DR GUANQUN(GAVIN) CHEN (Orcid ID : 0000-0001-5790-3903)

Article type : Original Article

Characterization of the diversification of phospholipid:diacylglycerol acyltransferases in the green lineage

Lucas J. Falarz[1,2], Yang Xu[1], Kristian Mark P. Caldo[1], Colin J. Garroway[2], Stacy D. Singer[3], Guanqun Chen[1,*]

[1] *Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Alberta, Canada, T6G 2P5*

[2] *Department of Biological Sciences, University of Manitoba, Winnipeg, Manitoba, Canada, R3T 2N2*

[3] *Agriculture and Agri-Food Canada, Lethbridge Research and Development Centre, Lethbridge, Alberta, Canada T1J 4B1*

[*]Corresponding author: Email: gc24@ualberta.ca; Phone: +1 (780) 492-3148; Fax: + 1 (780) 492-4265 (G. Chen)

Running title: PDAT diversification in green lineage

KEYWORDS

Arabidopsis thaliana, *Chlamydomonas reinhardtii*, phospholipid:diacylglycerol acyltransferase, lecithin:cholesterol acyltransferase, functional divergence, plant, algae, lipid biosynthesis

SUMMARY

Triacylglycerols have important physiological roles in photosynthetic organisms and are widely used as food, feed and industrial materials in our daily life. Phospholipid:diacylglycerol acyltransferase (PDAT) is the pivotal enzyme catalyzing the acyl-CoA-independent biosynthesis of triacylglycerols, which is unique in plants, algae and fungi but not in animals, and has essential functions in plant and algal growth, development and stress response. Currently, this enzyme has yet to be examined in an evolutionary context at the green lineage level. Some fundamental questions remain untapped, such as how PDATs were evolved in photosynthetic organisms and whether the evolution of terrestrial plant PDATs from a lineage of charophyte green algae diverges enzyme function. As such, we used molecular evolutionary analysis and biochemical assays to address these questions. Our results indicated that PDAT underwent divergent evolution in the green lineage. PDATs exist in a wide range of plants and algae, but not cyanobacteria. While PDATs exhibit conservation of several features, phylogenetic and selection pressure analyses revealed that overall they evolved to be highly divergent driven by different selection constraints. Positive selection, as one major driving force, may have resulted in enzymes with a higher functional importance in land plants than green algae. Further structural and mutagenesis analyses demonstrated that some amino acid sites under positive selection are critically important to PDAT structure and function, and may be central in lecithin:cholesterol acyltransferase family enzymes in general.

INTRODUCTION

Triacylglycerols (TAGs) produced by photosynthetic organisms represent an important storage compound in nature and have been widely used in our daily life (Xu *et al.*, 2018). TAG not only serves as an energy and carbon reserve in plants and algae, but also participates in many essential physiological processes (Yang and Benning, 2018; Hu *et al.*, 2008). For instance, TAG is evolved in the synthesis of cell building blocks and ATP production during seed germination, cell division, stomatal opening, pollen germination, membrane lipid remodeling and stress response in plant vegetative tissues, and photo-oxidative protection and response to adverse environmental conditions in microalgae (Hu *et al.*, 2008; Yang and Benning, 2018). TAG is synthesized in plants and microalgae via the acyl-Co-enzyme A (CoA)-dependent pathway (also known as the Kennedy pathway) and the acyl-CoA-independent pathway (Chapman and Ohlrogge, 2012; Xu *et al.*, 2018). The Kennedy pathway for TAG biosynthesis is well-conserved among animals, plants and microorganisms. However, in plants, algae and fungi, but not in animals, TAG is also synthesized in an acyl-CoA independent manner through the catalytic action of phospholipid:diacylglycerol acyltransferase (PDAT) (Dahlqvist *et al.*, 2000). In addition to its substantial contribution to TAG biosynthesis, PDAT also plays essential roles in mediating membrane lipid homeostasis, channelling unusual fatty acids to TAG, and stress response in plants and microalgae (Fan, Yan and Xu, 2013; Yoon *et al.*, 2012; Pan *et al.*, 2013; Fan, Yan, Zhang, *et al.*, 2013; Xu *et al.*, 2018). For instance, plant PDAT is essential for normal pollen and seed development and maintaining lipid homeostasis in growing tissues (Fan, Yan and Xu, 2013; Fan, Yan, Zhang, *et al.*, 2013; Fan *et al.*, 2014; Zhang *et al.*, 2009), and a couple of algal PDATs have been found to have possible roles in membrane remodeling and vigorous growth under stress conditions (X.,-Y., Liu *et al.*, 2016; Yoon *et al.*, 2012)□.

PDAT belongs to the lecithin:cholesterol acyltransferase (LCAT) family and shares structural similarity with human LCAT (Pan *et al.*, 2015; Ståhl *et al.*, 2004; Xu *et al.*, 2018; Dahlqvist *et al.*, 2000). Interestingly, the evolutionary fate of LCAT-family enzymes, which comprises not only PDATs, but also phospholipid:sterol O-acyltransferase (PSAT), LCAT-type phospholipase A (PLA hereafter), and LCAT-related proteins (a term used to describe members of the LCAT family not covered by the PDAT, PSAT and PLA subfamilies), differs substantially among kingdoms. While plants contain both PDAT and other LCAT-like enzymes, animals possess only LCAT-like enzymes

and fungi possess only PDATs, which suggests that distinct PDAT and LCAT-like enzymes were likely present in the common ancestor of these three groups and that losses in the respective kingdoms occurred more recently (Dahlqvist *et al.*, 2000). Although the maintenance of PDATs appears to be important for photosynthetic organisms, our understanding of the evolution of this enzyme, as well as its importance in fitness and TAG biosynthesis, across the green lineage is lacking.

Therefore, the aim of this study was to analyze the evolution of PDATs from photosynthetic organisms as a means of gaining insight into their functional roles across the green lineage. Our results revealed that while PDATs broadly exist in plant genomes and to a lesser extent in algae, they are non-existent in cyanobacteria. Algal and plant PDATs have typical LCAT features, but are highly divergent which is likely caused by different selection constraints. Among them, positive selection may have led to the increased functional importance of these enzymes in land plants compared to green algae. Furthermore, subsequent structural and biochemical analyses indicated that some positively selected sites within PDAT sequences have critical roles in PDAT function, which suggests that these sites may also have important functions in other LCAT family enzymes.

RESULTS

Algal and plant PDATs have typical LCAT features but are highly divergent

Available genome databases from photosynthetic organisms were searched for genes encoding putative LCAT-family enzymes including PDAT, PSAT, PLA, and LCAT-related proteins (Supplemental Table 1). In total, 37 and 69 putative LCAT-family genes were identified in 22 algal and 17 plant genomes, respectively, with a higher number per genome in plants than algae (4.1 sequences/ plant species versus 1.7 sequence/ algal species) (Figure 1). Multiple putative LCAT-family genes were identified in all 17 plant genomes but in only 22 of the 39 algal genomes. Although no *PDAT* genes were identified in cyanobacteria, 18 and 45 putative *PDAT*s were found in 15 of the 39 algal genomes and 16 of the 17 plant genomes assessed, respectively (Figure 1). In general, green algal PDATs were longer than their paralogs in plants and red algae/chromists (Supplemental Table 2 and Supplemental Table 3), which may be related to the unique structural and multifunctional features of green algal PDATs □(Yoon *et al.*, 2012; Dahlqvist *et al.*, 2000; Ståhl *et al.*, 2004)□.

Comparison of putative algal and plant PDAT protein sequences with human LCAT (HsaLCAT) (Glukhova *et al.*, 2015; Piper *et al.*, 2015; Peelman *et al.*, 1998) revealed that PDATs generally have typical conserved features of LCAT enzymes, including the "lid" region between two cysteines with a tryptophan in the middle, salt bridge, and catalytic S-D-H triad (S254, D573, and H626 in AtPDAT1; Figure 2A and Supplemental Figure 1)□. The positions of these features are all highly conserved, with regions of the PDATs comprising α/β hydrolase folds with the catalytic triad buried in the bottom of the active site pocket, as shown in the predicted 3D structures of PDATs from the model green alga *Chlamydomonas reinhardtii* and model plant *Arabidopsis thaliana* (Figure 2B). Based on the conserved locations of these, and a large number of identified motifs (Supplemental Figure 2), the PDAT amino acid sequences could be divided into three sections, with the middle section (between the "lid" domain and the catalytic serine, and containing the salt bridge and part of the α/β hydrolase domain) including motifs 5, 9, 14, 18, 1, and 12 being the most highly conserved (Supplemental Figure 3). Motif 12, which contains the serine residue of the catalytic triad, was conserved in all PDAT sequences (Supplemental Figure 2 and Supplemental Figure 3). In addition, the last five amino acids within this motif ([P/S/T/C/A/G]HS[M/L/Y/W]G) were highly conserved in all LCAT-family enzymes from algae and plants (Supplemental Figure 2 and Supplemental Figure 3).

Despite the high level of conservation observed with respect to particular motifs, PDAT have a lot of variation in their protein sequences, which suggests that the evolution of PDAT enzymes has been highly divergent (Figure 2 and Supplemental Figure 1). For example, many amino acid substitutions can be observed in the "lid" region, many amino acids other than the two cysteines and the tryptophan were changed during evolution. Moreover, the motifs containing the D and H residues were also highly diverse in LCAT-family proteins: despite of a few individual amino acids, the residues surrounding these sites varied substantially, which might lead to the formation of different enzyme subfamilies with new features during evolution (Supplemental Figure 3). Furthermore, the region downstream of the catalytic serine contains 24 of the 32 unique motifs in algal and plant PDATs and is thus very divergent.

