STRUCTURAL AND EVOLUTIONARY CHARACTERIZATION OF PHOSPHOLIPID:DIACYLGLYCEROL ACYLTRANSFERASE IN PHOTOSYNTHETIC ORGANISMS

by

Lucas José Falarz

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Biological Sciences

University of Manitoba

Winnipeg

ABSTRACT

Phospholipid:diacylglycerol acyltransferase (PDAT) catalyzes the last step in acyl-CoA independent triacylglycerol biosynthesis in algae and higher plants. Although PDAT has been characterized in some algal and plant species, the evolution and structural properties of this important enzyme at the broad level of photosynthetic organisms is yet to be studied. In this study, a reliable fluorescence-based PDAT assay was first developed to replace the costly and lab-extensive radiolabelling method. The novel method, as well as various biotechnological and *in silico* analyses, were then used to explore the structural and evolutionary properties of PDATs. The results showed that functional divergence and positive selection present in the evolution of PDAT in green algae and plants. The identified positive selection sites are important in PDAT activity and thus would be candidate sites for modifying PDAT via protein engineering. In addition, PDAT evolved to be more important for plant metabolism and fitness than algae.

ACKNOWLEDGEMENTS

I would like to thank Dr. Guanqun (Gavin) Chen for accepting me as his student and for the support, mentoring, friendship and opportunities during my studies. His guidance helped improve my critical thinking and my abilities as a researcher. I would also like to thank Dr. Colin Garroway and Dr. Pingzhao Hu for the support in this project and for participating in the supervisory committee.

I would like to thank the current and the past members from Dr. Chen's lab: Dr. Yang (Sarena) Xu, Kethmi Jayawardhane, Dr. Kristian Caldo, Dr. Saleh Shah, Junhao Lu, and Trinh Nguyen. Special thanks to Sarena and Kethmi for teaching me the techniques that were so valuable in my research. Also, thanks to Kristian for helping me when I started in the laboratory. I would also like to thank Dr. Stacy Singer for the feedback and contributions to the project. Special thanks to Dr. Chen, Dr. Singer and Dr. Randall Weselake for allowing me to contribute to the book "*Plant Bioproducts*".

I would like to thank the staff from the Department of Biological Sciences at the University of Manitoba, Faculty of Graduate Studies, and from the Department of Agricultural, Food & Nutritional Science from the University of Alberta for their contribution and support in my graduate life and research. Special thanks to the Natural Sciences and Engineering Research Council of Canada (NSERC), University of Manitoba Graduate Enhancement of Tri-Council Stipends (GETS) program, Department of Biological Sciences, and Faculty of Graduate Studies for providing funding for my research and for the participation in scientific conferences.

I thank my family in Brazil for their support and especially for their understanding when I could not be present on many occasions because I am more than ten thousand kilometers away. I would especially like to thank my wife Juliana for sharing the idea of going abroad for my studies and for all the love, support and patience during this project.

This work is dedicated to the loving memory of my father,

Eugênio José Soares Falarz

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

+G: rate heterogeneity among sites

+I: rate of invariable sites

16:0,[^{14}C]18:1-PtdCho: 1-palmytoil-2-[^{14}C]oleoyl-*sn*-3-glycero-phosphocholine

16:0,16:0-PtdCho: 1,2-dihexadecanoyl-*sn*-glycero-3-phosphocholine

16:0,18:1-DAG: 1-palmitoyl-2-oleoyl-*sn*-glycerol

18:1,18:1-PtdCho: 1,2-dioleoyl-*sn*-glycero-3-phosphocholine

18:2,18:2-PtdCho: 1,2-dilinoleoyl-*sn*-glycero-3-phosphocholine

ACCase: acetyl-CoA carboxylase

ACP: acyl carrier protein

AtPDAT1: *Arabidopsis thaliana* phospholipid:diacylglycerol acyltransferase 1

BEB: Bayes empirical Bayes

BI: Bayesian inference

BLAST: basic local alignment search tool

BnDGAT1: *Brassica napus* acyl-CoA:diacylglycerol acyltransferase 1

CO: carbon monoxide

CO_2: carbon dioxide

CoA: Co-enzyme A

DAG: diacylglycerol

df: degree of freedom

DGAT: acyl-CoA:diacylglycerol acyltransferase

DHA: docosahexaenoic acid

dN: non-synonymous substitutions

dS: synonymous substitutions

EAR: enoyl-ACP reductase

ER: endoplasmic reticulum

ESS: estimated sample size

FAS: fatty acid synthase

FAT: acyl-ACP thioesterase

G3P: glycerol-3-phosphate

GHG: greenhouse gas

GPAT: glycerol-3-phosphate acyltransferase

GTR: general time reversible

HD: hydroxyacyl-ACP dehydrase

HDL: high-density lipoprotein

KAR: ketoacyl-ACP reductase

KAS: ketoacyl-ACP synthase

ℓ: log-likelihood

LACS: long-chain acyl-CoA synthase

LCAT: lecithin:cholesterol acyltransferase

LD: lipid droplet

LPA: lysophosphatidic acid

LPAT: lysophosphatidic acid acyltransferase

LRT: likelihood ratio test

MCMT: malonyl-CoA:ACP malonyltransferase

MGDG: monogalactosyldiacylglycerol

MHP: mammalian heme peroxidase

ML: maximum likelihood

MPO: myeloperoxidase

MSA: multiple sequence alignment

MW: molecular weight

NBD: nitrobenzoxadiazole

NBD-DAG: 1-palmitoyl-2-dodecanoyl-nitrobenzoxadiazole-*sn*-glycero-3-glycerol

NBD-PtdCho: 1-palmitoyl-2-dodecanoyl-NBD-*sn*-glycero-3-phosphocholine

NJ: neighbor-joining

NNI: nearest neighbor interchange

PA: phosphatidic acid

PAP: phosphatidic acid phosphatase

PDAT: phospholipid:diacylglycerol acyltransferase

pI: isoelectric point

PLA: phospholipase A

PLA2: phospholipase A2

PMT: photomultiplier tube

PMT: photomultiplier tube

PSAT: phospholipid:sterol O-acyltransferase

PSRF: potential scale reduction factor

PtdCho: phosphatidylcholine

PtdEth: phosphatidylethanolamine

SAD: stearoyl-ACP desaturase

SH-aLRT: Shimodaira-Hasegawa approximate likelihood ratio test

SPR: subtree prune and regraft

TAG: triacylglycerol

TLC: thin-layer chromatography

TMD: transmembrane domain

WT: wild type

FOREWORD

Chapter 1 of this thesis includes the thesis introduction and the literature review. Chapters 2 and 3 contain the results of the work developed during my studies. The title of Chapter 2 is *"A fluorescence-based assay for quantitative analysis of phospholipid:diacylglycerol acyltransferase activity*". A manuscript with the results in Chapter 2 has been submitted for publication in Lipids. Dr. Yang (Sarena) Xu, Dr. Stacy Singer, and Dr. Guanqun (Gavin) Chen contributed to this work. The title of Chapter 3 is *"The evolution of PDAT in photosynthetic organisms was marked by functional divergence and positive selection which were essential to higher plants but not green microalgae"*, which will be submitted later.

CHAPTER 1. Thesis introduction

Bioproducts are derived from renewable biomass, which can be obtained, among other sources, from plants and algae (Weselake et al., 2018). Some of these organisms accumulate high contents of carbohydrates, proteins, lipids, and other metabolites, which allow their commercial exploitation. The components of the biomass can be used as feedstock for a variety of products, such as biofuels, bioplastics, pharmaceuticals, adhesives, and other materials and chemicals (Arif et al., 2018; Chen et al., 2019; Dyer et al., 2008; Falarz et al., 2018; Hu et al., 2008; Markou and Nerantzis, 2013; Singer and Weselake, 2018a). For instance, green algae *Haematococcus pluvialis* and *Chromochloris zofingiensis* can accumulate astaxanthin, a high-value carotenoid with strong antioxidant activity which is suitable as a nutraceutical with anti-cancer, anti-diabetes, and anti-inflammatory properties (Ambati et al., 2014; Fakhri et al., 2018; Han et al., 2013; Higuera-Ciapara et al., 2006; Roth et al., 2017).

Plant- and algal-based bioproducts may reduce the negative impacts of fossil-fuel based products in society. Among them, climate change is the most remarkable, which is caused by the accumulation of high amounts of carbon dioxide (CO_2) and other greenhouse gases (GHGs; Allen et al., 2014). GHGs accumulation is mostly driven by fossil fuel extraction and consumption. To avoid further increases in the accumulation of these gases, new actions and products are necessary. The biomass from plants and algae could alleviate the burden of GHG emissions because these organisms recover CO_2 during photosynthesis, which could thus become a sustainable alternative of fossil fuel.

Vegetable oils are a source of energy and have the potential to be transformed into environment-friendly products. Indeed, seed oils have already been converted into biodiesel. Brazil, for instance, produces 4.3 million m^3 of biodiesel and 65% is derived from soybeans

(*Glycine max*; Empresa de Pesquisa Energética - EPE, 2018). Algae are also able to accumulate a high content of lipids and have great potential to provide materials for the biodiesel industry. For instance, *Nannochloropsis salina* is able to accumulate 46% of dry cell weight as lipids under specific growth conditions (Liu et al., 2018). In addition to serving as biodiesel precursor, algal and plant lipids can also be used as food products, animal feed, pharmaceuticals, lubricants, and chemical intermediates (Dyer et al., 2008; McKeon, 2016; Spolaore et al., 2006).

As many algal and plant species can accumulate high amounts of lipids in their cells or certain organs, lipids may play an important role in the evolution of algae and plants. The relevance of oil accumulation in algae and higher plants and the need for diversified compounds are present in their genome and in the regulation of lipid-related genes. The final step of the biosynthesis of the major storage lipid triacylglycerol (TAG) can be catalyzed by at least four enzymes, acyl-CoA:diacylglycerol acyltransferases type 1, type 2 and type 3 (DGAT1, DGAT2 and DGAT3, respectively) and the phospholipid:diacylglycerol acyltransferase (PDAT) which share no homology (Liu et al., 2012; Xu et al., 2018a). Moreover, their copy numbers vary largely among species and can be very high in some species. For instance, flax (*Linum usitatissimum*) has six *PDAT* copies (Pan et al., 2015), while the microalga *Nannochloropsis oceanica* may have 15 copies of TAG-synthesizing genes (eleven *DGAT2,* two *DGAT1* and two *PDATs;* Vieler et al., 2012; Wang et al., 2014). Another particularity of some species is their ability to preferentially accumulate certain fatty acids, which are important for their application as bioproducts and are related to DGAT and PDAT. The related genes from those species could be used in genetic engineering of other species to modify their oil profile. For instance, heterologous coexpression of *Ricinus communis PDAT1A* and *FATTY ACID HYDROXYLASE12* in *Arabidopsis thaliana* seeds resulted in oil containing 27% hydroxylated fatty acids (van Erp et

al., 2011). Such diversity in the lipid metabolism might reflect the relevance of lipids in the evolution of both algae and plants and directly associate with their survival and fitness.

A better understanding of the composition of genomes and the genetic regulation of lipid in algal and plant species would not only expand our knowledge of the lipid-related genes in the survival and fitness of algae and plants, but also allow us to efficiently manipulate lipid production. The results would help increase total lipid contents or produce special fatty acids in algae and plants by molecular biotechnology. Therefore, it is necessary to understand the complexity of lipid metabolism in plants and algae from both evolution and biotechnology perspectives (Chen et al., 2019). In this study, I combined *in silico* analysis and on-bench experiments to characterize PDATs, a class of key enzymes in TAG biosynthesis as mentioned above, in algae and plants. I firstly developed a faster, cheaper and reliable fluorescence-based methodology to characterize PDATs and then combined this methodology with bioinformatics to characterize the structure and evolution of PDATs from algal and plant species.

1.1. Literature review

The literature review section will briefly introduce relevant information to help understand the relationship among the organisms studied, their economic importance, and the materials and methodologies applied in this study. Section 1.1.1 will show how algal and plant lipids can be exploited as bioproducts. The evolution of the main groups of algae and plants will be succinctly introduced in Section 1.1.2. In addition, the lipid metabolism in these organisms is explained in Section 1.1.3. Section 1.1.4 will briefly present bioinformatics methods, databases and programs commonly used for homologous sequence search, multiple sequence alignment, phylogenetic analysis, and selection pressure analysis. Finally, Section 1.1.5 introduces the basics of

enzymatic assays, the study of membrane enzymes and the application of fluorescent chemicals in enzymatic studies.

