seeds (El-Azaz et al., 2018) and chloroplast division (Bross et al., 2017). It was also identified that *Arabidopsis* ADT2 has the ability to utilize prephenate as substrate, indicating the presence of PDT activity in this isoform (Bross et al., 2011).

In our study, three domains namely: PDT domain, ACT domain and PDTcs and two motifs including ESRP and TRF were found to be nearly identical in all the selected species. Functionality analysis of domains revealed that in *E. coli* PDTcs is critical for efficient catalysis of the PDT enzyme. The activity of PDT enzyme was determined fundamentally on the regulatory domain (El-Azaz et al., 2016). PDT domain was found not to be directly associated with specific catalytic functionality in the PDT enzyme but may possess some indirect functions. This mutational study was, however, conducted solely on the *E. coli* PDT. Unlike this observation, we found PDT domain having the most conserved region in ADT (Fig. 4) which suggests the necessity of PDT domain for catalytic functionality.

Again, mutational analysis and isothermal titration calorimetry of the *E. coli* P-protein, identified two candidate regulatory regions of the ACT domain predicted to be involved in phenylalanine binding and feedback inhibition, *i.e.,* GALV and ESRP respectively (Liberles et al., 2005; Pohnert et al., 1999). Yamada et al. (Yamada et al., 2008) reported that GALV motif is conserved in all *Arabidopsis* ACT domain expect *Arabidopsis* ADT2, where this motif is present as GQLF. But in our study, GALV motif of ACT domain was found not to be conserved. On the other hand, ESRP motif, which is the most hydrophilic region in the ACT regulatory domain, is conserved across all six *Arabidopsis* ADTs and other selected species expect the lycophytes. Therefore, in *Arabidopsis* ADTs feedback inhibition mechanism is shown by the ESRP motif of ACT domain.

The PDTcs is an overlapping domain between PDT and ACT domains. This domain contains a conserved motif TRF that is found in all selected species expect lycophytes. ADT enzyme from lycophytes is not a full-length protein and that's the reason behind this exception. Mutation within this motif at the position of Thr showed no PDT activity in *E. coli* (Zhang et al., 2000). Therefore, PDTcs may be indirectly associated with the catalytic activity of PDT domain in land plants. The fact of no variation in domain architecture and existence of all domains in the selected species indicates the essentiality of these three domains throughout the evolution.

ADT protein family, the enzyme of interest in our study had evolved before the establishment of land plants as it was found to be present in red algae and green algae. This enzyme was also found to be present in fungus. Through literature study, the existence of this enzyme in bacteria was also traced in the form of PDT. It also became apparent that plant and bacterial sequences were quite distinct from each other evolutionarily. Differences in domains between plant ADT and bacterial PDT enzyme were identified as bacterial PDT enzyme has extra some domains (Cho et al., 2007). Independent loss of ADT enzyme was observed in fern and hornworts.

In our study, CNV was observed in many species which plays significant roles in plant evolution and adaptation (Duvaux et al., 2015). It results from unbalanced DNA modifications and trigger changes in the number of copies of a particular DNA sequence. This gene duplication event could be single gene or whole genome based. High level of CNV has been observed in different species of angiosperms. *Arabidopsis* ADT protein was found to form four different clusters which was in line with the findings of Cho et al. (2007) who reported similar pattern of clusters. Differences in cluster formation render the functionality differences of *Arabidopsis* ADT protein family. However, the exact reason behind the variation in cluster formation could not be identified in our study as no phylogenetic signal was detected. Having focused on the expression patterns of *Arabidopsis* ADT protein family, we have tried to figure out the functional variation of *Arabidopsis* ADT protein family in our next step.

Identifying the subcellular localization of proteins helps to define their functional role and subsequently leads to the identification of new unexpected characters (Sparkes & Brandizzi, 2012). When dissecting and differentiating the biological roles of members within protein families, this approach of finding subcellular location can particularly be useful (Karve et al., 2008). The stroma and stromules of chloroplasts are mostly targeted by most *Arabidopsis* ADTs (Köhler & Hanson, 2000). This subcellular localization lies in consistent with the enzymatic role of *Arabidopsis* ADTs in phenylalanine biosynthesis and also with the proposed role of stromules in increasing metabolite transport (Natesan et al., 2005).