PDATs evolved to be functionally divergent in plants and algae through the facilitation of different selection constraints

Phylogenetic analysis of algal and plant LCAT-family proteins indicated that algal LCAT-related enzymes are most closely related to algal PSATs, whereas plant LCAT-related enzymes form a monophyletic group with plant PLAs and are phylogenetically distinct from plant PSATs (Figure 3 and Supplemental Figure 4). Since land plants and green algae belong to the *Viridiplantae* and share an ancient green algal organism as their common ancestor (Clerck *et al.*, 2012; Leliaert *et al.*, 2012), their PDATs were chosen in subsequent analyses to study the functional divergence of PDATs. The divergent evolution of PDATs does not follow the evolution of plants and green algae, indicating the enzymes may have diverse functions (Figure 1, Figure 3 and Supplemental Figure 4). In line with previous findings, plant PDATs could be separated into two large groups, with Group A proteins comprising PDATs from basal land plants, all monocots and group V and VI eudicots, and Group B proteins consisting of PDATs from group VII eudicots, which are basal to other eudicot and monocot PDATs (Pan *et al.*, 2015).

To explore whether the phylogenetic distinction between plant and green algal PDATs might translate into functional divergence between the two groups, selection pressure analysis was performed with Clade model C as the alternative model (H1) and M2a_rel as the null model (H0) (Bielawski and Yang, 2004; Yang *et al.*, 2005; Weadick and Chang, 2012). The likelihood-ratio test (LRT) reveals that H1 fits the data significantly better than H0 ($p < 0.001$) (Supplemental Table 4 and Supplemental Figure 5). When comparing green algal PDATs with those from Group A and Group B plants, 58.7% and 53.5% of the analyzed sites with a divergent evolution evolved under different selection pressures (site class 2), respectively (Supplemental Table 4). These analyses, together with the very different characteristics (e.g. protein length and pI values; Supplemental Table 2 and Supplemental Table 3) and distinct regions/motifs in PDAT sequences from green algae and land plants (Supplemental Figure 2 and Supplemental Figure 3), strongly suggest functional divergence in the evolutionary history of PDATs from *Viridiplantae.*

Considering the high proportion of amino acids that have undergone divergent evolution in PDATs, it is likely that some of them may have evolved under positive selection and thus have served as a driving force behind functional divergence. To test this hypothesis, a branch-site model and associated Bayes Empirical Bayes (BEB) analysis (Yang and Nielsen, 2002; Zhang *et al.*, 2005; Yang *et al.*, 2005) were used to estimate the amino acid sites under positive selection and their posterior

probability□. As shown in Supplemental Table 5, there is a high proportion of residues classified in site class 2 ($p < 0.001$). This high number of sites under positive selection corroborated the functional divergence among the green algal and plant PDATs, which indicated that positive selection pressure facilitated their divergent evolution.

Selection pressure analysis also indicated that the ω ratios (the ratio of non-synonymous substitution rate and synonymous substitution rate or dN/dS; a non-synonymous substitution means that the mutation causes the codon to encode a new amino acid, while the synonymous substitution means that the mutation does not cause a change in the amino acid encoded) of the green algal branch at site class 2 were higher than that of Group A and Group B plants, which means that evolution was more relaxed in the algal group (Supplemental Table 4). Consistently, the branch-site model analysis indicated plant PDATs have more sites under positive selection than algae (Supplemental Table 5). In addition, BEB analysis results indicated that plant PDATs accumulated more sites than green algae that underwent positive selection after their divergence from the common ancestor of Chlorophyta and Streptophyta (Supplemental Table 5, Figure 4, Supplemental Figure 6 and Supplemental Figure 7). Indeed, in the comparisons of Group A plants and green algae PDATs, 42 and 12 sites were identified to have undergone positive selection, respectively (Figure 4, Supplemental Figure 7 and Supplemental Table 6). Because positive selection of enzymes is advantageous to the organism, these results indicated PDATs likely evolved to be more relevant to the metabolisms in plants than green algae and positive selection facilitated in the process.

Positively selected sites may have critical roles to the enzymatic function of PDATs

Sites that were predicted to have undergone positive selection were analyzed to investigate whether positive selective pressure could serve as a driving force behind the divergent evolution of PDATs in the green lineage. Since most positive selections were identified between comparisons of Group A plants and green algae, the analyses were mainly performed on these sites (Figure 4, Supplemental Figure 7 and 8, and Supplemental Table 6). Interestingly, the sites at which positive selection has occurred were largely variable. Indeed, none of the positive selection sites present in plant PDATs were also under positive selection in green algae and vice versa, and many of the sites were occupied by amino acids of different groups with different molecular structures between the two groups. For

instance, four positively selected sites in green algae were made up of aliphatic/hydroxyl amino acids (V291, V489, A728, and I1029 in CrPDAT) while their corresponding sites in plants consisted of aromatic amino acids. Similarly, the only two aromatic amino acids under positive selection in Group A plants (Y259 and Y603 in AtPDAT1) have corresponding sites in green algae that are occupied by aliphatic amino acids.

This large variability in sites that have undergone positive selection between plant and algal PDATs could feasibly contribute to differences in their substrate binding and specificity properties, as has been found previously in human LCAT (Wang *et al.*, 1997; Peelman *et al.*, 1998). In line with this, some of the positive selection sites are predicted to be located at important positions within the PDAT enzymes. For instance, three basic residues in a highly conserved region upstream of the catalytic serine and lying partially within the "lid" region were under positive selection in plant PDATs (R149, R200 and K246 in AtPDAT1), whereas the corresponding sites in green algal PDATs, which were not under positive selection, were occupied with acidic amino acids (Figure 4 and Supplemental Table 6). Similarly, a positively selected site within the "lid" region of green algal PDATs (V291 in CrPDAT) comprised hydroxyl or aliphatic side-chain amino acids while its plant counterpart consisted of an aromatic amino acid. Moreover, three highly conserved amino acids located within close proximity to the catalytic serine (+1, +3 and +7 positions; Y402, E404 and R408 in CrPDAT) that were under positive selection in green algae (Figure 4 and Supplemental Figure 7), differed in plant, chromist and *S. cerevisiae* PDATs (Figure 4, Supplemental Figure 7 and Supplemental Figure 8). Y402 appears to occupy a large space at the bottom of the catalytic pocket in the predicted 3D structure of CrPDAT and E404 and R408 may form a salt bridge in the helix (Supplemental Figure 9 and Supplemental Figure 10), which can affect the conformation of the active site and change PDAT's substrate preferences or catalytic activities (Donald *et al.*, 2011).

Further analysis of positively selected sites in Arabidopsis AtPDAT1 indicated that certain sites may possibly affect the secondary and tertiary structures of this enzyme. For example, the side chains of S158, T222, T228, and T242 have the potential to form polar contacts with the main chain, which could affect the stabilization of the protein's secondary structure (Figure 5). Moreover, the side chains of R149, S176, Q336, and S610 (Figure 5A and 5C) may be involved in polar contacts to influence the tertiary structure of AtPDAT1. Indeed, R149 located at the C-terminus of a helix has its

guanidinium group forming polar contact with the carbonyl oxygen of F143, an amino acid in a preceding helix, and the guanidinium-carboxyl interaction may be involved in stabilizing a helix-loop-helix motif (Figure 5C). Furthermore, positively selected amino acids located between the catalytic aspartic acid and histidine in AtPDAT1 (Figure 5), which are in the same region as the functionally important L372 and T347 residues in human LCAT (Supplemental Figure 11)(Calabresi *et al.*, 2005; Klein *et al.*, 1993; Piper *et al.*, 2015), and their substitution may disrupt catalytic activity and/or affect substrate binding.

Structural and directed mutagenic analyses of the first residue upstream the catalytic serine confirmed its critical function in PDATs

The first residue upstream of the catalytic serine is predicted to be under positive selection in green algal PDATs (Figure 4 and Supplemental Figure 7 and 8). Considering the fact that this site is very close to the catalytic triad, the +1 position of the catalytic serine (Y402 and M255 in CrPDAT and AtPDAT, respectively) was chosen for further *in silico* and enzymatic analysis. In the CrPDAT model, Y402 forms a non-polar interaction with the side chains of M507 and F260; however, a Y402M conversion was predicted result in the loss of such interactions (Supplemental Figure 10). In contrast, while M255 in AtPDAT1 is also in close proximity to the active site pocket (Figure 6A), it likely does not affect the position or orientation of the catalytic triad (S254, D573 and H626) since neither M255W or M255A conversions were predicted to affect amino acids within the active site pocket (Figure 6B-D) or alter the distances between atoms involved in polar interactions with the three residues (Figure 6E). Nevertheless, M255 and two nearby hydrophobic residues (M351 and I116) may form a network of hydrophobic interactions, whereby the distance between each involved atom is less than 4 angstroms (Figure 6F). It was showed that the three residues interact because of the overlapping of their van der Waals radii, which confirmed the existence of these hydrophobic interactions (Figure 6G). The surface view of the active site pocket also indicated that M255, I116, and M351 might form a hydrophobic patch on the wall of the pocket (Figure 6H). Interestingly, M255W and M255A conversions both appear to disrupt the aforementioned network of hydrophobic interactions in the active site pocket (Figure 6C-D) and alter the landscape of the hydrophobic patch (Figure 6I-J), which could negatively affect PDAT activity since the two major substrates of

AtPDAT1 (PC and DAG) are both hydrophobic. To confirm our *in silico* results, two AtPDAT1 mutants (M255W and M255A) were generated using site-directed mutagenesis for enzymatic assay. As expected, both single amino acid replacements at M255 were found to abolish PDAT activity compared to the wild-type sequence to levels that were in line with the negative control (Figure 7). Considering highly conserved amino acids were present in the +1 position, the amino acid appears to be vital for the function of the LCAT-family enzymes in various organisms, including algal PDATs such as CrPDAT (Figure 8, Supplemental Figure 1, and Supplemental Table 7).