1.1.1. Potential of algae and plants as sources of biodiesel and other lipid bioproducts

Biomass from algae and plants are sources of different compounds, which may have high value or be used in the manufacturing of various bioproducts. Many species are able to accumulate lipids in high contents. Canola (*Brassica napus*), soybean, sunflower (*Helianthus sp.*), and flax can accumulate lipids in their seeds and are currently cultivated for the production of vegetable oil and other products (such as protein and meals; Singer and Weselake, 2018a; Weselake et al., 2017). Palm (*Elaeis guineensis*) also accumulates oil in the mesocarp and kernel and is largely cultivated in Southeast Asia. Many oil crops are used in the production of vegetable oil for food usage. Some oil plant species also have commercial potential due to their ability to synthesize unusual fatty acids, which have uncommon or modified fatty acid chain and special application in the oleochemical industry. For instance, castor (*R. communis*) and tung (*Vernicia fordii*) seed oils contain hydroxy fatty acids and α-eleostearic acids, respectively, with high industry values (Dyer et al., 2002; McKeon, 2016; Shockey et al., 2006). Unlike plants that accumulate most of the oil in one or a few tissues (seed, mesocarp, kernel, etc.), many algal species are unicellular organisms which store TAG in specialized structures known as lipid droplets in the whole cell (Xu et al., 2016). The biosynthesis and storage of oils in microalgae usually occur when the cells are grown under adverse conditions, such as excessive light or nutrient deprivation (Han et al., 2013; Hu et al., 2008; Li-Beisson et al., 2019, 2015).

Biodiesel is an important bioproduct from vegetable oils. Biodiesel is produced from the transesterification of TAG, which reacts with an alcohol (usually methanol or ethanol) in the

presence of a catalyst (an acid or an alkali) to produce fatty acid methyl esters (biodiesel) and a by-product glycerol (Durrett et al., 2008). The properties of biodiesel, such as cetane number, cloud point, pour point, oxidative stability, heat of combustion and lubricity, are impacted by the fatty acid composition of the vegetable oil (Singer and Weselake, 2018b). In addition, the presence of unsaturation must be adjusted depending on the region it will be used. Biodiesel that is rich in α-linolenic acid is adequate for cold regions since the unsaturation decreases the cloud point, which is the temperature at which crystallization occurs in the fuel (Knothe, 2005; Singer and Weselake, 2018b). However, the unsaturation also decreases the cetane number, which measures the speed of combustion after fuel injection. The higher the cetane number, the higher the ignition quality. Therefore, biodiesel derived from oleic acid may be more adequate to meet biofuel regulations in some regions. Biodiesel has been mostly produced from plant oils and animal fat. Due to the high oil content and rapid growth of microalgae, algal oil has been used in biodiesel production. However, due to the variation of lipid content and fatty acid composition, not every plant or algal oil are suitable for biodiesel production. Therefore, the ability to manipulate the oil composition and content in plant and algae would greatly benefit the biodiesel industry.

Biodiesel derived from plant lipids has several advantages over fossil fuels. The primary benefits are cleaner production with lower GHG emissions and elevated energetic security since biodiesel production is not geographically limited (Weselake et al., 2018). Plant oils can be produced in high quantities since many crops have well-established cultivation methods. In addition, oil crops can produce other co-products (e.g. protein-rich grain meals), which can contribute to the economic viability of biodiesel projects. The major limitations of using seed oil

for biodiesel production are the competition with food crops, the seasonal variability, and the intensive use of fertilizers (Durrett et al., 2008; Weselake et al., 2018).

Many algal species can accumulate high amounts of lipids and thus have the potential to supply biodiesel production in the future. Algal lipids could be preferred to plants because algae grow faster and can accumulate oil rapidly. In addition, they can be cultivated in ponds or bioreactors setting up on marginal lands, which would release arable land for food crops. However, the viability of algal biodiesel projects is currently restricted, especially due to the low oil yields obtained in algal growth. Algal oil yield can be affected by algal biomass production and cellular oil content. Many algae need to be cultured under diluted medium with relatively low cell density under phototrophic conditions. But the diluted culture conditions increase the costs of algal biomass production because the recovery of biomass from diluted medium is expensive (Show et al., 2019). Moreover, it is attractive to increase oil content in algal cells which can increase algal oil yield. For this perspective, it is crucial to understand how TAG is produced in these organisms. Comparing with high plants, algae are relatively unexplored organisms and more information is necessary to design better technologies to allow algal biodiesel projects economically viable (Chen et al., 2019).

1.1.2. Evolution of algae and land plants

Green algae (Chlorophyta) and land plants (Embryophyta) were able to explore and adapt to all types of environments during their evolution and a diversified set of organisms was generated into each group. They belong to the same taxonomic group called Viridiplantae (Leliaert et al., 2012). Modern green algae and land plants have a common ancestor, which is an ancient green alga, which is the precursor of organisms that had different evolutionary paths down the road.

The term alga is not used for taxonomic characterization, but it usually comprises eukaryotic photosynthetic organisms that form a polyphyletic group without the differentiation into complex tissues (Barsanti and Gualtieri, 2014). Plants can be distinguished from algae because they exhibit differentiation into several tissues (e.g. leaves, xylem, roots). In addition, the life cycle of plants is well defined between haploid and diploid cells. In terms of classification, some important algal species such as the model alga *Chlamydomonas reinhardtii* and the carotenoid-rich *H. pluvialis* and *Dunaliella salina* are included in the Chlorophyta. Land plants are generally divided into bryophytes and vascular plants (Bowman et al., 2017). The latter group contains many plant species with great commercial interest, such as maize (*Zea mays*), rice (*Oryza sativa*), soybean, and potato (*Solanum tuberosum*). Vascular plants can be further subdivided into two groups, monocots, and eudicots.

Viridiplantae is a subclade of Archaeplastida, which also includes red algae (Rhodophyta) and Glaucophytes (Figure 1.1). In Archaeplastida, the organisms are considered primary endosymbionts, which means their ancestral was a heterotrophic organism that engulfed a cyanobacterium (De Clerck et al., 2012). In the primary endosymbiont, the cyanobacterium became the plastid that provides the photosynthetic ability to eukaryotes. The secondary and tertiary endosymbionts are the algae that engulfed a primary endosymbiont (Archibald, 2012), among which diatoms are a group of important species. Diatom belongs to the Stramenopila and has a plastid derived from a red alga. The representatives of secondary and tertiary endosymbionts are paraphyletic and are classified in the Kingdom Chromista (hence, their members are the chromists; Cavalier-Smith, 2018). Other relevant groups in this kingdom are the Haptophyta, Cryptophyta, Rhizaria, and Alveolata. *Nannochloropsis* sp., which is becoming a model algal species for the study of lipid biosynthesis, belongs to the Stramenopiles and is

suggested to have undergone multiple secondary endosymbioses, since it has genes from green algae and red algae in their genomes (Du and Benning, 2016; Wang et al., 2014).

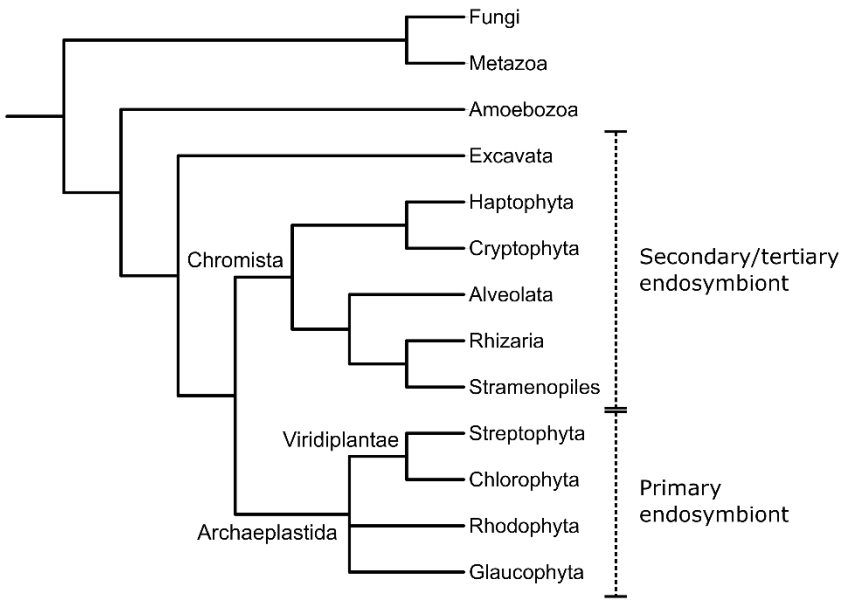


Figure 1.1. The relationship among primary endosymbionts, secondary/tertiary endosymbionts and other eukaryotic lineages. Based on (Archibald, 2012; De Clerck et al., 2012).

1.1.3. Lipid metabolism in algae and plants

Lipid metabolisms in algae and higher plants are similar in general but have many differentiations (Figure 1.2). The enzymes that participate in the lipid biosynthesis of algae and plants are usually similar. However, their regulation and expression are not. In addition, they may be present in different copy numbers or be expressed uniquely in some tissues. Several plants, for instance, accumulate a high content of oil in their seeds as the major storage compound (Xu et al., 2018a). Algae, in general, are able to accumulate a large amount of oil solely under adverse conditions. In both groups, lipid biosynthesis starts with the *de novo* fatty acid synthesis, in which the first committed step involves the carboxylation of acetyl-Coenzyme

A (CoA) with bicarbonate to produce malonyl-CoA (Li-Beisson et al., 2019; Ohlrogge and Browse, 1995). This reaction is catalyzed by acetyl-CoA carboxylase (ACCase), which has two forms: one comprises separated subunits (heteromeric) and one is a single polypeptide (homomeric). Then, malonyl from malonyl-CoA is transferred to an acyl carrier protein (ACP) to form malonyl-ACP, which is catalyzed by malonyl-CoA:ACP malonyltransferase (MCMT). The overexpression of *MCMT* in *A. thaliana* increased 20% of the oil content in seeds (Jung et al., 2019). Malonyl-ACP is then elongated through the addition of acetyl groups in a cycle synthesized by the fatty acid synthase (FAS) complex until palmitoyl-ACP is formed (Li-Beisson et al., 2019; Ohlrogge and Browse, 1995). In the process of palmitoyl-ACP synthesis, acetyl-CoA and malonyl-ACP are initially converted into 3-ketoacyl-ACP, which is catalyzed by ketoacyl-ACP synthase (KAS). This reaction is followed by a reduction, a dehydration, and a second reduction, which are catalyzed by ketoacyl-ACP reductase (KAR), hydroxyacyl-ACP dehydrase (HD) and enoyl-ACP reductase (EAR), respectively. The acyl-ACP formed will re-enter the cycle through the addition of a new acetyl-CoA. Once palmitoyl-ACP is synthesized, it can be further elongated to stearoyl-ACP by the action of KASII and this compound can undergo a desaturation to produce stearoyl-ACP through the action of a stearoyl-ACP desaturase (SAD; Li-Beisson et al., 2019). Acyl-ACP thioesterase (FAT) can release the fatty acid from ACP, so the acyl chain can be transported from the plastid. The free fatty acids are then converted to acyl-CoA, which can be used in the biosynthesis of TAG and other lipids, in the outer envelope of the plastid with the enzyme long-chain acyl-CoA synthase (LACS; Li-Beisson et al., 2019; Li et al., 2016).

The sequential reactions leading to the formation of TAG are described in the Kennedy pathway (Li-Beisson et al., 2019; Liu et al., 2012; Weiss et al., 1960; Weiss and Kennedy, 1956;

Xu et al., 2018a). TAG is composed of a glycerol backbone with three esterified acyl chains. In the first reaction of TAG assembly, glycerol-3-phosphate (G3P) is acylated in the *sn*-1 position by glycerol-3-phosphate acyltransferase (GPAT), which forms *sn*-1 lysophosphatidic acid (LPA). Then, a lysophosphatidic acid acyltransferase (LPAT) transfers another acyl moiety to the *sn*-2 position of the *sn*-1 LPA to form phosphatidic acid (PA). In both cases, acyl-CoA is the acyl donor. In the next step, DAG is formed by the removal of the phosphate group by the action of the phosphatidic acid phosphatase (PAP). The last step in the TAG biosynthesis may vary, which depends on the substrate and the enzyme catalyzing the reaction such as DGAT and PDAT. DGATs transfer a fatty acid to the *sn*-3 position of DAG using acyl-CoA as the donor. Membrane phospholipids are the acyl donors in the reactions catalyzed by PDAT. The first mechanism is known as acyl-CoA-dependent, while the second is the acyl-CoA-independent. Most of these reactions take place in the endoplasmic reticulum. Parts of this organelle will detach to become the lipid droplets (LD), which are composed of TAG and phospholipid monolayers that surround the TAG (Li-Beisson et al., 2019; Xu et al., 2018a). In *C. reinhardtii*, the formation of diacylglycerol (DAG) also occurs in the chloroplast (Fan et al., 2011). Thus, it is possible that the plastid contributes to the formation of LDs in green algae in general (Goold et al., 2015; Xu et al., 2018a).