Expression analysis of *Arabidopsis* ADT1 in both developmental stages and different anatomical parts unveils specific function of this isomer in leaf. This finding is consistent with Galili et al. who also reported the incredibly higher expression of ADT1 in leaves (Galili et al., 2016).

Arabidopsis ADT2 has the highest expression potential in germinated seeds, mature seeds and seedlings during both development and anatomical analysis. This implies that *Arabidopsis* ADT2 plays a more active role during seed development phase of *Arabidopsis*. No other ADTs have been found to be triggered in this stage. El-Azaz et al. also reported similar role of ADT2 in seed development in *Arabidopsis* (El-Azaz et al., 2018). It was investigated that mutation in SALKseq_044042 (insert position at exon number 6) and SALKseq_081342 Q (insert position at –34 bp from the open reading frame start codon) genes of *Arabidopsis* ADT2 caused seed lethality.

In both development and anatomical analysis, *Arabidopsis* ADT3 showed functionality in leaf, hypocotyl and replum. On the other hand, Para et al. reported expression potentiality of *Arabidopsis* ADT3 in stem and cotyledon (Para et al., 2016). In both development and anatomical analysis, *Arabidopsis* ADT4 exhibited higher expression potential in root, flower and leaf. Anatomy and developmental analysis revealed higher expression potential of *Arabidopsis* ADT5 among all other *Arabidopsis* ADTs, especially in root, stem, leaf and petal. *Arabidopsis* ADT6 showed expression potential in leaf, siliques and stem. A recent study has found that mutation in ADT reflect significant defects in germination hypocotyl elongation and root development responses during the seed to seedling transition suggesting their importance in growth and development process (Muhammad et al., 2023).

Co-expression analysis revealed AAT, ADH1-2 and CM1-3 genes to be highly co-expressed and these genes are integral part of the shikimate pathway. AAT transaminate prephenate into arogenate (Maeda et al., 2011) and ADH1-2 are involved in the biosynthesis of tyrosine (Connelly et al., 1986). As both AAT and ADH are the enzymes

of shikimate pathway, expression of ADT triggers the co-expression of those enzymes. As described earlier, shikimate pathway takes place within chloroplast and CM1, CM2, CM3 found to be involved in chloroplast biogenesis (Widhalm & Dudareva, 2015). Therefore, it also can be suggested that the expression of CM and ADT is positively correlated.

Finally, it can be concluded that *Arabidopsis* ADT2 plays a very specific role during the lifetime of *Arabidopsis* while the rest of the ADTs showed spatiotemporal expression. Spatiotemporal gene expression involves the activation of specific genes at particular times during development within specific tissues of an organism. This spatiotemporal expression indicated those ADTs to be highly variable.

Conclusion

Our phylogenetic study revealed that all the domains of ADT had evolved before the evolution of land plants as it was found in red algae, green algae and in fungus while expression analysis revealed the dominative role of ADT2 in seed development, whereas the rest of the ADTs showed spatiotemporal expression. Our final findings disclosed the conservation of all ADT domains in all the selected species; therefore, proper phylogenetic signal was absent. Another major limitation that needs to be addressed is the occurrence of poor bootstrap value in the case of some species. As our findings focused mainly on *Arabidopsis* ADT enzyme, further research could be carried out to co-relate *Arabidopsis* ADT with bacterial PDT and unveil the fact why PDT is completely conserved only in bacteria not in plants. Since some plant's species exhibit both ADT and PDT activity, further study could disclose the actual mechanism and reliable reason behind the utilization of both arogenate and prephenate as substrate by those species. Till now, none of the ADT enzyme related research has focused on the accurate reason behind the absence of such an important enzyme in ferns and hornworts. In future, we could lighten up on this and elucidate how this land plants synthesize phenylalanine in the absence of ADT. Understanding the core metabolic evolution of ADT toward the biosynthesis of phenylalanine-the core aromatic amino acid of plant will therefore help to reveal how plants adapted primary metabolism support diversified biological functions during the stressed environmental conditions.