DISCUSSION

TAG is the major storage lipid in most organisms and has great nutritional and industrial value (Xu *et al.*, 2018). PDAT, which is absent in animals, catalyzes the last and committed step in the acyl-CoA-independent biosynthesis of TAG and plays crucial physiological roles in many photosynthetic organisms (Xu *et al.*, 2018; Hu *et al.*, 2008; Yang and Benning, 2018). However, its evolution and structural features have not been characterized in depth in the context of the green lineage and other algal species. Our phylogenetic analyses indicated that PDATs are broadly existent in plant and algae, but not cyanobacteria, which suggests that PDATs evolved following the divergence of cyanobacteria and other algae, but before terrestrial plants diverged from charophyte green algae around 474–515 million years ago (Morris *et al.*, 2018).

Although both green algae and plants belong to *Viridiplantae* and share a common ancestor (Leliaert *et al.*, 2012; Clerck *et al.*, 2012)□, the evolution of their PDATs is not closely aligned to the evolution of species within these groups (Figure 1, Figure 3, Supplemental Figure 4). Instead, algal and plant PDATs appear to have divergently evolved (Supplemental Table 2, Supplemental Table 5, Supplemental Figure 2 and Supplemental Figure 3), which may have led to the appearance of different enzymatic functions in green algal and plant PDATs. Functional differences appear to exist between PDATs of the two groups. Indeed, CrPDAT (but not AtPDAT1) exhibits lipase and phospholipase activities with broad substrate preferences (van Erp *et al.*, 2011; Pan *et al.*, 2013; Dahlqvist *et al.*, 2000; Ståhl *et al.*, 2004; Yoon *et al.*, 2012)□. It should be noted that CrPDAT contains the lipase catalytic motif (GXSXG) but this motif is less conserved in AtPDAT1 and 2,

where the replacement of the first Gly by a Pro likely results in the absence of two β-sheets that may be essential for lipase activity (Yoon *et al.*, 2012). In addition, although the knockout of *CrPDAT* has been found to reduce TAG content by up to 25% in *C. reinhardtii*, the mutant algal cells were still viable (Boyle *et al.*, 2012). Conversely, disruption of *AtPDAT1* in *A. thaliana* led to a more than 50% decrease in oil accumulation in growing tissues (Fan, Yan and Xu, 2013; Fan, Yan, Zhang, *et al.*, 2013), and the mutant plants also displayed necrosis, growth retardation and gametophytic defects (Zhang *et al.*, 2009; Fan, Yan and Xu, 2013; Fan, Yan, Zhang, *et al.*, 2013). It should be pointed out that CrPDAT contains a predicted chloroplast transit peptide and was previously found in the chloroplast of *C. reinhardtii* based on proteomics analysis (Terashima *et al.*, 2011). The chloroplast localized CrPDAT appears to be physiological important as evidence showed that it may be involved in chloroplast membrane remodeling and TAG biosynthesis in the chloroplast (Yoon *et al.*, 2012). Distinct subcellular location was also predicted for some other plant and algal PDATs (Supplemental Table 8), which indicated the possible physiological importance and uniqueness of PDAT.

In line with this, our analyses also suggested that PDATs may play a more critical role in plants than green algae (Supplemental Table 4 and Supplemental Table 5) since PDATs were found to be absent in many algal species, whereas almost all plant species assessed contained at least one PDAT (Figure 1). Moreover, genetic duplication of algal PDATs was only observed in one of the green algal species assessed (Figure 3), but has occurred at least twice among plant PDATs followed by a more recent duplication in some species (Pan *et al.*, 2015). The retention of these extra PDAT copies during the evolution of eudicots may indicate that they contribute to increased fitness in these plants, though we can not role out the possibility that sub-functionalization, non-functionalization and/or redundancy may have occurred in at least a proportion of the additional PDAT copies (Panchy *et al.*, 2016; Pan *et al.*, 2015). Furthermore, given that the ω ratios of green algal PDAT branches were higher than those of both Group A and Group B plants (Supplemental Table 4), green algal PDATs likely evolved under more relaxed constraints and could thus accumulate a higher number of mutations with minimal impact on fitness than their plant homologs during evolution (Yang, 2007; Yang *et al.*, 2000). This apparent diminishment in the necessity of PDATs in green algae may be directly related to the higher numbers of diacylglycerol acyltransferases (DGATs), which catalyze the last and committed step in the acyl-CoA-dependent biosynthesis of TAG (Chen and Smith, 2012) and

may compensate for a lack of PDAT, that are present in this group than plants. Indeed, *C. reinhardtii* DGATs substantially contribute to TAG biosynthesis, whereas only one copy of *PDAT* is found in its genome and its down-regulation had a limited effect on TAG accumulation (Yoon *et al.*, 2012; Liu *et al.*, 2016). Since Arabidopsis, PDAT and DGAT1 have functional complementation in TAG biosynthesis (Zhang *et al.*, 2009), it would be a useful next step in the investigation of the possible function complementation of algal DGAT and PDAT. However, although multiple *DGAT* genes have been reported in green algae, not all the encoded enzymes are functional. For instance, in *C. reinhardtii DGAT2s* (also known as DGTTs), CrDGTT1, CrDGTT1, and CrDGTT3 are functional but CrDGTT4 has no function (Liu *et al.*, 2016). Similarly, most DGATs from *Chlorella zofingiensis* have DGAT activity except CzDGTT3 and CrDGTT8 (Mao *et al.*, 2019). This should be taken into consideration when studying algal DGATs and PDAT.

Both algal and plant PDATs have maintained many of the typical conserved LCAT features, including the "lid" region with two conserved cysteines and one conserved tryptophan, a salt bridge, and the catalytic S-D-H triad (Peelman *et al.*, 1998; Glukhova *et al.*, 2015; Piper *et al.*, 2015) (Figure 2 and Supplemental Figure 1-3) □. However, despite the conservation of such regions, overall they are highly divergent in their structures, which may have allowed PDATs to evolve different enzymatic properties that would benefit the organisms in terms of adapting to different environments. For example, the high level of amino acid differences except a few ones within the "lid" region could create a certain amount of functional and substrate plasticity to PDATs, as has been reported previously in human LCAT, CrPDAT and human lysosomal PLA2 (Adimoolam and Jonas, 1997; Glukhova *et al.*, 2015; Piper *et al.*, 2015; Yoon *et al.*, 2012). The region between the "lid" domain and the catalytic serine is also almost certainly important for PDAT activity and substrate preference, as is the case in human LCAT (Glukhova *et al.*, 2015; Piper *et al.*, 2015). Within this region, motif 12 (containing the serine residue of the catalytic triad) contains the typical [P/S/T/C/A/G]HS[M/L/Y/W]G domain of LCAT-family enzymes, which is analogous to the lipase motif (G-X-S-X-G) in hydrolytic enzymes including lipases and serine proteases (Derewenda and Derewenda, 1991; Ollis *et al.*, 1992). Considering that the first amino acid residue of this motif is highly conserved as glycine in hydrolases (Derewenda and Derewenda, 1991)□, it is possible that

variation at this site could allow LCAT-family enzymes to use different substrates and has contributed to divergent catalytic activities.

Our analyses indicated that positive selection, which is usually associated with a beneficial mutation leading to functional divergence that was retained by a population due to increased fitness (Hughes, 2007), served as a driving force in the divergent evolution of PDATs in green algae and plants (Supplemental Table 5).

Our observation that 42 positively selected sites were present in PDATs from Group A plants compared to 12 in green algae (Figure 4) suggested more beneficial substitutions were retained in plant PDATs. In addition, the higher ω ratio obtained in the algal group in the Clade model C analysis (Supplemental Table 4) suggests a more relaxed evolution in the algal PDATs. These results combined support the notion that these enzymes may have a higher functional importance in plants than algae (Supplemental Table 4 and Supplemental Table 5). Indeed, the large amount of variability in positive selection sites, which were often located at important positions in terms of protein structure (Figure 4, Supplemental Figures 7-10, and Supplemental Table 6), as well as the amino acids occupying these sites, between plant and algal PDATs (Figure 4, Supplemental Figures 7 and 8, and Supplemental Table 6) may well contribute to different substrate binding capacities and specificities of plant and algal PDATs, as has been found previously in human LCAT (Peelman *et al.*, 1998; Wang *et al.*, 1997), as well as *Streptomyces lividans* xylanase A (Roberge *et al.*, 1999) and xylose isomerase (Meng *et al.*, 1993).

The +1 position of the catalytic serine may be particularly important for the function of PDATs. This position is highly conserved as a bulky aromatic amino acid tyrosine in green algae but a hydroxyl methionine residue in plants (Figure 4, Supplemental Figure 7 and Supplemental Figure 8). The *in silico* analysis indicated that replacing tyrosine with methionine in the green algal CrPDAT at this location (Y402M) would likely lead to loss of non-polar interactions with other amino acids, thus disrupting its function (Supplemental Figure 10). Similarly, replacing the hydroxyl methionine in AtPDAT1 with the aromatic amino acid tyrosine (M255W) or an aliphatic one (M255A) abolished the PDAT activity (Figure 6, Figure 7). Therefore, to retain the aromatic tyrosine at the +1 position of the catalytic serine, which has undergone positive selection during evolution and is supposed to be beneficial to green alga PDATs, other conserved sites would be required. Indeed, the +3 and +7

positions to the catalytic serine are both under positive selection and highly conserved in green algae, which may help properly adjust the conformation of the catalytic pocket (Figure 4). Interestingly, amino acids at this position are highly conserved in the LCAT-family proteins in general (Figure 8, Supplemental Figure 1, and Supplemental Table 7). Mutations in this highly conserved +1 position may negatively affect the activity of LCAT-family enzymes. For example, in human LCAT, this site is occupied by a leucine (L206) and mutagenesis at this site (L206M and L206Y) would likely change the conformation of the active site pocket and negatively affect the enzyme activity (Supplemental Figure 11). Therefore, this position may also be particularly important for the function of LCAT-family enzymes in various organisms.

In conclusion, this study revealed the divergent evolution of PDAT in plants and algae and identified positive selection as a driving force in the process, which may have increased the functional importance of PDAT enzymes in land plants. Moreover, some amino acids sites under positive selection are important to PDAT and may be central in other LCAT family enzymes in general, which are valuable in enzyme designing for the production of lipids and unusual fatty acids, and shed light upon the study of lecithin:cholesterol acyltransferase family enzymes.