DGATs and *PDATs* are good candidates for genetic engineering since they catalyze the last and committed step in TAG biosynthesis. For instance, *DGAT* overexpression has been reported to be able to increase seed oil content in some plant species (Liu et al., 2012). In addition, DGAT and PDAT were also shown to play important physiological roles in plant and algal development, which may contribute in abiotic and biotic stress tolerance (Arisz et al., 2018; Fan et al., 2013a, 2013b; Tan et al., 2018; Weselake et al., 2008; Xu et al., 2018a). DGATs

belong to three different families (Cases et al., 1998; Lardizabal et al., 2001; Saha et al., 2006). DGAT1 and DGAT2 are membrane-bound proteins, while DGAT3 was characterized as cytosolic. These enzymes share no homology, showing a putative convergent evolution. In plants, DGAT1 is usually associated to oil accumulation in seeds, while DGAT2 is linked to the incorporation of unusual fatty acids (Liu et al., 2012; Xu et al., 2018a). In algae, DGAT1 may have a minor role compared with DGAT2 because many algal species only have one *DGAT1* but multiple copies of *DGAT2s* (Chen and Smith, 2012). Plant *PDATs* are clustered into different phylogenetic clades with some groups having different properties (Pan et al., 2015). Ancient and recent duplications are present in the evolution of eudicot *PDATs*. In this study, I will characterize PDATs in algae and plants and their possible functions will be discussed in Chapter 3.

1.1.4. Bioinformatics in the characterization of genes and proteins

Similar to the analysis of other genes, the characterization of PDATs in algae and plants requires bioinformatic analysis and on-bench experiments. Various bioinformatics tools and methods have been developed and used in the characterization of genes and proteins. In this section, some bioinformatic methods were summarized, with the focus on the ones being used in the current project.

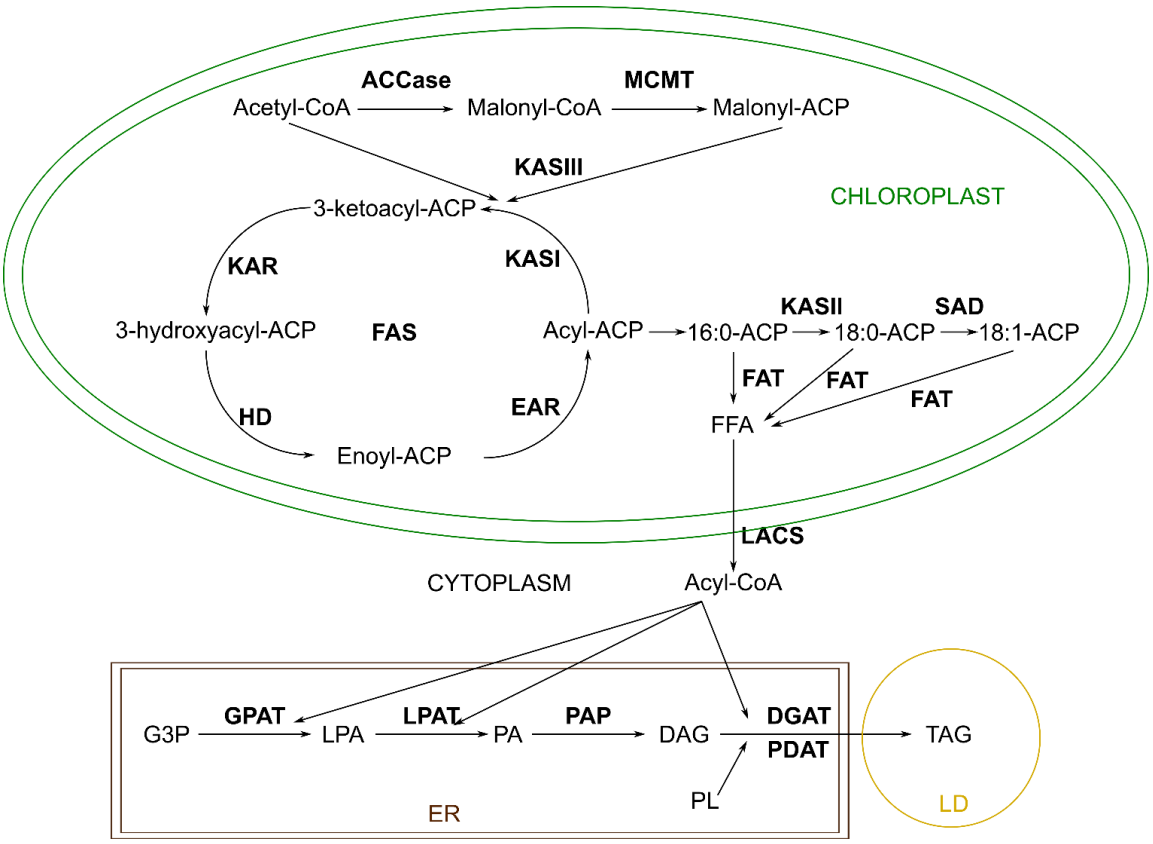


Figure 1.2. Simplified scheme of lipid biosynthesis pathways. Based on (Li-Beisson et al., 2019, 2015; Xu et al., 2018a). 16:0-ACP, palmitoyl-ACP; 18:0-ACP, stearoyl-ACP; 18:1-ACP, oleoyl-ACP; ACCase, acetyl-CoA carboxylase; ACP, acyl carrier protein; CoA, coenzyme A; DAG, diacylglycerol; DGAT, acyl-CoA:diacylglycerol acyltransferase; EAR, enoyl-ACP reductase; ER, endoplasmic reticulum; FAS, fatty acid synthase; FAT, acyl-ACP thioesterase; FFA, free fatty acid; G3P, glycerol-3-phosphate; GPAT, glycerol-3-phosphate acyltransferase; HD, hydroxyacyl-ACP dehydrase; KAR, ketoacyl-ACP reductase; KAS, ketoacyl-ACP synthase; LACS, long-chain acyl-CoA synthase; LD, lipid droplet; LPA, lysophosphatidic acid; LPAT, lysophosphatidic acid acyltransferase; MCMT, malonyl-CoA:ACP malonyltransferase; PA, phosphatidic acid; PAP, phosphatidic acid phosphatase; PDAT, phospholipid:diacylglycerol acyltransferase; PL, phospholipid; SAD, stearoyl-ACP desaturase; TAG, triacylglycerol.

1.1.4.1.Databases

Biological databases that are relevant for bioinformatics are formed of nucleotide sequences, protein sequences, protein structure, genomes, proteomes, transcriptomes, lipidomes, annotations and/or biological function and others. They facilitate the management of data since sequences and other information can be searched and retrieved using taxonomic information and sequence homology (Bottu, 2009). Genbank (https://www.ncbi.nlm.nih.gov/genbank/), European Nucleotide Archive – ENA (https://www.ebi.ac.uk/ena), and DNA Data Bank of Japan – DDBJ (https://www.ddbj.nig.ac.jp/index-e.html) contain most of the publicly available data and are thus the most broadly used databases. However, search in these systems sometimes may generate an enormous amount of results, which may be overwhelming and not as specific as we expected for certain analysis. Thus, sometimes it is reasonable and even better to carry out the search in curated or specialized databases, where the information is easier and faster to access. Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html), for instance, provides access to annotated plant and algal genomes and is popularly used in plant biology research. Studies aiming only plant or algae could easily search entire genomes or download them for local usage. Genome Portal (https://genome.jgi.doe.gov/portal/) and Ensembl (http://uswest.ensembl.org/index.html) also have several genomes, which can be readily retrieved or searched. Moreover, Ensembl is also organized into main sub-groups of species, such as bacteria, fungi, plants, protists and metazoan, which is convenient for bioinformatics analysis with the focus on certain species. Protein sequences and proteomes can be accessed at UniProt (https://www.uniprot.org/), while the crystal structure of proteins can be retrieved from Protein Data Bank – PDB (https://www.rcsb.org/). There are other databases for different applications or different

organisms. Thus, the selection of the best databases for an individual project in biological sciences could result in better analysis and save time and resources.

1.1.4.2.Sequence similarity search

Sequence similarity search is frequently used in comparing homologous genes in different species and other analyses. This search can be carried out in various databases with implemented programs in them. By comparing the sequence of the target genes with other genes or proteins in databases, a researcher can obtain clues regarding the possible functions and conserved motifs of the encoded proteins. In general, searching for homologous genes are broadly used, and sometimes serve as the first step of sequence similarity search, in various analysis. Homolog sequences can be obtained with the alignment of the query sequence with those from the database (Bottu, 2009). The alignment method usually assigns positive scores for matching nucleic acids or amino acids and negative scores for mismatching or gaps. Gaps are assigned when there are insertions or deletions. Then, scores and probabilities are calculated to select sequences with higher similarity.

The most used program to search for homologs is Basic Local Alignment Search Tool (BLAST; Altschul et al., 1990), in which the default parameters usually produce acceptable results. The users could change the parameters for a special purpose or to improve the accuracy of the searching results. In some analysis, the scores for matches/mismatches and the matrix used in protein searches must be adjusted depending on how many hits one expects and how similar to the query the hit should be. For instance, increasing the ratio of match/mismatch scores in a nucleotide search returns sequences with a higher divergence, which is a good strategy to obtain more hits (States et al., 1991). In BLAST, the E-value (expectation value) represents the number

of sequences with a certain score that can be obtained by chance in the searched database (Bottu, 2009). E-Value is another parameter that can be adjusted to limit the hits - the lower the E-value, the higher the confidence in the hit sequence.

BLAST also has different searching strategies to allow the search of nucleotide or proteins sequences. BLASTN is used to search a nucleic acid database with a nucleic acid sequence as the query. BLASTP searches a protein database with a protein sequence. BLASTX searches a protein database with the six translated frames of a nucleic acid sequence query. TBLASTN compares the six translated frames of a nucleotide database with a protein sequence as the query. TBLASTX is used to compare the six translated frames of a nucleotide database with the six translated frames of a nucleotide sequence (Bottu, 2009). Choosing the proper search strategy in BLAST is important for effectively getting the expected information in the sequence similarity search.

1.1.4.3.Multiple sequence alignment (MSA)

The alignment of two or more sequences, also known as multiple sequence alignment (MSA) is one of the most necessary activities in bioinformatics and molecular biology. It may be used to find or predict conserved motifs, investigate sequence similarity, and design primers (Doğan and Otu, 2014; Higgins and Lemey, 2009; Nalbantoglu, 2014). In addition, MSA is usually used as the first step in phylogenetic analysis to provide columns with homologous sites for the tree reconstruction process. For instance, in the study of isolating the first DGAT2 from *Mortierella ramanniana*, an MSA was generated to show that this enzyme is also present in other eukaryotes (Lardizabal et al., 2001). In another study, the comparison of *Chlorella ellipsoidea* DGAT1 with

its homologs from other algae and plants in an MSA showed relevant conserved motifs and regions (Guo et al., 2017).

The same matrices used in BLAST can also be used here to help find the alignment with the best score. The scores in the matrices represent the frequency of change of each amino acid as well as their biochemical properties (Higgins and Lemey, 2009). Nucleic acid sequences with a high divergence may be difficult to align since there are only four bases (A, T, G, and C), which can generate incorrect or uninformative alignments. In addition, nucleic acid sequences can accumulate more mutations than proteins because the genetic code is degenerate. Thus, it is better to align proteins or nucleotide sequences at the codon level (Higgins and Lemey, 2009; Wernersson and Pedersen, 2003). When parts or sites of the alignment cannot find a match, a gap is added. Gaps are important in MSA because they introduce information in the positions of putative insertions and/or deletions.

Dynamic programming is an algorithm that can find the best scoring alignment for a pair of sequences (Higgins and Lemey, 2009; Nalbantoglu, 2014), which is used to find the best global alignment, as implemented in the Needleman-Wunsch algorithm, or the best local alignment, as implemented in the Smith-Waterman algorithm (Nalbantoglu, 2014; Needleman and Wunsch, 1970; Smith and Waterman, 1981). The local alignment is important to discover functional motifs or similar regions between proteins. Using a weighted sum-of-pairs function, an MSA with the best score can be found. However, the processing memory and time grow exponentially with the number of sequences with dynamic programming (Higgins and Lemey, 2009; Sievers et al., 2011). Thus, this method is ideal solely for pairwise alignment.

The modern algorithms for MSAs use phylogenetic relationship among the sequences to prepare the alignment (Higgins and Lemey, 2009). In the progressive alignment, a tree is quickly

generated using a distance method. Then, a pairwise alignment is generated using the two closest sequences according to the tree and the result is treated as one single sequence. The next closer sequence is subsequently added to a new pairwise alignment. This iteration continues until the last sequence is included in the alignment. ClustalW is a tool that uses progressive alignment (Thompson et al., 1994). Mistakes introduced early during the alignment cannot be corrected after the introduction of new sequences, which may diminish the overall quality and biological meaning of the alignment (Higgins and Lemey, 2009; Sievers et al., 2011; Thompson et al., 1994). The consistency-based scoring approach may overcome this problem (Doğan and Otu, 2014; Higgins and Lemey, 2009). Although this approach is slower than progressive alignment it introduces new information in the alignment to calculate the final MSA. T-Coffee, for instance, generates a library of alignments with the sequence pairwise alignments (Notredame et al., 2000). Later, this library will be used as a guide for the MSA. Iterative approaches include iterative cycles that modify the initial alignment towards a better solution (Higgins and Lemey, 2009). MAFFT, which uses a fast Fourier transform to find homologous regions, iterates through the alignment by diving it into two groups and realigning them until the convergence of the score of both groups (Higgins and Lemey, 2009; Katoh et al., 2002; Katoh and Standley, 2013). The choice of method depends on the needs of the user. For instance, a program, such as ClustalΩ, would be preferred for very large datasets because it can align the sequences only in a few hours (Sievers et al., 2011).