References

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., & Huala, E. (2015). The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis*, *53*(8). https://doi.org/10.1002/dvg.22877

Bonner, C. A., & Jensen, R. A. (1987). A selective assay for prephenate aminotransferase activity in suspension-cultured cells of Nicotiana silvestris. *Planta*, *172*(3). https://doi.org/ 10.1007/BF00398672

Bross, C. D., Corea, O. R. A., Kaldis, A., Menassa, R., Bernards, M. A., & Kohalmi, S. E. (2011). Complementation of the pha2 yeast mutant suggests functional differences for arogenate dehydratases from Arabidopsis thaliana. *Plant Physiology and Biochemistry*, *49*(8). https://doi.org/10.1016/j.plaphy.2011.02.010

Bross, C. D., Howes, T. R., Rad, S. A., Kljakic, O., & Kohalmi, S. E. (2017). Subcellular localization of Arabidopsis arogenate dehydratases suggests novel and non-enzymatic roles. *Journal of Experimental Botany*, *68*(7). https://doi.org/10.1093/jxb/erx024

Cho, M. H., Corea, O. R. A., Yang, H., Bedgar, D. L., Laskar, D. D., Anterola, A. M., Moog-Anterola, F. A., Hood, R. L., Kohalmi, S. E., Bernards, M. A., Kang, C., Davin, L. B., & Lewis, N. G. (2007). Phenylalanine biosynthesis in Arabidopsis thaliana: Identification and characterization of arogenate dehydratases. *Journal of Biological Chemistry*, *282*(42). https://doi.org/10.1074/jbc.M702662200

Connelly, J. A., Conn, E. E. (1986). Tyrosine biosynthesis in Sorghum bicolor: isolation and regulatory properties of arogenate dehydrogenase. Z Naturforsch C J Biosci, 41(1-2). https://doi: 10.1515/znc-1986-1-212

Corea, O. R. A., Ki, C., Cardenas, C. L., Kim, S. J., Brewer, S. E., Patten, A. M., Davin, L. B., & Lewis, N. G. (2012). Arogenate dehydratase isoenzymes profoundly and differentially modulate carbon flux into lignins. *Journal of Biological Chemistry*, *287*(14). https://doi.org/10.1074/jbc.M111.322164

Dixon, R. A., & Paiva, N. L. (1995). Stress-induced phenylpropanoid metabolism. *Plant Cell*, *7*(7). https://doi.org/10.2307/3870059

Duvaux, L., Geissmann, Q., Gharbi, K., Zhou, J. J., Ferrari, J., Smadja, C. M., & Butlin, R. K. (2015). Dynamics of copy number variation in host races of the pea aphid. *Molecular Biology and Evolution*, *32*(1). https://doi.org/10.1093/molbev/msu266

El-Azaz, J., Cánovas, F. M., Á vila, C., & de la Torre, F. (2018). The arogenate dehydratase adt2 is essential for seed development in arabidopsis. *Plant and Cell Physiology*, *59*(12). https://doi.org/10.1093/pcp/pcy200

El-Azaz, J., de la Torre, F., Ávila, C., & Cánovas, F. M. (2016). Identification of a small protein domain present in all plant lineages that confers high prephenate dehydratase activity. *The Plant Journal : For Cell and Molecular Biology*, *87*(2). https://doi.org/10.1111/tpj.13195

Galili, G., Amir, R., & Fernie, A. R. (2016). The Regulation of Essential Amino Acid Synthesis and Accumulation in Plants. In *Annual Review of Plant Biology* (Vol. 67). https://doi.org/10.1146/annurev-arplant-043015-112213

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2012). Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research*, *40*(D1). https://doi.org/10.1093/nar/gkr944