METHODS

Identification and characterization of PDAT paralogs in algae and plants

PDAT paralogs were retrieved from the available genomes of 17 plant and 39 algal species (Supplemental Table 1). Algal genomes, as well as *Brassica napus*, *B. rapa*, *B. oleracea capitata*, *Macrozamia polymorpha*, and *Ricinus communis* genomes were searched locally with TBLASTN v2.6.0 and BLASTP v2.6.0 from the Blast+ suite (Camacho *et al.*, 2009) using PDAT homologs from *A. thaliana* and *C. reinhardtii* (AT5G13640, AT3G44830, Cre02.g106400) as queries with an e-value cut-off of $1x10^{-9}$. Additional PDAT homologs were directly retrieved from the Phytozome database as previously reported (Pan *et al.*, 2015). PDAT paralogs in cyanobacteria were searched against 376 genomes in CyanoBase (http://genome.microbedb.jp/cyanobase; accessed on July 23, 2019) using BLASTP with CrPDAT and AtPDAT1 sequences as the queries with an e-value cutoff of $1x10^{-5}$. The complete list of genes used in this study and their acronyms are found in Supplemental Table 7.

All sequences were further classified based on PANTHER (Mi *et al.*, 2017) implemented in InterProScan v5.22.61 (Jones *et al.*, 2014). PANTHER subfamilies PTHR11440:SF4 and PTHR11440:SF54 represent PDATs in algae and plants, respectively. Sequences classified in PTHR11440:SF53 and PTHR11440:SF7 represent phospholipases (PLAs) with the same evolutionary origin as LCAT and PSAT, respectively. The proteins classified in other LCAT subfamilies or not having a subfamily are shown as LCAT-related in this study. Although only a few PLAs were classified in the LCAT family, they form a monophyletic group in other analyses. Therefore, they are classified as a separate PLA group in this study, except in Figure 1 where PLAs were included in the LCAT-related group for ease of comparison. In-detail characterization of PDAT paralogs were conducted as described in the Supplemental Methods.

Phylogenetic analysis

Phylogenetic tree construction was carried out using maximum likelihood (ML) and Bayesian inference (BI) methods, which were conducted in PhyML v3.0 (Guindon *et al.*, 2010) and MrBayes V3.2.6 (Ronquist *et al.*, 2012), respectively. The complete description of tree reconstruction is in the Supplemental Methods.

Selection pressure analysis

Selection pressure analyses were performed using the CODEML program implemented in the PAML v4.9 package (Yang, 2007). Clade model C and branch-site models were used to test for the presence of functional divergence and positive selection, respectively (Yang *et al.*, 2005; Yang and Nielsen, 2002; Zhang *et al.*, 2005). The detailed procedure is described in the Supplemental Methods.

Protein structure prediction

The models of native and variant AtPDAT1 and CrPDAT were obtained through structural homology modeling using SWISS-MODEL (Waterhouse *et al.*, 2018) and PHYRE2 software (Kelly *et al.*, 2015) with human lysosomal phospholipase A2 and human LCAT as templates (Glukhova *et al.*, 2015), respectively. The structure of HsaLCAT was previously obtained in complex with a Fab fragment from a tool antibody and was used to determine the structure of the native and variant

HsaLCAT using SWISS-MODEL (Piper *et al.*, 2015). The structural models were viewed using PYMOL software, which was also used to determine the possible interactions formed by the positive selection sites identified.

Site-directed mutagenesis of AtPDAT1, heterologous protein expression, and *in vitro* PDAT activity assay

The full-length coding sequence of AtPDAT1 (AT5G13640) was previously isolated in our laboratory from A. thaliana (Col-0). Site-directed mutagenesis of *AtPDAT1* was carried out using overlap extension PCR (Heckman and Pease, 2007) with the primers in Supplemental Table 9. Heterologous protein expression and *in vitro* PDAT activity assay were conducted with the procedure described in our previous study (Falarz *et al.*, 2019). In brief, the wild-type AtPDAT1, two mutant AtPDAT1, and LacZ negative control were cloned into the pYES2.1/V5-His TOPO vector (Invitrogen), respectively, according to the manufacture's manual. Yeast transformation was performed using the lithium acetate/single-stranded carrier DNA/PEG method to introduce the expression vectors, respectively. The *S. cerevisiae* strain used in this study was the quadruple knock-out H1246 strain (*MATα are1-Δ::HIS3, are2-Δ::LEU2, dga1-Δ::KanMX4, lro1-Δ::TRP1 ADE2*), which lacks the ability to synthesize neutral lipids (Sandager *et al.*, 2002). As for the *in vitro* PDAT assays, 1-palmitoyl-2-dodecanoyl-nitrobenzoxadiazole-*sn*-glycero-3-glycerol (NBD-DAG), 1,2-dihexadecanoyl-*sn*-glycero-3-phosphocholine (16:0,16:0-PC), 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (18:1,18:1-PC), and 1,2-dilinoleoyl-*sn*-glycero-3-phosphocholine (18:2,18:2-PC) were chosen as substrates, where the acyl groups at *sn*-1 position of DAG and *sn*-2 position of PC are naturally present in Arabidopsis seed oil.

DATA AVAILABILITY STATEMENT

All relevant data can be found within the manuscript and its supporting materials.

ACKNOWLEDGMENTS

The authors thank Dr. Sten Stymne of Swedish University of Agricultural Sciences for the yeast H1246 strain and Dr. Xue Pan of the University of California at Riverside for her critical comments. This work was supported by Alberta Innovates, Alberta Agriculture and Forestry, Natural Sciences

and Engineering Research Council of Canada Discovery Grant (RGPIN-2016-05926), the Canada Research Chairs Program, the University of Alberta Start-up fund and the University of Manitoba Graduate Enhancement of Tri-Council Stipends (GETS) program.

AUTHOR CONTRIBUTIONS

G.C. conceptualized, designed and supervised the experiment. L.J.F., Y.X., and K.M.P.C. performed the research and analysed the data. C.G. contributed to the discussion. L.J.F. and G.C. prepared the initial draft of the manuscript with the contribution of Y.X. and K.M.P.C. S.D.S. significantly contributed to the research conception and manuscript editing. All authors were instrumental in the preparation of the final article.

CONFLICTS OF INTEREST STATEMENT

The authors declare no conflicts of interest.

SHORT SUPPORTING MATERIALS LEGENDS

Supplemental methods

Supplemental Figure 1. Sequence alignment of selected PDATs with human LCAT (HsaLCAT).

Supplemental Figure 2. The sequence logos of the identified conserved motifs

Supplemental Figure 3. Predicted transmembrane domains and conserved motifs on PDATs.

Supplemental Figure 4. The Bayesian inference phylogenetic tree of algal and plant lecithin-cholesterol acyltransferase (LCAT)-family enzymes.

Supplemental Figure 5. Phylogenetic trees used in the Clade model C analyses.

Supplemental Figure 6. Phylogenetic trees used in the branch-site model analyses.

Supplemental Figure 7. Positive selection sites of identified in phospholipid:diacylglycerol acyltransferases (PDATs) of all green algae and Group A plants in this study.

Supplemental Figure 8. Positive selection sites of identified in phospholipid:diacylglycerol acyltransferases (PDATs) of green algae and Group B plants.

Supplemental Figure 9. Predicted structures of *Arabidopsis thaliana* PDAT1 and *Chlamydomonas reinhardtii* PDAT with the sites under positive selection shown in red.

Supplemental Figure 10. Structure analysis of *Chlamydomonas reinhardtii* PDAT (CrPDAT).

Supplemental Figure 11. Structure analysis of human lecithin-cholesterol acyltransferase

Supplemental Table 1. List of genomes used in this study.

Supplemental Table 2. Properties of lecithin:cholesterol acyltransferase (LCAT)-family enzymes identified in the BLAST search from plants and algal sequences.

Supplemental Table 3. Properties of individual sequences of lecithin:cholesterol acyltransferase (LCAT)-family proteins used in this study.

Supplemental Table 4. Parameter estimates of Clade model C analyses to understand the selection pressures that acted on the evolution of PDATs in green algal and plant.

Supplemental Table 5. Parameter estimates of branch-site models to verify the presence of positive selection in the evolution of green algal and plant PDATs.

Supplemental Table 6. Comparison of amino acids under positive selection in PDATs.

Supplemental Table 7. List of the protein identifiers and acronyms used in this study.

Supplemental Table 8. Predicted subcellular localization of PDAT from plants and algae.

Supplemental Table 9. Primers used to in this study.

REFERENCES

Adimoolam, S. and Jonas, A. (1997) Identification of a domain of lecithin-cholesterol acyltransferase that is involved in interfacial recognition. *Biochem. Biophys. Res. Commun.*, **232**, 783–787.

Bielawski, J.P. and Yang, Z. (2004) A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J. Mol. Evol.*, **59**, 121–132.

Boyle, N.R., Page, M.D., Liu, B., et al. (2012) Three acyltransferases and nitrogen-responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in *Chlamydomonas*. *J. Biol. Chem.*, **287**, 15811–15825.

Calabresi, L., Pisciotta, L., Costantin, A., et al. (2005) The molecular basis of lecithin: cholesterol acyltransferase deficiency syndromes: A comprehensive study of molecular and biochemical

findings in 13 unrelated Italian families. *Arterioscler. Thromb. Vasc. Biol.*, **25**, 1972–1978.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) BLAST+: Architecture and applications. *BMC Bioinformatics*, **10**, 421.

Chapman, K.D. and Ohlrogge, J.B. (2012) Compartmentation of triacylglycerol accumulation in plants. *J. Biol. Chem.*, **287**, 2288–2294.

Chen, J.E. and Smith, A.G. (2012) A look at diacylglycerol acyltransferases (DGATs) in algae. *J. Biotechnol.*, **162**, 28–39.