1.1.4.4.Phylogenetic analysis

Phylogenetic analysis provides the means to reconstruct the evolutionary relationship between organisms or sequences (Vandamme, 2009). Phylogenetic trees are the graphical representation

of these results. These analyses are important to investigate the relationship between sequences, infer functions, gene duplications, and horizontal gene transfers. In addition, they can be used to date evolutionary events or reconstruct an ancestral sequence. In the same study regarding the *M. ramanniana* DGAT2 as mentioned above, the researchers reconstructed the phylogenetic tree of the sequences after preparing the MSA (Lardizabal et al., 2001). The results showed that gene duplications are specific to some species, which means they did not happen in the ancestor of those organisms. Conversely, in a study about plant *PDAT* genes, the phylogenetic trees built with different methods for plant *PDATs* showed the presence of both ancestral and recent duplications in eudicots (Pan et al., 2015).

Nucleotide substitution model is an important concept in phylogenetic analysis. The distance between sequences can be roughly estimated by the number of substitutions found in a pairwise alignment. However, it may not always correctly describe what happened because multiple substitutions may hit the same site. The nucleotide substitution models provide the assumptions and the statistical means to describe the rate of mutations in the phylogenetic inference (Arenas, 2015; Strimmer and von Haeseler, 2009). The most used models are time reversible; i.e. the rate of change from base "a" to base "b" is the same from base "b" to base "a". In the simplest model, the JC69 model, the rate of equilibrium frequencies is the same for the nucleotides and the substitutions have the same probability for any nucleotide (Arenas, 2015; Jukes and Cantor, 1969; Strimmer and von Haeseler, 2009). The general time reversible (GTR) model has eight free parameters (the nucleotide frequencies and the rate of change may be different) and, thus, is the most complex model (Strimmer and von Haeseler, 2009). Although simpler models can be derived from GTR, the original complex GTR model usually fits better to the data due to the complexity of the evolutionary process (Arenas, 2015; Sumner et al., 2012).

To further improve the quality of the models, a rate heterogeneity among sites (+G) and a rate of invariable sites (+I) can be included in the phylogenetic analysis. The former improvement (+G) includes differences in certain sites in the model, such as the third nucleotide of codons, which has a higher change frequency (Yang, 1994). The latter improvement (+I) takes the most conserved nucleotides in the dataset into consideration because the behavior of these nucleotides cannot be explained by the common nucleotide substitution models and should be modeled differently (Shoemaker and Fitch, 1989). In general, nucleotide substitution model with less than eight free variables is adequate in fitting the data. The best-fit model for each dataset is usually estimated prior to the phylogenetic inference.

Various phylogenetic analysis methods have been developed and used in data analysis. Since the methods have different working theories and may result in different phylogenetic trees and thus different biological assumptions from the same dataset, it is important to understand the basics of the methods and choose the right one and their parameters for proper data analysis. Distance methods, maximum likelihood, and Bayesian inference are probably the most used methodologies and are thus further discussed here.

Distance methods are solely based on the genetic variability obtained in the pairwise alignment (Van de Peer and Salemi, 2009). The neighbor-joining (NJ) method is a commonly used distance method, which tries to find pairs of neighbor taxa that minimize the sum of branch-lengths (Saitou and Nei, 1987). It is an adequate method for quick construction of phylogenetic trees but may not handle very complex cases.

The maximum likelihood (ML) method uses the likelihood function to reconstruct the phylogenetic tree (Schmidt and von Haeseler, 2009; Yang and Rannala, 2012). This function can estimate the likelihood of the observed data using statistics. Specifically, the maximum

likelihood estimation can find the values that maximize the parameters that form the likelihood. It uses different trees in the model, attributes a probability to each tree, and obtain the best one. The advantage of the ML methods is that they include all the data provided in the model and generate multiple alternative trees to find the best solution. Although ML is more robust than NJ, it may take much longer processing times to assess all possible trees when the datasets contain several sequences (Schmidt and von Haeseler, 2009; Yang and Rannala, 2012). Different approaches have been developed to address this issue in ML. For example, PhyML v3.0 used both nearest-neighbor interchange (NNI) and subtree pruning and regrafting (SPR) to search for the best tree in a reasonable time (Guindon et al., 2010). However, the best tree found in the program may not always be the optimal tree (Schmidt and von Haeseler, 2009).

Both NJ and ML methods rely on bootstrap analysis to indicate the reliability of an obtained tree (Van de Peer and Salemi, 2009). In the analytical process, the columns of the alignment are resampled and a new tree is obtained in each cycle. The trees are later summarized to form the consensus tree with branch supporting values. Branches are considered well supported if they are present in more than 75% of the replicates (Zharkikh and Li, 1992). The Shimodaira-Hasegawa approximate likelihood ratio test (aLRT) implemented in PhyML v 3.0 is a faster alternative to bootstrap analysis (Anisimova et al., 2011; Anisimova and Gascuel, 2006; Guindon et al., 2010). In aLRT, the branches are compared to the null model, which defines the branch as having zero length. This aLRT method was shown to perform well with analyses showing strong model violations (Anisimova et al., 2011).

Bayesian inference (BI) is also widely used in phylogenetic analysis (Ronquist et al., 2012, 2009). The calculations are based in the Bayes' theorem and the results are shown as posterior probabilities, which are of easy interpretation. MrBayes is a widely used software that

reconstructs phylogenies with the Bayesian approach (Ronquist et al., 2012). BI is as powerful as the ML method because it can use complex evolutionary models and generates several trees, which are sampled every few generations. In addition, BI can produce robust results if information about the prior probability distribution is included (Yang and Rannala, 2012). Similar to ML, the computation takes a long time in the BI method where the majority rule consensus tree method is applied to summarize all trees sampled during the analysis (Ronquist et al., 2009).

The advantage of ML and BI is the production of several trees and the possibility to adjust several parameters. As both methods generate a log-likelihood (ℓ) for the trees, the ℓ values can be used in hypothesis tests to compare different tree topologies. In MrBayes, for instance, the harmonic mean can be used as a rough estimation of which tree is better (Ronquist et al., 2012). However, the stepping-stone sampling method is a better approach for the topological hypothesis test because the harmonic mean overestimates ℓ (Xie et al., 2011). Then, the alternative and null hypotheses can be compared with $2x(\ell_0-\ell_1)$. If the result is (i) higher than 10, (ii) between 6 and 10, (iii) between 2 and 6, and (iv) between 0 and 2, there is (i) very strong, (ii) strong, (iii) positive, and (iv) weak evidence in favor of the null hypothesis, respectively (Kass and Raftery, 1995).

1.1.4.5.Selective pressure analysis

Nucleic acid substitutions in codons can be used to estimate how evolutionary pressures act on protein-coding sequences. The degeneracy of the genetic code allows the accumulation of mutations. If the mutations do not change the encoded amino acids and do not affect the organism, they are considered synonymous mutations (dS). If the new nucleotide causes a

change in the encoded amino acid, they are considered non-synonymous (dN; Vandamme, 2009). The ratio of dN/dS is presented as ω. If $0<\omega<1$, it means more synonymous substitutions than non-synonymous ones have been accumulated during evolution, which is considered negative or purifying (Yang, 2007). Purifying selection indicates that non-synonymous mutations are harmful to the enzyme and are removed from the population due to a decrease in fitness. On the contrary, $\omega = 1$ and $\omega > 1$ represent neutral and positive selection, respectively, and the latter case suggests that non-synonymous mutations are beneficial to the organisms and are retained in a population (Vandamme, 2009; Yang, 2007; Yang et al., 2005).

The PAML package contains a CODEML program, which can run several codon analyses to test selective pressure among lineages or sites of protein-coding sequences (Yang, 2007). This program contains several models such as site models, branch models, branch-site models, and clade models. Site models can compare the selective pressures acting in the codons of a dataset and branch models may estimate ω values for selected branches of a lineage. Branch-site models are especially useful in estimating positive selection between foreground and background branches. Branch-site models allow the variation of ω among sequence sites and lineages of a phylogenetic tree. Clade models also compare clades in the foreground and background, but they are mostly used to test divergent selection between them (Bielawski and Yang, 2004). Each of these models has a ℓ, which can be compared with ℓ from the null model (ℓ_0) with the likelihood ratio test (LRT) statistics: $LRT = 2x(\ell_1-\ell_0)$ (Posada, 2009; Schmidt, 2009). The LRT follows a χ^2 distribution, which can be used to calculate a critical value to accept or reject an alternative hypothesis. In site and branch-site models, the Bayes empirical Bayes (BEB) method can be used to estimate the posterior probability of residues in the site class

with $\omega > 1$. Sites are considered under positive selection if their posterior probability is elevated (e.g. $p > 0.95$; Yang et al., 2005).

1.1.5. Enzymatic assays

Considering that enzymes catalyze most reactions in the cells, *in vitro* enzymatic assays are in the core of biochemistry and have broad applications (Punekar, 2018a, 2018b). Enzymatic reactions are reactions between organic compounds, which are catalyzed by enzymes (Koshland Jr., 1953). Important applications of enzymatic assays include the identification of enzyme function, the characterization of the enzymatic reaction, the study and diagnosis of diseases, and the identification of mutation effects. The characterization of enzymes with *in vitro* assays has been broadly applied in the study of lipid metabolism in many organisms. For instance, mutagenized variants of *B. napus* DGAT1 and *Camelina sativa* DGAT1B were studied with *in vitro* enzymatic activity (Xu et al., 2017) and the results showed that mutations in the ninth transmembrane domain could enhance DGAT1 performance. In another study, the diagnostic of diseases associated with lecithin:cholesterol acyltransferase (LCAT) involves the *in vitro* determination of LCAT activity in the blood serum using radiolabeled or fluorescent substrates (Sakurai et al., 2018; Vaisman and Remaley, 2013).

1.1.5.1.Enzymatic assay of membrane proteins

Some enzymes have transmembrane domains (TMDs) and bound in cellular membranes. Due to their hydrophobic properties, sometimes it is challenging to purify these enzymes without losing their activities (Tiefenauer and Demarche, 2012). This is particularly problematic for *in vitro* assays of enzymes in the lipid metabolism of plants and algae because many of the enzymes,

such as DGAT1, DGAT2, and PDAT, are membrane-bound proteins (Liu et al., 2012; Pan et al., 2015; Xu et al., 2018a). As an alternative approach, crude protein extracts containing the enzyme and the membrane fractions (microsomes) are used in *in vitro* assays.

However, microsomes directly extracted from plant tissue or algal cells are sometimes not suitable for the characterization of a target enzyme because the content of the target enzyme in the microsomes is generally low. In addition, the microsomes also contain other enzymes in the same pathways, which may interfere with the *in vitro* assay. With the development of molecular biology, the genes encoding the target enzymes can be identified, isolated and heterologously expressed in different hosts such as yeasts, bacteria, plants or animal cells for the accumulation of the target enzymes in a high concentration. As a result, microsomes extracted from the host cells contain a high amount of the target enzymes. The expression of the membrane-bound enzymes in eukaryotes is especially useful because eukaryotes have different membranes in their cells, which might be necessary for the correct folding and function of the enzymes.

Saccharomyces cerevisiae has been widely used as a host for the expression of recombinant lipid-related enzymes. *S. cerevisiae* grows well and fast in cheap media. As a model species, this yeast species has been extensively studied. Its genome sequences are available, various yeast mutants have been produced, various protocols for gene transformation and protein expression are generated, and many related products are commercially available (Gietz and Schiestl, 2007; Strausberg and Strausberg, 1995; Xie et al., 2018). The most popularly used yeast strain in DGAT and PDAT studies in plant and microalgae is *S. cerevisiae* strain H1246, which can not synthesize TAG due to the knockout of all four TAG-synthesizing genes, *DGA1*, *LRO1*, *ARE1*, and *ARE2* (Sandager et al., 2002). If the expression of a target *PDAT* and *DGAT* could

restore TAG biosynthesis in this strain, it indicated the function of the target enzyme (Caldo et al., 2018; Chen et al., 2017; Greer et al., 2016; Pan et al., 2013; Siloto et al., 2009; Xu et al., 2018b, 2017). In addition, microsomes of the *S. cerevisiae* H1246 hosting a target PDAT or DGAT protein can be extracted for extensive *in vivo* assay (Siloto et al., 2009; Xu et al., 2018b, 2017).