Holliday, G. L., Mitchell, J. B. O., & Thornton, J. M. (2009). Understanding the Functional Roles of Amino Acid Residues in Enzyme Catalysis. *Journal of Molecular Biology*, *390*(3). https://doi.org/10.1016/j.jmb.2009.05.015

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W., & Zimmermann, P. (2008). Genevestigator V3: A Reference Expression Database for the Meta-Analysis of Transcriptomes. *Advances in Bioinformatics*, *2008*(1). https://doi.org/ 10.1155/2008/420747

Karve, A., Rauh, B. L., Xia, X., Kandasamy, M., Meagher, R. B., Sheen, J., & Moore, B. D. (2008). Expression and evolutionary features of the hexokinase gene family in Arabidopsis. *Planta*, *228*(3). https://doi.org/10.1007/s00425-008-0746-9

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Köhler, R. H., & Hanson, M. R. (2000). Plastid tubules of higher plants are tissue-specific and developmentally regulated. *Journal of Cell Science*, *113*(1). https://doi.org/10.1242/ jcs.113.1.81

Kutchan, T. M. (1995). Alkaloid biosynthesis - The basis for metabolic engineering of medicinal plants. *Plant Cell*, *7*(7). https://doi.org/10.2307/3870057

Letunic, I., & Bork, P. (2019). Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, *47*(W1). https://doi.org/10.1093/nar/gkz239

Liberles, J. S., Thórólfsson, M., & Martínez, A. (2005). Allosteric mechanisms in ACT domain containing enzymes involved in amino acid metabolism. In *Amino Acids* (Vol. 28, Issue 1). https://doi.org/10.1007/s00726-004-0152-y

Liu, Q., Luo, L., & Zheng, L. (2018). Lignins: Biosynthesis and biological functions in plants. *International Journal of Molecular Sciences*, *19*(2). https://doi.org/10.3390/ijms19020335

Lovell, S. C., Davis, I.W., Arendall, W. B. 3rd, de Bakker, P.I., Word, J. M., Prisant, M.G., Richardson, J.S., Richardson, D.C. (2003). Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins, 50(3):437-50. doi: 10.1002/prot.10286

Maeda, H., Shasany, A. K., Schnepp, J., Orlova, I., Taguchi, G., Cooper, B. R., Rhodes, D., Pichersky, E., & Dudareva, N. (2010). RNAi suppression of Arogenate Dehydratase1 reveals that phenylalanine is synthesized predominantly via the arogenate pathway in Petunia petals. *Plant Cell*, *22*(3). https://doi.org/10.1105/tpc.109.073247

Maeda, H., Yoo, H., & Dudareva, N. (2011). Prephenate aminotransferase directs plant phenylalanine biosynthesis via arogenate. *Nature Chemical Biology*, *7*(1). https://doi.org/ 10.1038/ nchembio.485

Morgat, A., Lombardot, T., Coudert, E., Axelsen, K., Neto, T. B., Gehant, S., Bansal, P., Bolleman, J., Gasteiger, E., de Castro, E., Baratin, D., Pozzato, M., Xenarios, I., Poux, S., Redaschi, N., & Bridge, A. (2020). Enzyme annotation in UniProtKB using Rhea. *Bioinformatics*, *36*(6). https://doi.org/10.1093/bioinformatics/btz817

Muhammad, D., Alameldin, H. F., Oh, S.,Montgomery, B. L., & Warpeha, K. M. (2023). Arogenate dehydratases: unique roles in light-directed development during the seed-to-seedling transition in Arabidopsis thaliana. Front. Plant Sci., 14. https://doi.org/10.3389/fpls.2023.1220732

Natesan, S. K. A., Sullivan, J. A., & Gray, J. C. (2005). Stromules: A characteristic cell-specific feature of plastid morphology. In *Journal of Experimental Botany* (Vol. 56, Issue 413). https://doi.org/10.1093/jxb/eri088

Pal, D., & Chakrabarti, P. (2002). On residues in the disallowed region of the Ramachandran map. *Biopolymers*, *63*(3). https://doi.org/10.1002/bip.10051