Clerck, O. De, Bogaert, K.A. and Leliaert, F. (2012) Diversity and evolution of algae: Primary endosymbiosis. In G. Piganeau, ed. *Genomic Insights into the Biology of Algae*. Advances in Botanical Research. Academic Press, pp. 55–86.

Dahlqvist, A., Stahl, U., Lenman, M., Banas, A., Lee, M., Sandager, L., Ronne, H. and Stymne, S. (2000) Phospholipid:diacylglycerol acyltransferase: An enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 6487–6492.

Derewenda, Z.S. and Derewenda, U. (1991) Relationships among serine hydrolases: Evidence for a common structural motif in triacylglyceride lipases and esterases. *Biochem. Cell Biol.*, **69**, 842–851.

Donald, J.E., Kulp, D.W. and DeGrado, W.F. (2011) Salt bridges: Geometrically specific, designable interactions. *Proteins*, **79**, 898–915.

Erp, H. van, Bates, P.D., Burgal, J., Shockey, J. and Browse, J. (2011) Castor phospholipid:diacylglycerol acyltransferase facilitates efficient metabolism of hydroxy fatty acids in transgenic *Arabidopsis*. *Plant Physiol.*, **155**, 683–693.

Falarz, L., Xu, Y., Singer, S.D. and Chen, G. (2019) A fluorescence-based assay for quantitative analysis of phospholipid:diacylglycerol acyltransferase activity. *Lipids*, **54**, 571–579.

Fan, J., Yan, C., Roston, R., Shanklin, J. and Xu, C. (2014) Arabidopsis Lipins, PDAT1 acyltransferase, and SDP1 triacylglycerol lipase synergistically direct fatty acids toward β -oxidation, thereby maintaining membrane lipid homeostasis. *Plant Cell*, **26**, 4119–4134.

Fan, J., Yan, C. and Xu, C. (2013) Phospholipid:diacylglycerol acyltransferase-mediated triacylglycerol biosynthesis is crucial for protection against fatty acid-induced cell death in

growing tissues of Arabidopsis. *Plant J.*, **76**, 930–942.

Fan, J., Yan, C., Zhang, X. and Xu, C. (2013) Dual role for phospholipid:diacylglycerol acyltransferase: enhancing fatty acid synthesis and diverting fatty acids from membrane lipids to triacylglycerol in Arabidopsis leaves. *Plant Cell*, **25**, 3506–18.

Glukhova, A., Hinkovska-Galcheva, V., Kelly, R., Abe, A., Shayman, J.A. and Tesmer, J.J. (2015) Structure and function of lysosomal phospholipase A_2 and lecithin:cholesterol acyltransferase. *Nat Commun*, **6**, 6250.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Asessing the performance of PhyML 3.0. *Syst. Biol.*, **59**, 307–321.

Heckman, K.L. and Pease, L.R. (2007) Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat Protoc*, **2**, 924–932.

Hu, Q., Sommerfeld, M., Jarvis, E., Ghirardi, M., Posewitz, M., Seibert, M. and Darzins, A. (2008) Microalgal triacylglycerols as feedstocks for biofuel production: Perspectives and advances. *Plant J.*, **54**, 621–639.

Hughes, A.L. (2007) Looking for Darwin in all the wrong places: The misguided quest for positive selection at the nucleotide sequence level. *Heredity (Edinb).*, **99**, 364–373.

Jones, P., Binns, D., Chang, H.Y., et al. (2014) InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, **30**, 1236–1240.

Kelly, L.A., Mezulis, S., Yates, C., Wass, M. and Sternberg, M. (2015) The Phyre2 web portal for protein modelling, prediction, and analysis. *Nat. Protoc.*, **10**, 845–858.

Klein, H.G., Lohse, P., Duverger, N., Albers, J.J., Rader, D.J., Zech, L.A., Santamarina-Fojo, S. and Brewer, H.B. (1993) Two different allelic mutations in the lecithin:cholesterol acyltransferase (LCAT) gene resulting in classic LCAT deficiency: LCAT (tyr^{83}→stop) and LCAT (tyr^{156}→asn). *J. Lipid Res.*, **34**, 49–58.

Leliaert, F., Smith, D.R., Moreau, H., Herron, M.D., Verbruggen, H., Delwiche, C.F. and Clerck, O. De (2012) Phylogeny and molecular evolution of the green algae. *CRC. Crit. Rev. Plant Sci.*, **31**, 1–46.

Liu, J., Han, D., Yoon, K., Hu, Q. and Li, Y. (2016) Characterization of type 2 diacylglycerol

acyltransferases in *Chlamydomonas reinhaardtii* reveals their distinct substrate specificities and functions in triacylglycerol biosynthesis. *Plant J.*, **86**, 3–19.

Liu, X.-Y., Ouyang, L.-L. and Zhou, Z.-G. (2016) Phospholipid: diacylglycerol acyltransferase contributes to the conversion of membrane lipids into triacylglycerol in Myrmecia incisa during the nitrogen starvation stress. *Sci. Rep.*, **6**, 26610.

Mao, X., Wu, T., Kou, Y., Shi, Y., Zhang, Y. and Liu, J. (2019) Characterization of type I and type II diacylglycerol acyltransferases from the emerging model alga *Chlorella zofingiensis* reveals their functional complementarity and engineering potential. *Biotechnol. Biofuels*, **12**, 28.

Meng, M., Bagdasarian, M. and Zeikus, J.G. (1993) The role of active-site aromatic and polar residues in catalysis and substrate discrimination by xylose isomerase. *Proc. Natl. Acad. Sci. U. S. A.*, **90**, 8459–8463.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D. and Thomas, P.D. (2017) PANTHER version 11: Expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.*, **45**, D183–D189.

Morris, J.L., Puttick, M.N., Clark, J.W., et al. (2018) The timescale of early land plant evolution. *Proc. Natl. Acad. Sci.*, **115**, E2274–E2283.

Ollis, D.L., Cheah, E., Cygler, M., et al. (1992) The α/β hydrolase fold. *Protein Eng.*, **5**, 197–211.

Pan, X., Peng, F.Y. and Weselake, R.J. (2015) Genome-wide analysis of *PHOSPHOLIPID : DIACYLGLYCEROL ACYLTRANSFERASE (PDAT)* genes in plants reveals the eudicot-wide *PDAT* gene expansion and altered selective pressures acting on the core eudicot PDAT paralogs. *Plant Physiol.*, **167**, 887–904.

Pan, X., Siloto, R.M.P., Wickramarathna, A.D., Mietkiewska, E. and Weselake, R.J. (2013) Identification of a pair of phospholipid:diacylglycerol acyltransferases from developing flax (*Linum usitatissimum L.*) seed catalyzing the selective production of trilinolenin. *J. Biol. Chem.*, **288**, 24173–24188.

Panchy, N., Lehti-Shiu, M.D. and Shiu, S.-H. (2016) Evolution of gene duplication in plants. *Plant Physiol.*, **171**, 2294–2316.

Peelman, F., Vinaimont, N., Verhee, A., et al. (1998) A proposed architecture for lecithin cholesterol acyl transferase (LCAT): Identification of the catalytic triad and molecular modeling.

Protein Sci., **7**, 587–599.

Piper, D.E., Romanow, W.G., Gunawardane, R.N., et al. (2015) The high-resolution crystal structure of human LCAT. *J Lipid Res*, **56**, 1711–1719.

Roberge, M., Shareck, F., Morosoli, R., Kluepfel, D. and Dupont, C. (1999) Characterization of active-site aromatic residues in xylanase A from *Streptomyces lividans*. *Protein Eng.*, **12**, 251–257.

Ronquist, F., Teslenko, M., Mark, P. Van Der, et al. (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.*, **61**, 539–542.

Sandager, L., Gustavsson, M.H., Stahl, U., Dahlqvist, A., Wiberg, E., Banas, A., Lenman, M., Ronne, H. and Stymne, S. (2002) Storage lipid synthesis is non-essential in yeast. *J Biol Chem*, **277**, 6478–6482.

Ståhl, U., Carlsson, A.S., Lenman, M., Dahlqvist, A., Huang, B., Banás, W., Banás, A. and Stymne, S. (2004) Cloning and functional characterization of a phospholipid:diacylglycerol acyltransferase from Arabidopsis. *Plant Physiol.*, **135**, 1324–1335.

Terashima, M., Specht, M. and Hippler, M. (2011) The chloroplast proteome: A survey from the *Chlamydomonas reinhardtii* perspective with a focus on distinctive features. *Curr. Genet.*, **57**, 151–168.

Wang, J., Gebre, A.K., Anderson, R.A. and Parks, J.S. (1997) Amino acid residue 149 of lecithin:cholesterol acyltransferase determines phospholipase A_2 and transacylase fatty acyl specificity. *J. Biol. Chem.*, **272**, 280–286.

Waterhouse, A., Bertoni, M., Bienert, S., et al. (2018) SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.*, **46**, W296–W303.

Weadick, C.J. and Chang, B.S.W. (2012) An improved likelihood ratio test for detecting site-specific functional divergence among clades of protein-coding genes. *Mol. Biol. Evol.*, **29**, 1297–1300.

Xu, Y., Caldo, K.M.P., Pal-Nath, D., Ozga, J., Lemieux, M.J., Weselake, R.J. and Chen, G. (2018) Properties and biotechnological applications of acyl-CoA:diacylglycerol acyltransferase and phospholipid:diacylglycerol acyltransferase from terrestrial plants and microalgae. *Lipids*, **53**, 663–688.

Yang, Y. and Benning, C. (2018) Functions of triacylglycerols during plant development and stress. *Curr. Opin. Biotechnol.*, **49**, 191–198.

Yang, Z. (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.*, **24**, 1586–1591.

Yang, Z. and Nielsen, R. (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.*, **19**, 908–917.

Yang, Z., Nielsen, R., Goldman, N. and Pedersen, A.-M.K. (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, **155**, 431–449.

Yang, Z., Wong, W.S.W. and Nielsen, R. (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.*, **22**, 1107–1118.