1.1.5.2.Use of fluorescent substrates in *in vitro* assay of lipid-related enzymes

Similar to many other *in vitro* enzymatic assays, the conventional *in vitro* DGAT and PDAT assays using C14-radiolabeled substrates because it is sensitive and accurate (Cases et al., 2001, 1998; Xu et al., 2018b). However, radiolabeled chemicals are usually expensive and strictly regulated. Safe and cheap fluorescent compounds have been used in the *in vitro* assays of phospholipases and DGATs (Chen et al., 2012; McFie and Stone, 2011; Moreau, 1989; Sanderson and Venable, 2012). Detection of fluorescence can be carried out in laser scanners, spectrophotometers or microplate readers. Similar to the assay with radiolabelled substrates, the fluorescent substrates and products of some enzymatic assays can be easily separated on a TLC plate. When a TLC plate is scanned, the fluorescent intensity of the spots representing the enzymatic reaction products can be quantified with image-processing software, such as ImageJ (Schneider et al., 2012). For instance, acyl-CoA or DAG labeled with nitrobenzoxadiazole (NBD) have been used as fluorescent substrates in *in vitro* DGAT1 assay. Several lipids containing an NBD group are commercially available, which can be used in the assay of other lipid-related enzymes (Amaro et al., 2016). However, fluorescent-labeled substrates may not work as similar as C14-labeled ones because the fluorescent chemical groups are fairly large and may change the structure of the substrate and subsequently affect or even prevent the enzymatic

reaction (Sakurai et al., 2018). Therefore, although fluorescent-labeled substrates have many advantages, the method of using them in the assay of a target enzyme must be established and validated first.

1.1.5.3.Michaelis-Menten kinetics

Enzymatic reaction kinetics is one of the most important parameters in enzymology. During enzymatic assays, several factors are determined, such as optimal pH, temperature, enzyme concentration, and time to produce a specified amount of products. The reaction kinetics, which can be calculated from some of the above factors, offers an explanation on the mechanism of reaction or whether and how substrates or other molecules influence the enzyme activity. Some kinetics equations have been established based on various enzymatic assays. The Michaelis-Menten equation is one of the well-studied equations. If on-bench experiments show that any enzymatic reaction fits this equation, the relationship between the substrate concentration and the initial velocity of the enzymatic reaction can be defined as below (Johnson and Goody, 2011; Punekar, 2018c).

$$V_0 = \frac{V_{max}[S]}{K_m + [S]} \qquad \text{(Michaelis-Menten equation)}$$

V_0 is the initial reaction velocity, [S] is the substrate concentration, V_{max} is maximal reaction velocity, and K_m is the Michaelis constant. V_{max} is the maximal rate of a reaction when the increase in [S] concentration no longer affects the V_0. K_m is defined as the concentration of substrate where V_0 is half of V_{max}. These parameters can be obtained using the Lineweaver-Burk plot, which plots $1/V_0$ and 1/[S] (Lineweaver and Burk, 1934).

$$\frac{1}{V_0} = \frac{1}{V_{max}} + \frac{K_m}{V_{\max}[S]} \text{ (Lineaweaver-Burk equation)}$$

V_{max} and K_m are important parameters for the description of an enzymatic reaction mechanism and for the comparison of reaction conditions or different enzymes. Moreover, the Michaelis-Menten equation can be further modified to understand other effects and conditions, such as product inhibition (Johnson and Goody, 2011).

1.1.6. The biochemistry and physiological roles of PDAT

1.1.6.1.The biochemical properties of PDATs

PDATs belong to the LCAT family, which also includes the LCATs, phospholipases, and phospholipid:sterol O-acyltransferase (PSAT; Pan et al., 2015; Yoon et al., 2012). PDATs use phospholipids as the acyl donors and DAGs as the acyl acceptors to produce TAGs and lysophospholipids. The enzyme preferentially catalyzes the transfer of the acyl group from the *sn-2* position of the phospholipid to the *sn-3* position of TAG (Stahl et al., 2004). While enzymes in the LCAT family generally prefer phosphatidylcholine (PtdCho) as the acyl-donor (Liu et al., 1998), PDATs from different organisms might have different head group preferences. It was shown that *S. cerevisiae* PDAT and *A. thaliana* PDAT1 prefer phosphatidylethanolamine (PtdEth) over PtdCho as a substrate (Dahlqvist et al., 2000; Stahl et al., 2004). In addition to head groups, PDAT also has a different preference for the acyl groups on phospholipids. As for AtPDAT1, the acyl groups were shown to impact the enzyme activity more than the head groups (Stahl et al., 2004). The algal PDAT from *Myrmecia incisa* prefers PtdCho as a substrate (Liu et al., 2016). Conversely, the PDAT from a model green microalga *C. reinhardtii* prefers anionic phospholipids: TAG formation was higher when phosphatidylinositol and phosphatidylserine were used at the acyl donor instead of PtdCho, respectively (Yoon et al., 2012). This might be a unique case since *C. reinhardtii* does not synthesize PtdCho *in vivo* (Giroud et al., 1988).

PDAT enzymes share structural and functional similarities with other members in the LCAT family, such as human LCAT and phospholipases. Important sites of the human LCAT (e.g. catalytic triad and lid region) were shown to be conserved among plant PDATs (Pan et al., 2015). Although the crystal structure of PDAT has not been elucidated, the high-resolution structure of human LCAT and lysosomal phospholipase A2 (PLA2) have been recently published (Glukhova et al., 2015; Piper et al., 2015), which could provide useful information to interpret the PDAT conformation. According to the crystal structure, the LCAT enzyme could be subdivided into three main regions: an α/β hydrolase domain and two subdomains (subdomain 1 and a subdomain 2; Piper et al., 2015). The model shows that the subdomain 1 contains a region used by the enzyme for interfacial activation ("lid" domain) when in contact with high-density lipoprotein (HDL) particle. PDAT enzymes might contain a similar domain, which would interact with membranes that supply the phospholipids for the reaction. The second subdomain, located on the top of the catalytic sites, contains amino acids that contribute to the formation of the catalytic pocket. The α/β hydrolase domain contains β-sheets linked by α-helices and is folded to arrange the catalytic triad into nearby loops (Ollis et al., 1992). The crystal structure of lysosomal PLA2 is similar to that of LCAT, which seems to be a characteristic of this family (Glukhova et al., 2015; Piper et al., 2015).

Based on the structural information of LCAT and lysosomal PLA2, the catalytic mechanism of PDAT was proposed (Adimoolam et al., 1998; Glukhova et al., 2015; Piper et al., 2015). PDAT catalysis may start by the interaction of the "lid" region with an intracellular membrane, which would activate PDAT and position the enzyme with the opening of the catalytic pocket towards the membrane. Then, a phospholipid from the membrane would enter the active site with the head group towards the bottom. The acyl-chain from the *sn-2* position of

the phospholipid will acylate the serine in the catalytic triad. As a result, a molecule of lysophospholipid is produced, which then leaves the pocket. The PDAT enzyme remains acylated by this acyl group until it contacts with a DAG. After that, the acyl group is transferred to the *sn-3* position of a DAG to form a TAG.

The presence of a unique transmembrane domain (TMD) in the N-terminal region is common to most plant PDATs, which are predicted to be located in the endoplasmic reticulum (ER) with the C-terminal end oriented towards the ER lumen (Pan et al., 2015; Stahl et al., 2004). However, in the microalga *C. reinhardtii*, the PDAT is predicted to be localized in the plastid and the *M. incisa PDAT* expressed in tobacco leaves is in the plasma membrane (Liu et al., 2016; Yoon et al., 2012). Thus, the functional roles of PDAT in plants and algae might be different.

1.1.6.2.The physiological roles of PDAT in high plants and microalgae

Previous studies showed that the physiological roles of PDAT are diverse among different organisms. PDATs are able to produce oils with different acyl chains, including unusual fatty acids in the TAG molecules (Banaś et al., 2013; Dahlqvist et al., 2000; Pan et al., 2013; Stahl et al., 2004). This characteristic is sometimes species-specific and likely contributes to different adaptation conditions. The expression levels of *PDAT* may be distinct in different tissues (Pan et al., 2015). Moreover, PDATs might be more active in distinct moments and tissues in the life cycle of the organisms (Pan et al., 2015). In yeasts, the enzyme is more active in the logarithmic phase, but the overexpression of *PDAT* also increases oil accumulation in the stationary phase (Dahlqvist et al., 2000). Camelina (*Camelina sativa*) has five *PDAT* paralogs with tissue-specific expression profiles (Yuan et al., 2017). For instance, the expression of *CsPDAT1-A* is high in

seeds whereas that of *CsPDAT1-C* high in leaves and flower. *CsPDAT2-B* and *CsPDAT2-A* have high expression levels in roots and stems. In addition, *CsPDATs* respond differently under abiotic stresses: *CsPDAT2-A* expression was stimulated by drought, whereas cold stress stimulated the expression of *CsPDAT1-A* and *CsPDAT1-C* (Yuan et al., 2017).

The functions of PDAT in lipid biosynthesis and related physiological roles have been reported in the model plant *A. thaliana*. *Atpdat1* knockout mutant line did not show any differences in TAG accumulation in seeds, because DGAT1 can fully complement the absence of PDAT1 (Mhaske et al., 2005; Zhang et al., 2009). On the contrary, the *Atdgat1* knockout line has low TAG content in seeds, indicating AtPDAT1 can only partially complement the absence of DGAT1 (Zhang et al., 2009). Moreover, AtPDAT1 also has a major role in TAG biosynthesis in leaves: *Atpdat1* and *Atdgat1* mutants can only accumulate 43% and 61% of the TAG being produced in wild-type Arabidopsis lines, respectively (Dahlqvist et al., 2000; Fan et al., 2013b). Indeed, PDAT plays an important role in lipid homeostasis in developing tissues of *A. thaliana* (Fan et al., 2013a). An *Arabidopsis* mutant line (*tgd1-1*) being defective in the transport of lipid precursors from the ER to the plastids and thus deficient in the eukaryotic thylakoid lipid synthesis pathway, was used to verify the effects of *PDAT1* disruption (Fan et al., 2013a; Xu et al., 2003). The results indicated that the lack of PDAT activity caused premature cell death due to the accumulation of toxic free fatty acids in the growing tissues. In addition, the levels of phospholipids (PtdCho and PtdEth) increased in the *tgd1-1* lines lacking *PDAT1*, indicating the role of PDAT in maintaining lipid homeostasis in the cells.

Although only a few algal PDATs have been well-studied, the results indicated they have some similarities with the homologs in other organisms. PDAT helps in the accumulation of TAG in *C. reinhardtii* under favorable conditions, which is similar to the role of PDATs in yeast

cells and *A. thaliana* leaves (Dahlqvist et al., 2000; Fan et al., 2013b; Yoon et al., 2012). Interestingly, although TAG is generally accumulated in a high amount under stress in cells, the absence of PDAT does not affect TAG content in *C. reinhardtii*, indicating other enzymes compensate PDAT function (Yoon et al., 2012). In addition, *C. reinhardtii* PDAT is responsible for lipid homeostasis in the plastid membrane, where phospholipids are the acyl donors for TAG biosynthesis. PDAT was also proposed to catalyze the membrane remodeling of *M. incisa* under nitrogen starvation because the relative abundance of PtdCho was significantly reduced after four days of cultivation under this stress (Liu et al., 2016). In addition, C. *reinhardtii* PDAT also has lipase and phospholipase activities, which suggest this enzyme might have other intracellular roles in this organism (Yoon et al., 2012).

1.2.Objectives

The overall objective of this thesis is to compare and analyze PDATs in plants and microalgae to expand our knowledge of this important enzyme, which is composed of two sub-objectives as described below.

The standard PDAT assay uses radiolabeled substrates in the reactions. Considering the costs of radiolabeled chemicals and the strict regulations that must be followed to deal with these compounds, replacing radiolabeled substrates with non-radiolabeled and cheaper ones would be attractive. NBD is a widely used fluorescent group in the study of lipids (Amaro et al., 2016). NBD-PtdCho containing the NBD group at the end of the acyl chain is commercially available. In addition, the conversion of NBD-PtdCho into NBD-DAG has already been developed (Sanderson and Venable, 2012). Considering the assumption that PDAT and LCAT have a similar mechanism of reaction, it was hypothesized that the position of the NBD group in the

NBD-DAG would not affect catalysis. This hypothesis is consistent with the model described previously where the glycerol backbone touches the bottom of the catalytic pocket and the acyl chains are oriented towards the entrance of the active site. In addition, the region of the enzyme that interacts with the acyl chains is considered to have greater plasticity to accommodate different substrates (Glukhova et al., 2015). Therefore, the first research objective of this thesis is to establish a reliable PDAT assay with fluorescent-labeled substrates (Chapter 2), which also serves as the analysis method in the further characterization of algal and plant PDATs (Chapter 3).