Para, A., Muhammad, D. S., Orozco-Nunnelly, D. A., Memishi, R., Alvarez, S., Naldrett, M. J., & Warpeha, K. M. (2016). The dehydratase ADT3 affects ROS homeostasis and cotyledon development. *Plant Physiology*, *172*(2). https://doi.org/10.1104/pp.16.00464

Pohnert, G., Zhang, S., Husain, A., Wilson, D. B., & Ganem, B. (1999). Regulation of phenylalanine biosynthesis. Studies on the mechanism of phenylalanine binding and feedback inhibition in the Escherichia coli P- protein. *Biochemistry*, *38*(38). https://doi.org/10.1021/ bi991134w

Rippert, P., Puyaubert, J., Grisollet, D., Derrier, L., & Matringe, M. (2009). Tyrosine and phenylalanine are synthesized within the plastids in arabidopsis. *Plant Physiology*, *149*(3). https://doi.org/10.1104/pp.108.130070

Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, *42*(W1), 320–324. https://doi.org/10.1093/nar/ gku316

Sahay, A., Piprodhe, A., & Pise, M. (2020). In silico analysis and homology modeling of strictosidine synthase involved in alkaloid biosynthesis in catharanthus roseus. *Journal of Genetic Engineering and Biotechnology*, *18*(1). https://doi.org/10.1186/s43141-020-00049-3

Smith, S. A., & Dunn, C. W. (2008). Phyutility: A phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics*, *24*(5), 715–716. https://doi.org/10.1093/bioinformatics/ btm619

Sparkes, I., & Brandizzi, F. (2012). Fluorescent protein-based technologies: Shedding new light on the plant endomembrane system. In *Plant Journal* (Vol. 70, Issue 1). https://doi.org/10.1111/j.1365-313X.2011.04884.x

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/ btu033

Tzin, V., & Galili, G. (2010). New Insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. In *Molecular Plant* (Vol. 3, Issue 6). https://doi.org/10.1093/mp/ssq048

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

Widhalm, J. R., & Dudareva, N. (2015). A familiar ring to it: Biosynthesis of plant benzoic acids. In *Molecular Plant* (Vol. 8, Issue 1). https://doi.org/10.1016/j.molp.2014.12.001

Yamada, T., Matsuda, F., Kasai, K., Fukuoka, S., Kitamura, K., Tozawa, Y., Miyagawa, H., & Wakasa, K. (2008). Mutation of a rice gene encoding a phenylalanine biosynthetic enzyme results in accumulation of phenylalanine and tryptophan. *Plant Cell*, *20*(5). https://doi.org/10.1105/tpc.107.057455

Yang, J., & Zhang, Y. (2015). I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Research*, *43*(W1). https://doi.org/10.1093/nar/gkv342

Yoo, H., Widhalm, J. R., Qian, Y., Maeda, H., Cooper, B. R., Jannasch, A. S., Gonda, I., Lewinsohn, E., Rhodes, D., & Dudareva, N. (2013). An alternative pathway contributes to phenylalanine biosynthesis in plants via a cytosolic tyrosine:phenylpyruvate aminotransferase. *Nature Communications*, *4*. https://doi.org/10.1038/ncomms3833

Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G., & Bhattacharya, D. (2004). A Molecular Timeline for the Origin of Photosynthetic Eukaryotes. *Molecular Biology and Evolution*, *21*(5), 809–818. https://doi.org/10.1093/molbev/msh075

Zhang, S., Wilson, D. B., & Ganem, B. (2000). Probing the catalytic mechanism of prephenate dehydratase by site- directed mutagenesis of the Escherichia coli P-protein dehydratase domain. *Biochemistry*, *39*(16). https://doi.org/10.1021/bi9926680

Supplementary

Table-1: List of species name, IDs and corresponding clades used in the phylogenetic analysis