Yoon, K., Han, D., Li, Y., Sommerfeld, M. and Hu, Q. (2012) Phospholipid:diacylglycerol acyltransferase is a multifunctional enzyme involved in membrane lipid turnover and degradation while synthesizing triacylglycerol in the unicellular green microalga Chlamydomonas reinhardtii. *Plant Cell*, **24**, 3708–3724.

Zhang, J., Nielsen, R. and Yang, Z. (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.*, **22**, 2472–2479.

Zhang, M., Fan, J.L., Taylor, D.C. and Ohlrogge, J.B. (2009) *DGAT1* and *PDAT1* acyltransferases have overlapping functions in *Arabidopsis* triacylglycerol biosynthesis and are essential for normal pollen and seed development. *Plant Cell*, **21**, 3885–3901.

FIGURE LEGENDS

Figure 1. Number of lecithin:cholesterol acyltransferase (LCAT) family sequences in algae and land plants. The sequences were identified using BLAST searches with *Chlamydomonas reinhardtii* and *Arabidopsis thaliana* phospholipid:diacylglycerol acyltransferases (PDATs) as queries. Sequences not used in further analyses are not included. PSAT, phospholipid:sterol O-acyltransferase.

Figure 2. Representative elements of phospholipid:diacylglycerol acyltransferase (PDAT) structure. (**A)** Sequence alignment of representative PDATs with human lecithin:cholesterol acyltransferase (HsaLCAT). Alignment conservation is illustrated above the diagram, and the lid region, salt bridge and catalytic triad are indicated beneath with their locations denoted with red lines. (**B)** Superposition of the predicted structures of HsaLCAT, *Chlamydomonas reinhardtii* PDAT (CrPDAT) and *Arabidopsis thaliana* PDAT1 (AtPDAT1). The "lid" region, salt bridge, and catalytic triad are shown in red, orange and green, respectively. Arg, arginine (R); Asp, aspartic acid (D); His, histidine (H); Ser, serine (S); Trp, tryptophan (W).

Figure 3. Maximum likelihood (ML) phylogenetic tree of algal and plant lecithin-cholesterol acyltransferase (LCAT)-family enzymes. Branch-supporting values are shown on the branches. PDAT, phospholipid:diacylglycerol acyltransferase; PLA, phospholipase; PSAT, phospholipid:sterol O-acyltransferase.

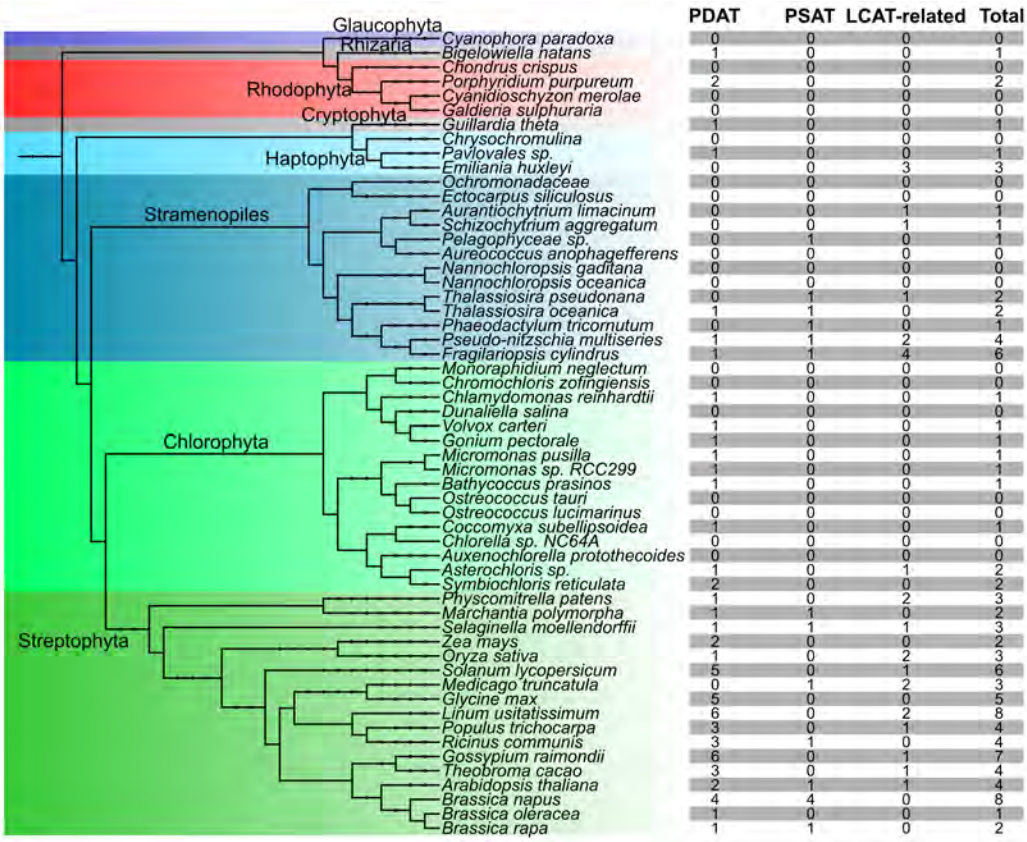
Figure 4. Positive selection sites identified in phospholipid:diacylglycerol acyltransferases (PDATs) from green algae and Group A plants. Positive selection sites in PDATs from representative green algae and Group A plants are highlighted in red and blue, respectively. The alignment consensus is shown on the top and the red bars below it represent the amino acid sequences in the alignment beneath with different fragments separated by black vertical lines. The amino acids of the catalytic triad (S-D-H) are marked with a circle (●). The amino acid position further studied in *Arabidopsis thaliana* PDAT1 (AtPDAT1) is marked with a star (★).

Figure 5. Predicted locations and interactions of positive selection sites with other amino acids in *Arabidopsis thaliana* PDAT1 (AtPDAT1). (A) AtPDAT1 structure showing the location of positive selection sites in orange. The catalytic triad is shown in green, red and blue. **(B)** Representative positive selection sites that may influence or stabilize secondary structure through polar contacts with the polypeptide backbone. **(C)** Representative positive selection sites that may influence or stabilize tertiary structure through polar contacts. The distance between the atoms participating in polar contacts (yellow dash) is indicated in angstroms. Arg, arginine (R); Asp, aspartic acid (D); Gln, glutamine (Q); Glu, glutamic acid (E); Gly, glycine (G); Leu, leucine (L); Met, methionine (M); Phe, phenylalanine (F); Ser, serine (S); Thr, threonine (T).

Figure 6. Analysis of M255 in *Arabidopsis thaliana* PDAT1 (AtPDAT1). **(A)** AtPDAT1 homology model obtained using SWISS-MODEL software. The predicted N-terminal transmembrane region is added to the structure. **(B)** Close-up view of the active site pocket of the wild-type (WT) AtPDAT1 wherein the side chains for the catalytic triad (S254, D573, and H626) are shown in different colors with green, red and blue corresponding to carbon, oxygen and nitrogen, respectively. The positive selection site (M255) is shown in yellow. **(C)** and **(D)** Close-up view of active site pockets of the AtPDAT1-M255W and -M255A mutants. **(E)** Alignment of the three models (WT, M255W and M255A) shows that the position and polar contacts made by the catalytic triad appears to be unchanged during the mutation of M255. The carbon atoms of the 3 models are shown in different shades of green whereas nitrogen and oxygen atoms are shown in blue and red, respectively. The polar contacts (orange dash) among the catalytic triad are indicated in angstroms. **(F)** M255 appears to form a network of hydrophobic contacts (brown dash in angstroms) with non-polar residues (Ile116 and Met351) lining the active site. The three non-polar residues may form a hydrophobic patch in the active site pocket. **(G)** The van der Waals surface of the non-polar residues shown in dots. **(H-J)** Surface view of the active site pocket wherein the M255, Ile116, and Met351 (shown in yellow) are found to form a hydrophobic patch with two ridges (indicated by arrows). The catalytic triads are shown in green.

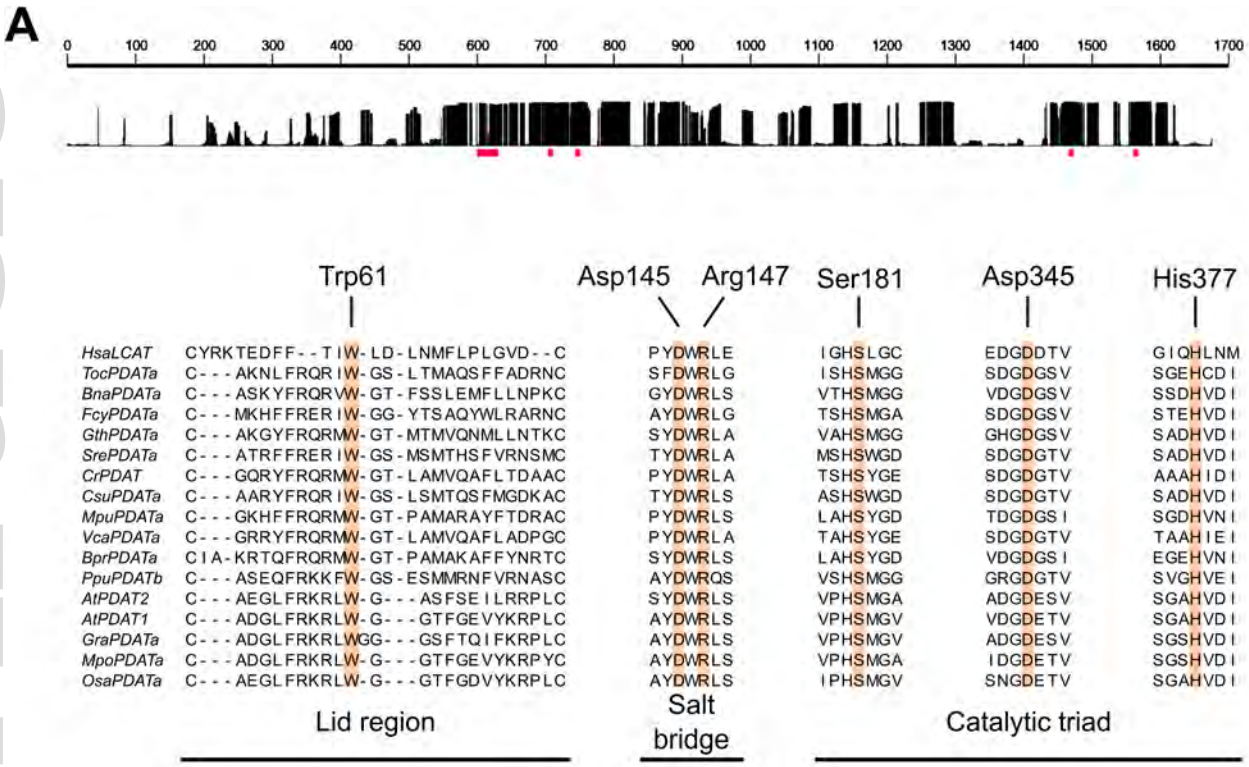
Figure 7. Assessment of the catalytic activity of microsomes containing *LacZ*, AtPDAT1, AtPDAT1-M255W and AtPDAT1-M255A. Asterisks indicate statistically significant differences in the production of NBD-TAG between yeast expressing AtPDAT1 (WT) and the mutants or LacZ (*, $p < 0.05$; **, $p < 0.01$). Ala, alanine (A); Asp, Aspartic acid (D); His, Histidine (H); Ile, Isoleucine (I); Met, Methionine (M) Ser, Serine (S); Trp, Tryptophan (W)