As explained previously, PDAT is an important enzyme in TAG biosynthesis and thus is a good candidate enzyme for manipulation with the aim to improve storage lipid content or modify fatty acid composition. in algae and plants. A good understanding of PDATs is, thus, necessary to pursue these approaches. While there are several studies about plant PDATs, algal PDATs are only well-studied in two species (Liu et al., 2016; Yoon et al., 2012). Moreover, PDATs also play important roles in the physiology of plants and algae. It is necessary to expand our understanding of PDATs in algae and higher plants. In addition, the relationship between PDATs in algae and plants is largely unknown. With the rich source of "omics" data the rapid development of molecular biology, the other objective of this thesis, therefore, is to combine *in silico* and *in vitro* approaches to advance the understanding of PDATs in algae and plants (Chapter 3).

CHAPTER 2. A fluorescence-based assay for quantitative analysis of phospholipid:diacylglycerol acyltransferase activity

Abstract

Phospholipid:diacylglycerol acyltransferase (PDAT) catalyzes the acyl-CoA-independent triacylglycerol (TAG) biosynthesis in plants and oleaginous microorganisms and thus is a key target in lipid research. The conventional *in vitro* PDAT activity assay involves the use of radiolabeled substrates, which, however, are expensive and demand strict regulation. In this study, a reliable fluorescence-based method using nitrobenzoxadiazole-labeled diacylglycerol (NBD-DAG) as an alternative substrate was established and subsequently used to characterize the enzyme activity and kinetics of a recombinant *A. thaliana* PDAT1 (AtPDAT1). I also demonstrate that the highly toxic benzene used in typical PDAT assays can be substituted with diethyl ether without affecting the formation rate of NBD-TAG. Overall, this method works well with a broad range of PDAT protein content and shows a linear correlation with the conventional method with radiolabeled substrates, and thus may be applicable to PDATs from various plant and microorganisms species.

2.1. Introduction

Storage lipids, which comprise mainly triacylglycerols (TAGs), are biologically and economically important molecules. TAG is composed of a glycerol backbone and three long-chain fatty acids, and is stored in high quantities (up to 70% of the microalgal biomass and seed weight in some oilseed plants) as an energy source in many organisms (Hu et al., 2008; Li-Beisson et al., 2016). Indeed, various species have been examined as potential sources of

vegetable oils and/or fatty acids, which can be used for food, feed, biofuel and industrial purposes. For instance, docosahexaenoic acid (DHA) produced in certain microalgae has important nutraceutical applications, while ricinoleic acid found in castor seeds is used as a feedstock for the production of high-performance polymers, coatings, varnishes, lubricants, cosmetics and surfactants (Dyer et al., 2008; McKeon, 2016; Mutlu and Meier, 2010).

The formation of different TAG species is largely controlled by the substrate specificity of the acyltransferases that sequentially transfer fatty acyl chains to the *sn*-1, 2, and 3 positions of a glycerol backbone (Xu et al., 2018a). While the first two acylation steps are catalyzed by enzymes with acyl-CoA as the sole acyl donor, the final step involving the acylation of diacylglycerol (DAG) to form TAG can be carried out in either acyl-CoA-dependent and acyl-CoA-independent reactions, catalyzed by DGAT and PDAT (EC 2.3.1.158), respectively. Although DGAT (especially DGAT1) appears to be the major contributor to TAG biosynthesis in oil crops, PDAT, which uses membrane phospholipids as acyl donors to acylate DAG (Figure 2.1; Dahlqvist et al., 2000), also makes a significant contribution and plays an important role in channeling unusual fatty acids into TAG in some plant species (Dahlqvist et al., 2000; Pan et al., 2013; Stahl et al., 2004; van Erp et al., 2011; Xu et al., 2018a). For example, PDAT from castor bean and hawk's-beard (*Crepis palaestina*) preferentially incorporate ricinoleoyl and vernoloyl moieties into TAG, respectively, while flax PDAT prefers substrates containing *α*-linolenic acid (Dahlqvist et al., 2000; Pan et al., 2013).

Given the important role of PDAT in TAG biosynthesis, this enzyme is increasingly being used as a target for the enhancement of oil content and development of designer specialty oils enriched in desirable fatty acids using genetic engineering approaches (Bates et al., 2014; Dahlqvist et al., 2000; Kim et al., 2011; Pan et al., 2013; van Erp et al., 2015, 2011; Xu et al.,

2018a). *In vitro* assays of PDATs to estimate their enzymatic activities, combined with *in vivo* studies, have been proven to be a powerful approach with which to characterize these valuable enzymes (Stahl et al., 2004; Yoon et al., 2012). However, *in vitro* assays of PDAT activities with radiolabeled substrates is a standard biochemical method, which depends on the use of expensive radiolabeled chemicals and may require separate laboratory areas, strict and tedious regulation processes, and extra training of personnel. Therefore, it would be advantageous to establish alternative methods for such assays that do not require the use of radiolabeled chemicals (McFie and Stone, 2011; Sakurai et al., 2018).

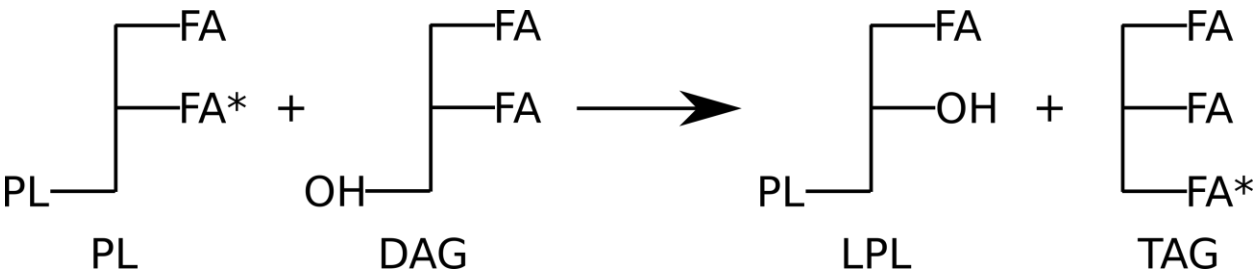


Figure 2.1. Schematic representation of the enzymatic reaction catalyzed by PDAT. PDAT catalyzes the transfer of an acyl chain from the *sn*-2 position of phospholipid to the *sn*-3 position of DAG. PL, phospholipid; DAG, diacylglycerol; LPL, lysophospholipid; TAG, triacylglycerol; FA, fatty acid

Fluorescent compounds have been used to replace those that are radiolabeled in certain *in vitro* enzymatic assays, including those involving DGAT, previously (Huang et al., 2018; McFie and Stone, 2011; Sakurai et al., 2018; Sanderson and Venable, 2012). However, the quantification of PDAT activity with non-radiolabelled substrates has yet to be established. The current study describes the development of a fluorescence-based *in vitro* method for the

quantification of PDAT activity. Using phosphatidylcholine (PtdCho) as the acyl donor and DAG labeled with nitrobenzoxadiazole (NBD) as the acyl acceptor, the activity of recombinant *A. thaliana* PDAT1 (AtPDAT1) was quantified. Moreover, alternative solvents to highly toxic benzene, which is typically used in PDAT assays, were also assessed. This fluorescence-based method for *in vitro* PDAT assay is safer, less costly and more convenient than the one using radiolabeled chemicals, and thus will increase the feasibility of future PDAT studies.

2.2. Materials and Methods

2.2.1. Genes, enzymes and chemicals

AtPDAT1 (AT5G13640) and *B. napus* DGAT1 (BnDGAT1; GenBank accession No.: JN224473; used as a positive control for the production of NBD-TAG) were used in enzymatic assays. The full-length *AtPDAT1* coding sequence was previously isolated in our laboratory using cDNA synthesized from total RNA extracted from *A. thaliana* (Col-0) siliques as template. The *AtPDAT1* coding sequence was cloned into the pYES2/NT vector (Invitrogen, Burlington, ON, Canada) for yeast heterologous expression using forward primer (5'- CAG AGC GGC CGC TAT GCC CCT TAT TCA TCG GAA AAA GCC GAC -3') and reverse primer (5'- GCT CTA GAT CAC AGC TTC AGG TCA ATA CGC TCC GAC C - 3'). Similarly, the *BnDGAT1* coding sequence was also previously isolated in our laboratory and was cloned into the pYES2.1/V5-His TOPO vector (Invitrogen; Xu et al., 2017).

Lipids, including the fluorescent substrate 1-palmitoyl-2-{12-[(7-nitro-2-1,3-benzoxadiazol-4-yl) amino] dodecanoyl}-*sn*-glycero-3-phosphocholine (NBD-PtdCho), 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (18:1,18:1-PtdCho), 1-palmitoyl-2-oleoyl-*sn*-glycerol (16:0,18:1-DAG), and oleoyl-CoA were purchased from Avanti Polar Lipids Inc. (Alabaster, AL,

USA). 1-palmytoil-2-[^{14}C]oleoyl-*sn*-3-glycero-phosphocholine (16:0,[^{14}C]18:1-PtdCho; 55 µCi/µmol) was acquired from American Radiolabeled Chemicals (St. Louis, MO, USA). Phospholipase C used for the synthesis of NBD-DAG was purchased from Sigma-Aldrich (Oakville, ON, Canada).

2.2.2. Yeast transformation

Constructs containing *AtPDAT1* and *BnDGAT1* were individually transformed into the *S. cerevisiae* H1246 strain (*MATα are1-Δ::HIS3, are2-Δ::LEU2, dga1-Δ::KanMX4, lro1-Δ::TRP1 ADE2*), which lacks the ability to synthesize TAG (Sandager et al., 2002), using the lithium acetate/single-stranded carrier DNA/PEG method (Gietz and Schiestl, 2007). Briefly, *S. cerevisiae* H1246 was cultivated in YPD medium (1% yeast extract, 2% peptone, 2% dextrose) at 30ºC with shaking at 220 rpm for 24 h. Subsequently, 2 ml of the yeast culture was used to inoculate 50 ml of YPD medium and the yeast culture was grown at 30ºC with shaking at 220 rpm until the cell concentration reached an OD_{600} of 0.6 - 0.8. Yeast cells were then harvested by centrifugation at 3,000 g for 3 min, washed twice with sterile water and once with LiTE solution (100 mM lithium acetate, 10 mM Tris, 1 mM EDTA, pH 7.5). Yeast cells were subsequently resuspended in 1ml of LiTE solution. For yeast transformation**,** 100 µl of the resuspended yeast cells were mixed with 5 µl of 2 µg/µl deoxyribonucleic acid sodium salt from salmon testes (Sigma-Aldrich), 500 µl 40% PEG3350-LiTE solution (40% PEG 3350, 100 mM lithium acetate, 10 mM Tris, 1 mM EDTA), 60 µl of DMSO, and 600 ng of the plasmid containing *AtPDAT1* or *BnDGAT1* coding sequences. The mixtures were incubated at room temperature for 15 min, followed by heat shock at 42ºC for 15 min. The yeast cells were then spread on agar plates composed of solid minimal media lacking uracil [0.67% yeast nitrogen base, 0.2% synthetic

complete medium lacking uracil (SC-Ura), 2% dextrose, and 2% (w/v) agar] and incubated at 30ºC. Transformants were identified as colonies that grew in the absence of uracil. A yeast strain containing *Lac*Z developed in our laboratory previously was used as a negative control in enzymatic assays (Xu et al., 2018a).

2.2.3. Yeast cultivation and microsomal preparation

Yeast cultivation and microsomal preparation were carried out using a previously described method with slight modifications (Xu et al., 2018a). Briefly, *S. cerevisiae* H1246 strains hosting *AtPDAT1*, *BnDGAT1* or *LacZ* were first grown in liquid minimal media lacking uracil (0.67% yeast nitrogen base, 0.2 % SC-Ura and 2% raffinose) at 30ºC with shaking at 220 rpm for 24 h. These yeast cells were then used to inoculate induction media (liquid minimal media containing 2% galactose and 1% raffinose) to an initial OD_{600} of 0.2. Cultures were then grown under the same conditions until the OD_{600} reached approximately 6.0. Yeast cells were harvested by centrifugation at 3,000 g for 5 min, washed once with distilled water, and resuspended in a lysis buffer containing 20 mM Tris-HCl (pH 7.9), 2 mM dithiothreitol, 10 mM magnesium chloride, 1mM EDTA, 5% glycerol (by volume), and 300 mM ammonium sulfate. Harvested yeast cells were disrupted through homogenization with 0.5 mm glass beads in a bead beater (Biospec, Bartlesville, OK, USA), followed by centrifugation at 10,000 *g* for 20 min. The supernatant was subsequently recovered and centrifuged at 100,000 *g* for 70 min, and the resulting pellet (microsomal fraction) was resuspended in 0.1M potassium phosphate buffer (pH 7.2). Protein concentration was measured using the Bradford method with bovine serum albumin as the standard (Bradford, 1976). All steps were performed at 4ºC.