Clades	Species Name	IDs used in Phylogeny
Angiosperm	*Arabidopsis thaliana*	RO_Q9SA96_AROD1_ARATH, RO_Q9SSE7_AROD2_ARATH, RO_Q9ZUY3_AROD3_ARATH, RO_O22241_AROD4_ARATH, RO_Q9FNJ8_AROD5_ARATH, RO_Q9SGD6_AROD6_ARATH
	Vitis vinifera	RO_E0CVS5_Vvinifera
	Juglans regia	RO_A0A2I4E2J5_Jregia
	Populus trichocarpa	RO_B9H107_Ptrichocarpa
	Jatropha curcas	RO_A0A067LA69_Jcurcas
	Ricinus communis	RO_B9RXK2_Rcommunis, RO_B9SN95_Rcommunis
	Eucalyptus grandis	RO_A0A059AAR7_Egrandis
	Gossypium raimondii	RO_A0A0D2PY58_Graimondii
	Nelumbo nucifera	RO_A0A1U7Z7Y6_Nnucifera

	Solanum lycopersicum	AS_46112970_Slycopersicum, AS_52689034_Slycopersicum
	Capsicum annuum	AS_A0A1U8FWM0_Cannuum
	Oryza sativa	MO_20169859_Osativa, MO_22542390_Osativa, MO_22564086_Osativa
	Zea mays	MO_113534989_Zmays, MO_166507370_Zmays, MO_59667944_Zmays
	Setaria italica	MO_K3ZT82_Sitalica
	Sorghum bicolor	MO_C5YFR9_Sbicolor
	Phoenix dactylifera	MO_A0A2H3YVE3_Pdactylifera
	Amborella trichopoda	BA_4332804_Atrichopoda
Gymnosperms	*Wollemia nobilis*	GY_A0A0C9QSN9_Wnobilis
	Araucaria cunninghamii	GY_A0A0D6R8Z6_Acunninghamii
	Keteleeria davidiana	GY_A0A1J1GBC9_Kdavidiana
	Pinus pinaster	GY_A0A1I9WKA3_Ppinaster
	Picea sitchensis	GY_B8LQ85_Psitchensis
Pteridophytes	*Selaginella moellendorffii*	LY_D8R0E4_Smoellendorffii
Bryophytes	*Physcomitrium patens*	MS_3437168_Ppatens, MS_8576463_Ppatens
	Marchantia polymorpha	LW_464888_Mpolymorpha, LW_126716_Mpolymorpha
Charophytes	*Klebsormidium nitens*	1 CA_kfl00456_0070_v1.1_Knitens
	Chara braunii	CA_A0A388MBG6_Cbraunii
Chlorophytes	*Chlorella sorokiniana*	CH_A0A2P6U5A5_Csorokiniana
	Trebouxia sp	CH_A0A5J4Y7C7_Tsp
	Bathycoccus prasinos	CH_K8E9V4_Bprasinos
	Monoraphidium neglectum	CH_A0A0D2N5H8_Mneglectum
	Auxenochlorella protothecoides	CH_A0A087SSZ4_Aprotothecoides
	Micromonas commode	CH_C1FED1_Mcommoda
	Chlamydomonas reinhardtii	CH_A8HXC5_Creinhardtii
	Micromonas pusilla	CH_t20072_Mpusilla
	Coccomyxa subellipsoidea	C CH_I0Z6Q1_Csubellipsoidea
	Micromonas sp	CH_Chr010856_Msp
	Ostreococcus lucimarinus	CH_1000098_Olucimarinus
	Chlamydomonas reinhardtii	CH_g261800_Creinhardtii

Gymnosperms	*Volvox carteri*	CH_0002s0472_Vcarteri
	Chromochloris zofingiensis	CH_Cz13g09120_Czofingiensis
	Botryococcus braunii	CH_0204s0015_Bbraunii CH_0204s0014_Bbraunii
Rhodophytes	*Porphyridium purpureum*	RA_A0A5J4YQA2_Ppurpureum
	Cyanidioschyzon merolae	RA_M1VDH6_Cmerolae
	Galdieria sulphuraria	RA_M2X4F9_Gsulphuraria
Fungus	*Fragilariopsis cylindrus*	FN_A0A1E7F413_Fcylindrus