Figure 8. Conservation of amino acids at the position +1 downstream of the catalytic serine in animal, fungi, plant and green algae LCAT-related enzymes. To understand the conservation of this residue over a broader range of organisms, a phylogenetic tree was built with available LCAT family proteins of various organisms including sponge (*Amphimedon queenslandica)*, fish (*Danio rerio)*, model worm (*Caenorhabditis elegans)*, fly (*Drosophila melanogaster*), horse (*Equus caballus*), domestic chicken (*Gallus gallus*), frog (*Xenopus tropicalis*), mouse (*Mus musculus*), human, filamentous fungi (*Aspergillus nidulans*), yeast (*Saccharomyces cerevisiae*, ScePDATa), plant (*Arabidopsis thaliana*) and green microalga (*Chlamydomonas reinhardtii*). Methionine (M), leucine (L) and tyrosine (Y) are shown in black, blue and red in the alignment, respectively. LCAT, lecithin:cholesterol acyltransferase; PDAT, phospholipid:diacylglycerol acyltransferase; PLA, phospholipase A.

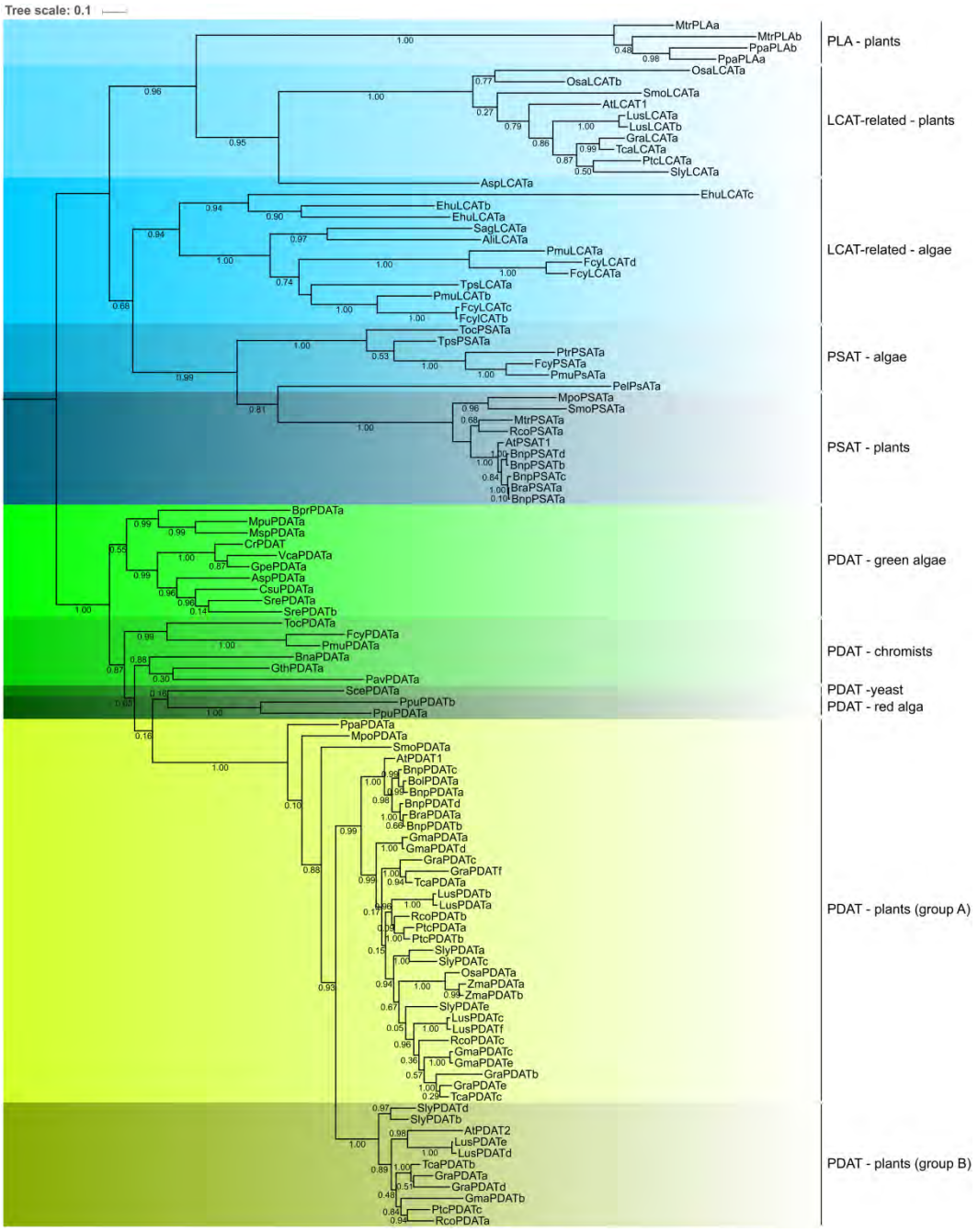


PDAT	PSAT	LCAT-related	Total
0	0	0	0
1	0	0	1
0	0	0	0
2	0	0	2
0	0	0	0
0	0	0	0
1	0	0	1
0	0	0	0
1	0	0	1
0	0	3	3
0	0	0	0
0	0	0	0
0	0	1	1
0	0	1	1
0	1	0	1
0	0	0	0
0	0	0	0
0	0	0	0
0	1	1	2
1	1	0	2
0	1	0	1
1	1	2	4
1	1	4	6
0	0	0	0
0	0	0	0
1	0	0	1
0	0	0	0
1	0	0	1
1	0	0	1
1	0	0	1
1	0	0	1
0	0	0	0
0	0	0	0
1	0	0	1
0	0	0	0
1	0	1	2
2	0	0	2
1	0	2	3
1	1	0	2
1	1	1	3
2	0	0	2
1	0	2	3
5	0	1	6
0	1	2	3
5	0	0	5
6	0	2	8
3	0	1	4
3	1	0	4
6	0	1	7
3	0	1	4
2	1	1	4
4	4	0	8
1	0	0	1
1	1	0	2