2.2.4. Synthesis of NBD-DAG

NBD-DAG was synthesized from NBD-PtdCho in a reaction catalyzed by phospholipase C as described previously (Sanderson and Venable, 2012). In brief, 1 mg of NBD-PtdCho was dissolved in 500 µl of a diethyl ether/ethanol mixture (98:2, by volume) in a screw cap culture tube**.** Subsequently, 15 µl of 0.02 M calcium chloride and 20 µl of phospholipase C (1 unit) were added to the glass tube. The mixture was then incubated at room temperature with agitation for 2 h. Following the reaction, lipids in the mixture were extracted using the Bligh and Dyer method (Bligh and Dyer, 1959), dried under nitrogen gas, resuspended in 100 µl of chloroform/methanol (2:1, by volume) and spotted on a pre-coated thin-layer chromatography (TLC) plate (SIL G-25, Macherey-Nagel, Düren, Germany). The TLC plate was developed in hexane/diethyl ether/methanol/acetic acid (70:30:5:1, by volume). The NBD-DAG band was visualized under UV light, scraped, and transferred to a screw cap culture tube. The synthesized NBD-DAG was extracted using the Bligh and Dyer method, dissolved in chloroform, purged with nitrogen and stored at -20ºC for further use.

2.2.5. *In vitro* DGAT assays

In vitro DGAT assays (used as a positive control for the production of NBD-TAG) were carried out using microsomes containing recombinant BnDGAT1, along with NBD-DAG and oleoyl-CoA as substrates as described previously with slight modifications (Sanderson and Venable, 2012; Xu et al., 2017). Briefly, reaction mixtures (50 µl) were composed of 238 mM HEPES-NaOH (pH 7.4), 3.85 mM $MgCl_2$, 400 µM NBD-DAG, 0.02% Tween 20 (by volume), 18 µM oleoyl-CoA and 10 µl of microsomes containing BnDGAT1 or LacZ (negative control). Reactions were initiated through the addition of the yeast microsomes, and were incubated at

30ºC for 60 min, followed by quenching with 100 µl of chloroform/methanol (2:1, by volume). All reactions (and hereafter) were performed in triplicate.

2.2.6. *In vitro* PDAT assays

Unless otherwise indicated, PDAT assays were conducted as described previously (Dahlqvist et al., 2000), except that NBD-DAG was used in place of radiolabeled DAG as the acyl acceptor. Yeast microsomes containing 40 µg of protein were aliquoted to reaction tubes, flash-frozen in liquid nitrogen and freeze-dried overnight. Subsequently, 1.5 nmol of NBD-DAG and 2.5 nmol of 18:1,18:1-PtdCho dissolved in 14 µl of benzene were added to the microsomes. The benzene was immediately evaporated under a stream of N_2 to avoid loss of enzymatic activity. The reaction was then initiated by adding 100 µl of 50 mM potassium phosphate buffer (pH 7.2). The reaction mix, which had 15 µM of NBD-DAG and 25 µM of 18:1,18:1-PtdCho, was incubated at 30ºC for 60 min, followed by quenching with 100 µl of chloroform/methanol (2:1, by volume).

For the determination of the time course of NBD-TAG production, samples were collected from 0 to 120 min. To verify the effects of yeast microsome amount in the assay, the equivalent of 0-240 µg of protein was individually used in the assays. To measure the enzymatic kinetics of AtPDAT1, different NBD-DAG concentrations in the range of 5 to 60 µM were used in the reactions. In order to validate the fluorescence-based assay, we also carried out the conventional PDAT assay using radiolabeled chemicals and a wide range of microsomal protein contents containing recombinant AtPDAT1 (from 0-160 µg of protein). In the conventional radioisotope-based PDAT assay, 1.5 nmol of 16:0,18:1-DAG and 2.5 nmol of 16:0,[^{14}C]18:1-PtdCho (equivalent of 15 µM of DAG and 25 µM of PtdCho in the final mix) were dissolved in 14 µl of benzene to be added in the dried microsomes. In the experiment to identify the ideal

alternative solvent to benzene, 14 µl of diethyl ether or ethanol were used to dissolve the substrates and all other parameters were kept the same.

2.2.7. Visualization of products formed in PDAT and DGAT assays

To separate the products formed in PDAT and DGAT assays with fluorescent chemicals, the chloroform phases (lipid fractions) of quenched reaction mixtures were spotted on TLC plates, which were then developed in diethyl ether/hexane/methanol/acetic acid (60:40:5:1, by volume). The TLC plates were then dried and 10 µl of NBD-PtdCho (0.01 mg/ml) was spotted at the top of each plate as an internal standard. The fluorescence of each spot was used for visualization purposes by scanning TLC plates in a Typhoon FLA 9500 (GE Healthcare, Mississauga, ON, Canada) with excitation and emission wavelengths of 495 nm and 519 nm, respectively. The photomultiplier tube (PMT) voltage was set to 250 V. In terms of the conventional PDAT assay using radiolabeled chemicals, the products of the enzymatic reactions were separated on TLC plates developed with hexane/diethyl ether/acetic acid (80:20:1.5, by volume). The radiolabelled TAG formed was located with phosphorimaging in the Typhoon FLA 9500 and quantified in a Beckman-Coulter LS6000.

2.2.8. Generation of a standard curve and quantification of NBD-TAG

A standard curve of NBD fluorescent intensity versus the amount of lipids was established as described previously (Sanderson and Venable, 2012). In brief, 10 µl aliquots of a dilution series of NBD-DAG (in the range of 0.5 pmol to 200 pmol) were loaded on TLC plates in triplicate. Fluorescence was detected on a Typhoon FLA 9500 as described above, and the intensity of each spot was quantified using ImageJ 1.52a (Schneider et al., 2012). The intensity of fluorescence

was then plotted against the amount of NBD-DAG, and the standard curve was generated by linear regression. To quantify NBD-TAG, fluorescent intensity was normalized using the NBD-PtdCho internal standard, and background fluorescence was subtracted from the total reading of the NBD-TAG band.

2.2.9. Statistical analyses

Plots and statistical analyses were carried out with GraphPad Prism 8. Error bars represent the standard deviation or standard errors of the results. Student's *t*-test was employed to compare the effects of different solvents in the PDAT assay.

2.3. Results and Discussion

2.3.1. Establishment of a PDAT assay with NBD-labelled substrate

In previous studies, PDAT assays have been conducted using phospholipids with a radiolabeled acyl chain at the *sn*-2 position and regular DAG as substrates (Dahlqvist et al., 2000; Stahl et al., 2004). To avoid the use of radioactive chemicals, I examined the effectiveness of a fluorescence-based method for assessing PDAT activity. The use of fluorescent substrates has also been estimated to reduce the cost of DGAT assays by 75% (McFie and Stone, 2011), which would provide another benefit to the development of a fluorescence-based PDAT assay. Lipids containing an NBD fluorescent group were chosen for the establishment of this assay since NBD has been successfully used previously in *in vitro* assays of other lipid biosynthetic enzymes, such as DGAT (McFie and Stone, 2011; Sanderson and Venable, 2012).

On the commercial NBD-PtdCho, the relatively big NBD molecular group is linked to the end of the acyl chain at the *sn*-2 position, which is transferred to the *sn*-3 position of DAG in the

PDAT reaction. To minimize the potential effects of the NBD molecular structure on PDAT assay, NBD-DAG and regular PtdCho were used in this study. Since NBD-DAG is not commercially available, this compound was synthesized from NBD-PtdCho. This synthesized NBD-DAG contains a palmitoyl chain at the *sn*-1 position and an NBD molecule linked to a dodecanoyl chain at the *sn*-2 position. Moreover, this method design would allow the testing of the enzyme preference for different acyl chains (e.g. palmitoyl, oleoyl, linoleoyl) in the donor PtdCho.

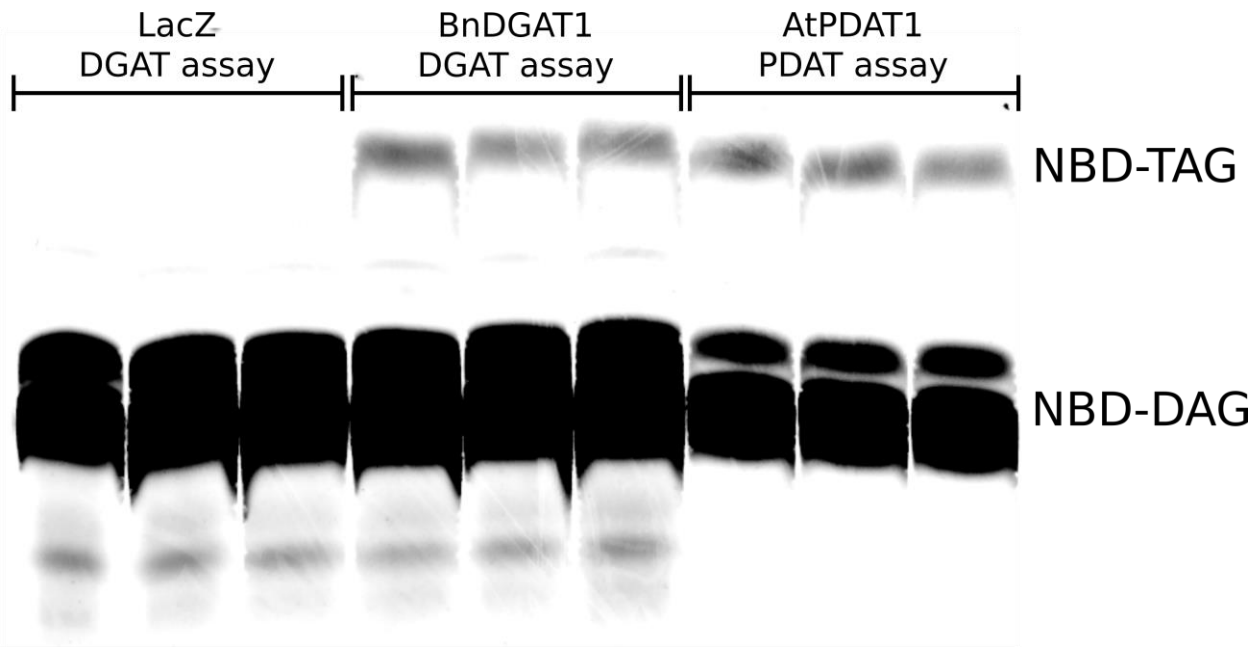
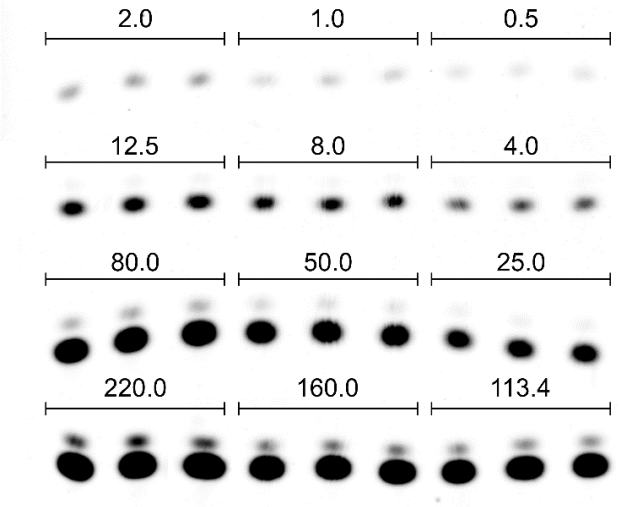


Figure 2.2. A representative TLC plate displaying the results of enzymatic assays using fluorescence-labeled substrate. Microsomal preparations from the *Saccharomyces cerevisiae* quadruple mutant H1246 transformed with either AtPDAT1 and BnDGAT1 were used as enzymes in the PDAT and DGAT assays, respectively. For DGAT assay, fluorescent 1-palmitoyl-2-dodecanoyl-NBD-*sn*-glycero-3-glycerol (NBD-DAG) and oleoyl-CoA were used as substrates. NBD-DAG spots are composed of 1,2-NBD-DAG and 1,3-NBD-DAG. Microsomes derived from yeast cells transformed with the *LacZ* gene were used as the negative control. Following enzymatic assays, lipid fractions were extracted and separated on TLC plates. NBD-TAG, NBD-labeled triacylglycerol. AtPDAT1, *Arabidopsis thaliana* phospholipid:diacylglycerol acyltransferase 1; BnDGAT1, *B. napus* acyl-CoA:diacylglycerol acyltransferase 1

A



B

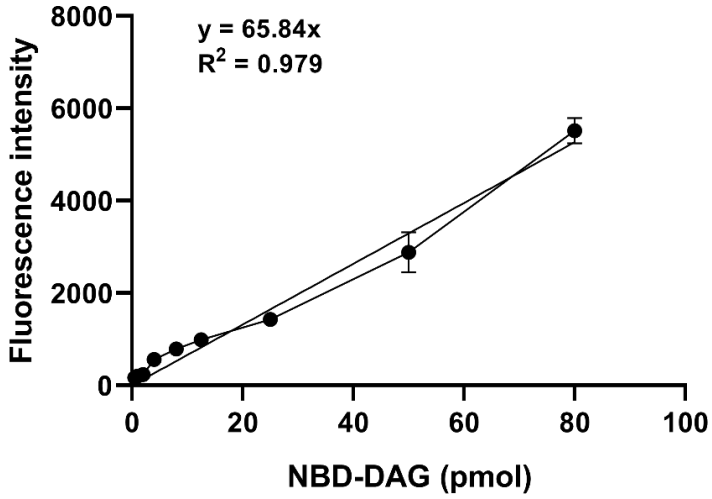


Figure 2.3. Establishment of a standard curve for the quantification of PDAT activity. (A) Scanned image of a representative TLC plate showing the fluorescence intensity of 1-palmitoyl-2-dodecanoyl-NBD-*sn*-glycero-3-glycerol (NBD-DAG) in a series of molar concentrations. (B) Standard curve for the quantification of NBD-labeled molecular NBD-DAG or NBD-triacylglycerol (NBD-TAG**).** The linear range varies from 0.5 pmol to 80 pmol of NBD-DAG.