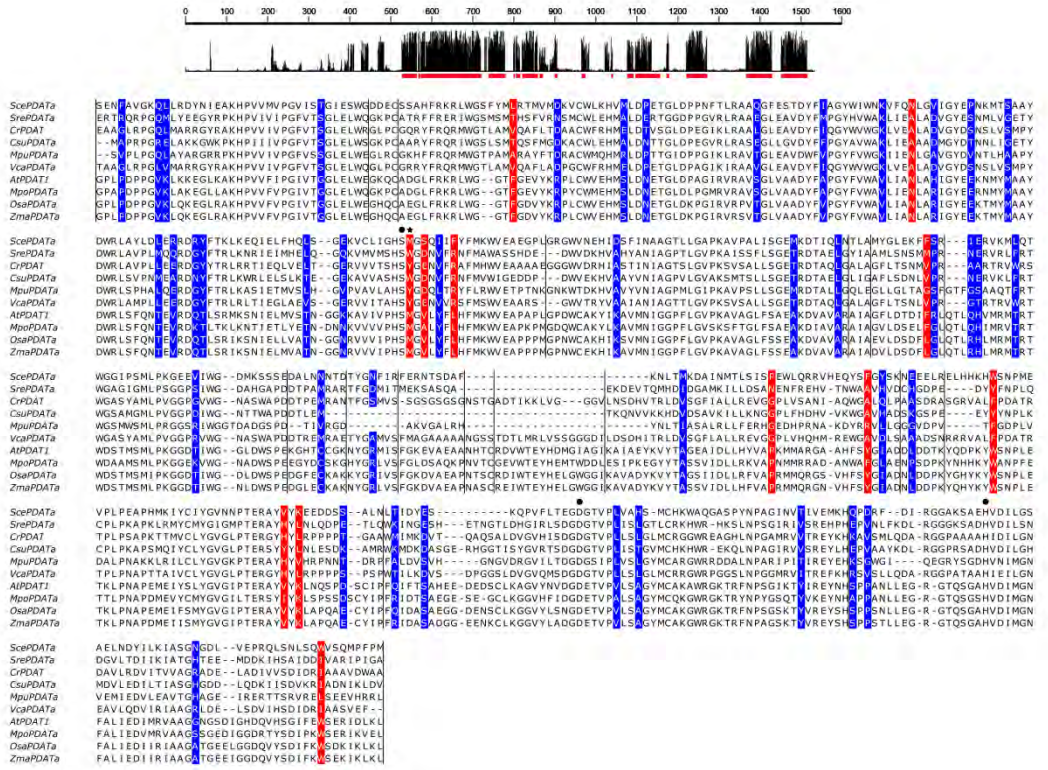
tpj_14880_f1.tif



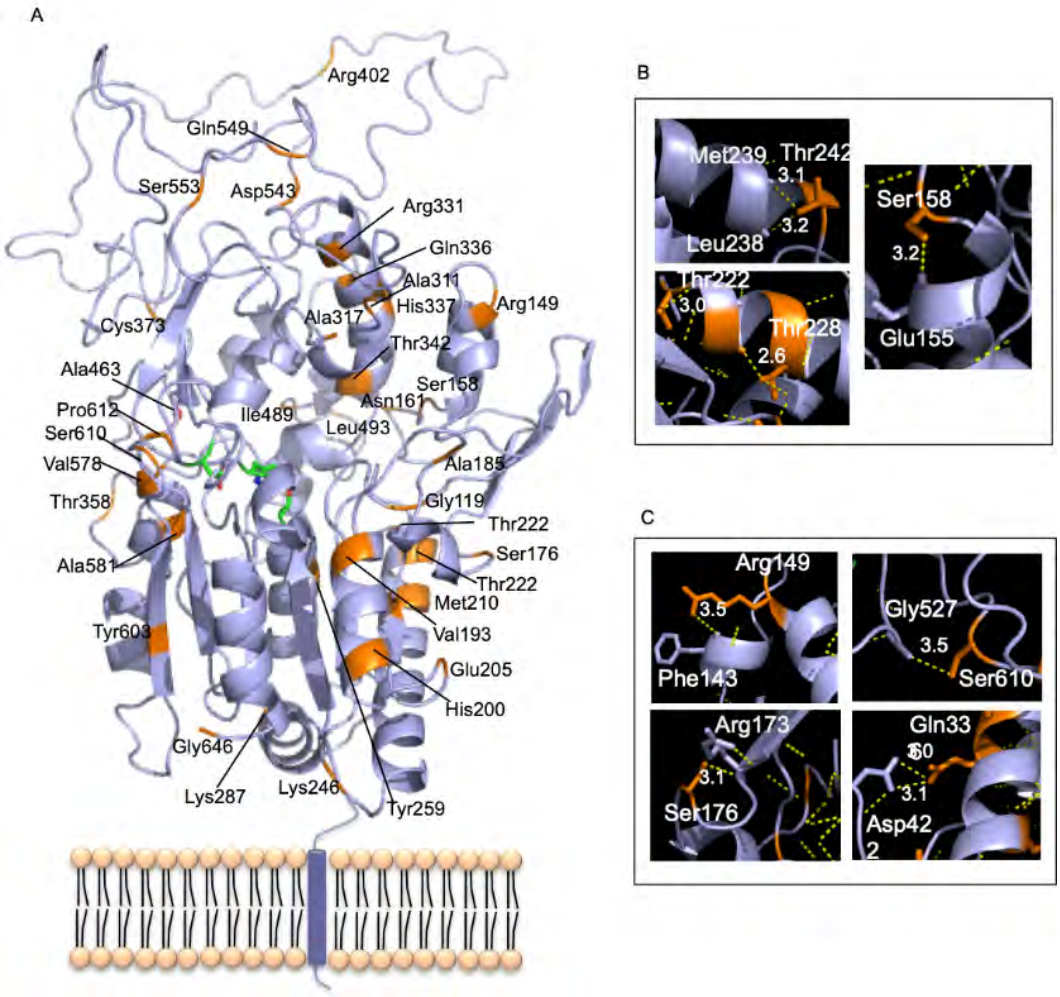
tpj_14880_f2.tif



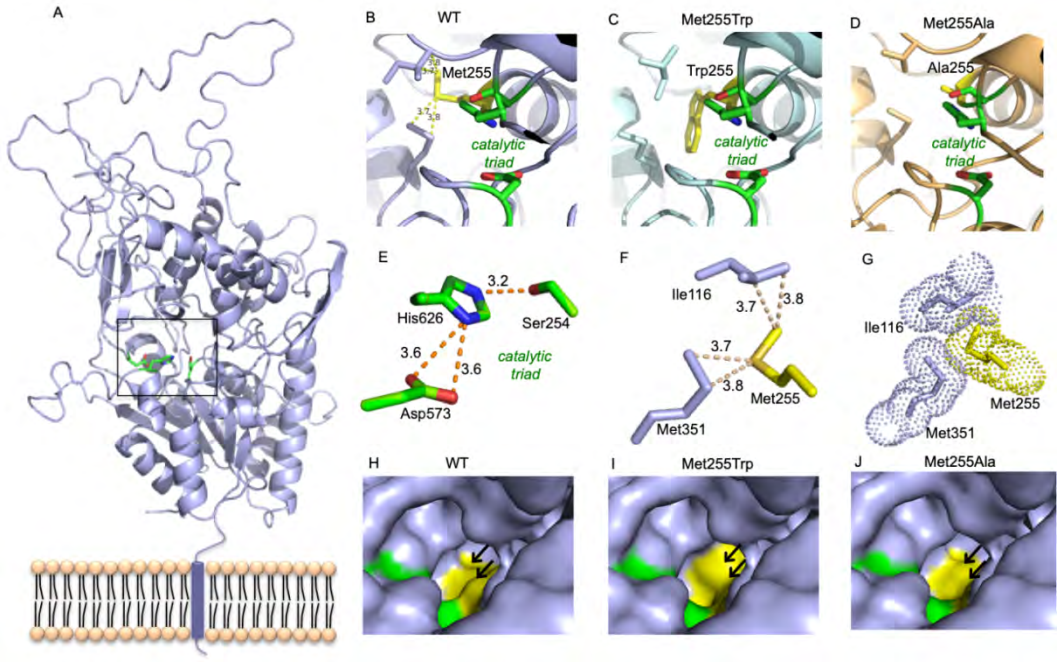
tpj_14880_f3.tif



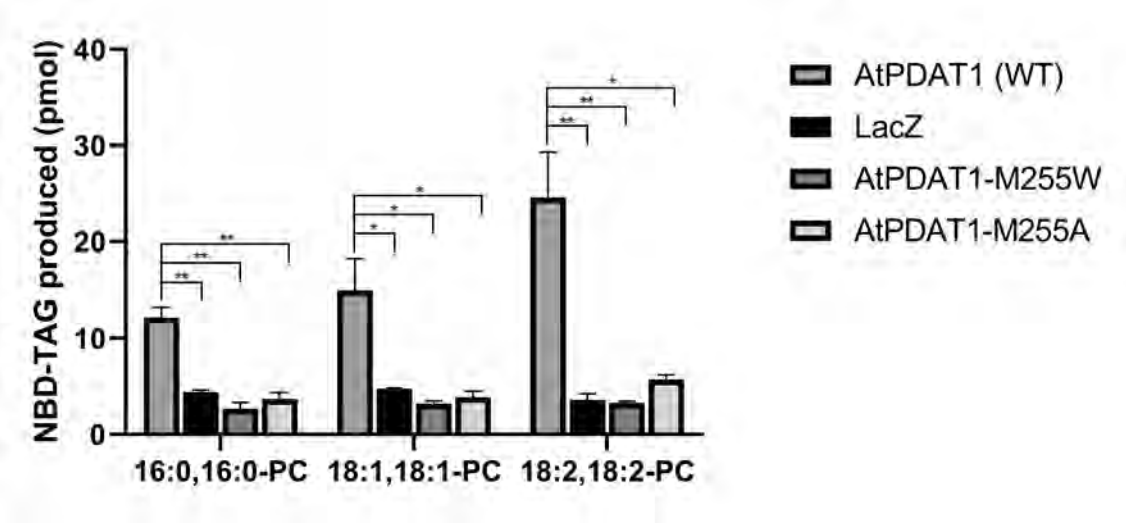
tpj_14880_f4.tif



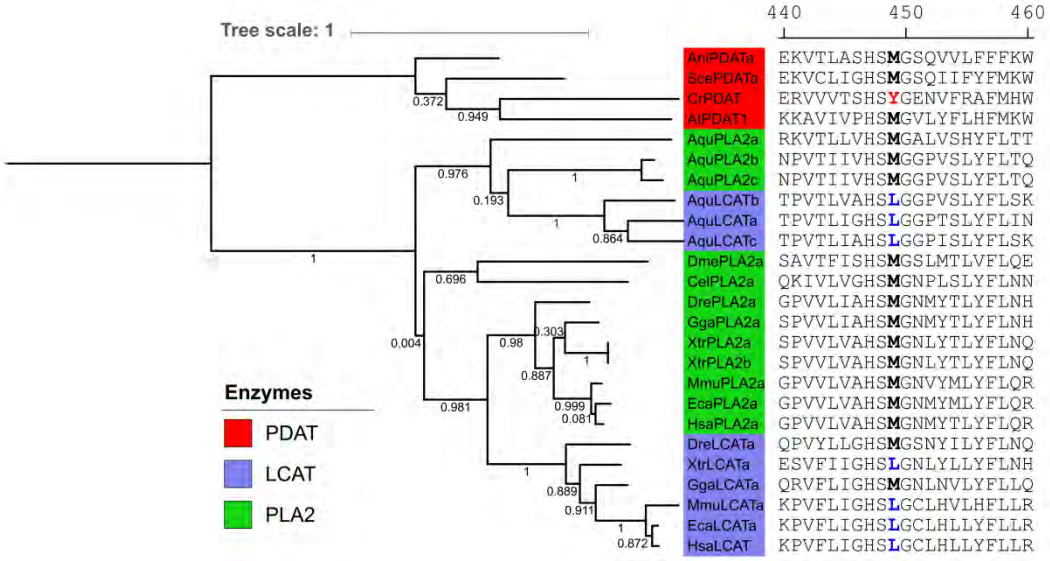
tpj_14880_f5.tif



tpj_14880_f6.tif



tpj_14880_f7.jpg



tpj_14880_f8.tif