Similar to NBD-DAG, NBD-TAG, which is required to confirm the success of the PDAT reaction with NBD-DAG as substrate, is also not commercially available. To overcome this, recombinant BnDGAT1 was used to convert NBD-DAG to NBD-TAG. As shown in Figure 2.2, NBD-TAG synthesized by BnDGAT1, which has been validated in the DGAT1 reaction, has the same migration distance as the product of recombinant AtPDAT1 on a TLC plate, confirming that the product of AtPDAT1 was indeed an NBD-TAG molecule.

In order to quantify NBD-TAG produced in PDAT assays, a standard curve was generated based on the fluorescence of known quantities of NBD-DAG (Figure 2.3). When different dilutions of NBD-DAG (0.5 - 220 pmol) were spotted on a TLC plate, the lowest detectable amount of NBD-DAG was 0.5 pmol (Figure 2.3A). This detection limit is similar to that (0.1 pmol) reported in a previous study (Sanderson and Venable, 2012), even though they worked with different fluorescence scanners and PMT voltage (250 V in the present study versus 420 V in the other study). In any case, the production of NBD-TAG by AtPDAT1 was substantially higher than 0.5 pmol in the enzymatic assay, and therefore the detection limit observed here was sufficiently sensitive for our purposes.

When fluorescence intensity was plotted against the amount of NBD-DAG, a standard curve was obtained by linear regression of the linear region from 0.5 pmol to 80 pmol of NBD-PtdCho (Figure 2.3B). The amount (pmol) of NBD-TAG produced in PDAT reactions could then be quantified using the equation *NBD-TAG = 0.01519 x corrected fluorescence intensity* (R^2=0.979).

The suitability of using NDB-DAG in the PDAT assay was further analyzed through enzyme structure analysis. Although the crystal structure of PDAT has not yet been reported, the recently released structures of human LCAT and lysosomal phospholipase A2, which belong to

the same family as PDAT, could provide some insights into the catalytic mechanisms of PDAT (Glukhova et al., 2015; Piper et al., 2015). These enzymes all have a catalytic triad (Ser181, Asp345, and His377 in LCAT) at the base of the catalytic pocket, which is covered by a lid-loop (Piper et al., 2015; Xu et al., 2018a). During catalysis, the lid-loop moves aside from the opening of the catalytic pocket, allowing phospholipids to enter and interact with the catalytic triad. The catalytic triad then attaches to the *sn*-2 acyl-chain of a phospholipid as a nucleophile to form an acyl-enzyme intermediate, followed by the transfer of the acyl-chain to cholesterol or DAG to generate an ester (Xu et al., 2018a). Based on this catalytic mechanism, the NBD group in the acyl chain of NBD-DAG would be far from the catalytic triad, and thus unlikely to cause substantial negative effects to the catalysis process. Therefore, in addition to our *in vitro* analyses (Figures 2.3-2.8), the enzymatic reaction mechanism also supports the suitability of using DAG with fluorescent-labeled acyl chains in *in vitro* PDAT assays.

2.3.2. Characterization of AtPDAT1 using a fluorescence-based PDAT assay

After confirming that a fluorescence-based method was suitable for *in vitro* PDAT assays, I further investigated if this method would be suitable for the characterization of AtPDAT1. The time course of the enzymatic assay indicated that the formation of NBD-TAG could increase in a linear fashion up to a maximum of 60 min of reaction time (Figure 2.4; $R^2 = 0.995$). Unexpectedly, the microsomes of *S. cerevisiae* H1246 expressing *LacZ* produced very low amounts of an unknown fluorescent compound with a similar migration distance to NBD-TAG on TLC plates (Figure 2.4), but PDAT reactions using boiled microsomes as the enzyme source did not (Figure 2.5), which suggests the presence of an unknown but very weak enzymatic reaction in our *LacZ* negative control microsomes. Since *S. cerevisiae* H1246 is a quadruple

mutant that lacks all genes necessary for neutral lipid biosynthesis (*DGA1*, *LRO1*, *ARE1*, and *ARE2*; Sandager et al., 2002), the resulting bands in the negative control microsomal samples were unlikely due to any presence of PDAT or TAG. These fluorescence signals generated in the negative control samples were thus treated as a background signal and subtracted from the fluorescence signals generated by the recombinant AtPDAT1.

The effects of protein content on PDAT activity was also subsequently analyzed. As shown in Figure 2.6 the NBD-TAG formation increasing in a linear fashion up to a maximum of 80 μg of microsomal protein (R^2 = 0.974). The biosynthesis of fluorescent TAG did not increase any further with higher enzyme amounts, which was likely due to a limitation of available substrates. This result was consistent with a previous study involving a fluorescence-based assay of DGAT activity, where a protein content above 50 μg did not yield any further increases in NBD-TAG production (McFie and Stone, 2011). In addition, there is a good correlation (R^2 = 0.979) between the fluorescence-based PDAT assay and the conventional assay that was carried out with 16:0,[^{14}C]18:1-PtdCho and 16:0,18:1-DAG as the substrates (Figure 2.7).

The enzyme kinetics of AtPDAT1 in response to different concentrations of NBD-DAG were also assessed. As shown in Figure 2.8, AtPDAT1 displayed a Michaelis-Menten response to an increasing concentration of NDB-DAG with apparent V_{max} and K_m values of 11.1 pmol NBD-TAG/min/mg protein and 32.2 μM, respectively. Similarly, LCAT, which belongs to the same family as PDAT, also shows Michaelis-Menten kinetics (Pan et al., 2015; Sakurai et al., 2018).

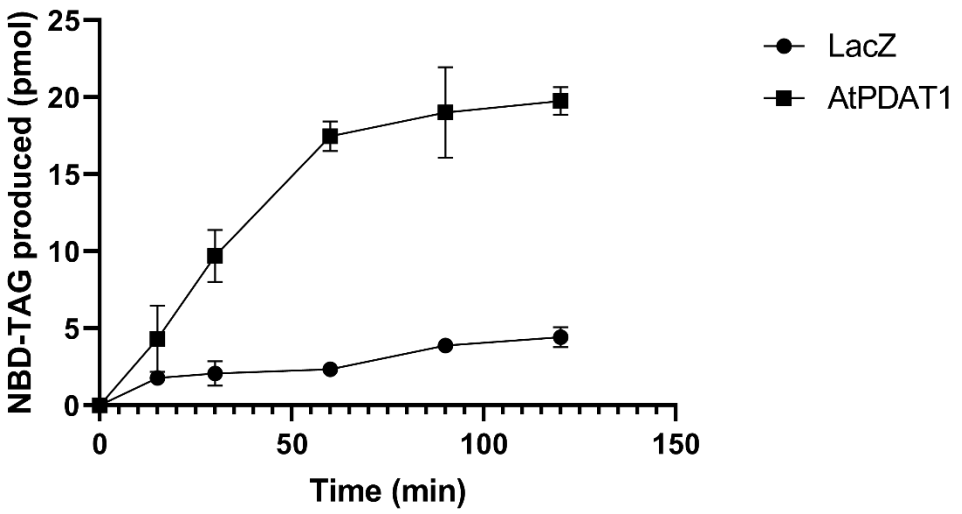


Figure 2.4. Time course curve of PDAT enzymatic reactions. Nitrobenzoxadiazole-labeled triacylglycerol (NBD-TAG) produced from NBD-diacylglycerol (NBD-DAG) over time in the presence of microsomal fractions from yeast heterologously expressing *Arabidopsis thaliana pdat1* (*Atpdat1*). Microsomes derived from yeast cells transformed with the *LacZ* gene were used as a negative control. The reactions were catalyzed using 40 µg of microsomal protein in each case. All reactions were carried out with 1.5 nmol NBD-DAG, 2.5 nmol 18:1,18:1-PtdCho, 100 µl of phosphate buffer (pH 7.2) at 30ºC. NBD-TAG production increases linearly up to 60 min (R^2 = 0.995). n = 3. Values represent mean ± standard deviation.

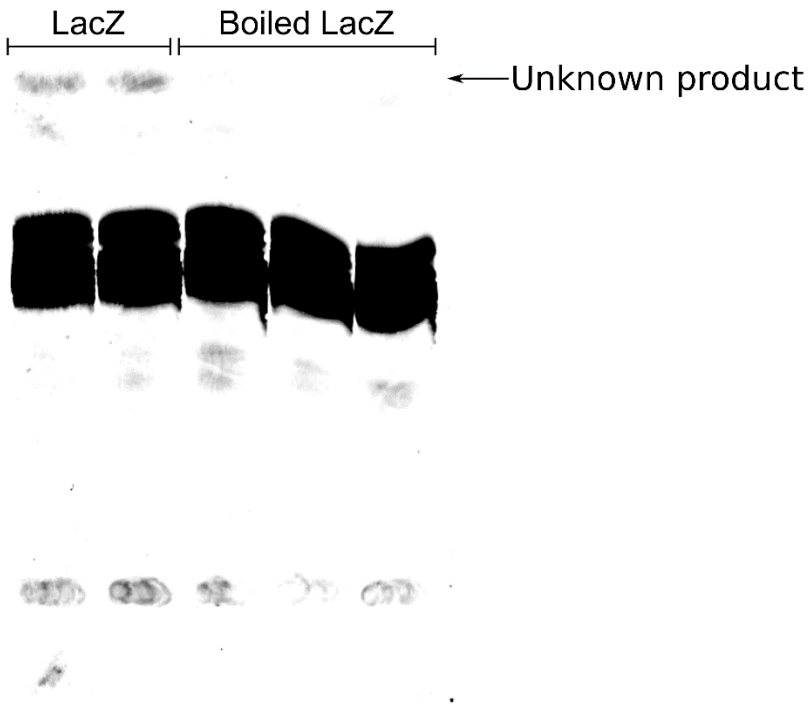


Figure 2.5. Production of an unknown product with similar migration distance to nitrobenzoxadiazole-labeled triacylglycerol (NBD-TAG). The compound is not present in boiled microsomal fractions from yeast transformed with *LacZ*, which suggests an enzyme-catalyzed reaction. The reactions were carried out using 40 µg of microsomal protein in each case. All reactions were added 1.5 nmol NBD-DAG, 2.5 nmol 18:1,18:1-PtdCho, and 100 µl of phosphate buffer (pH 7.2) and incubated at 30ºC for 60 min.

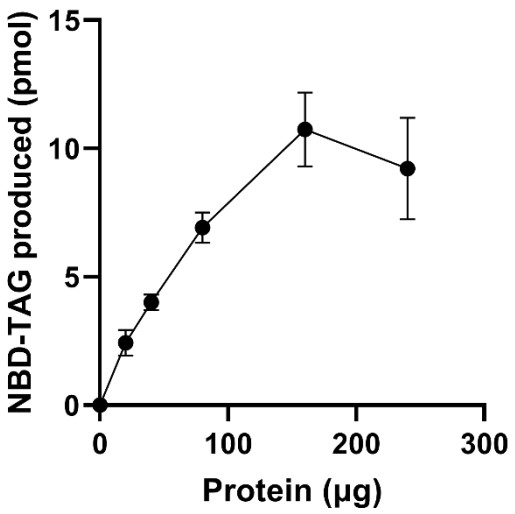


Figure 2.6. Production of nitrobenzoxadiazole-labeled triacylglycerol (NBD-TAG) in an enzymatic reaction catalyzed by *Arabidopsis thaliana* PDAT1 (AtPDAT1) at different concentrations. Microsomes derived from yeast cells transformed with the *LacZ* gene were used as the negative control and weak fluorescent signals generated were subtracted from the corresponding AtPDAT1 reactions as background. Microsome protein contents varied from 0 to 240 µg in the reactions. All reactions were carried out with 1.5 nmol NBD-DAG, 2.5 nmol 18:1,18:1-PtdCho, 100 µl of phosphate buffer (pH 7.2) at 30ºC for 60 min. n = 3. Values represent mean ± standard deviation

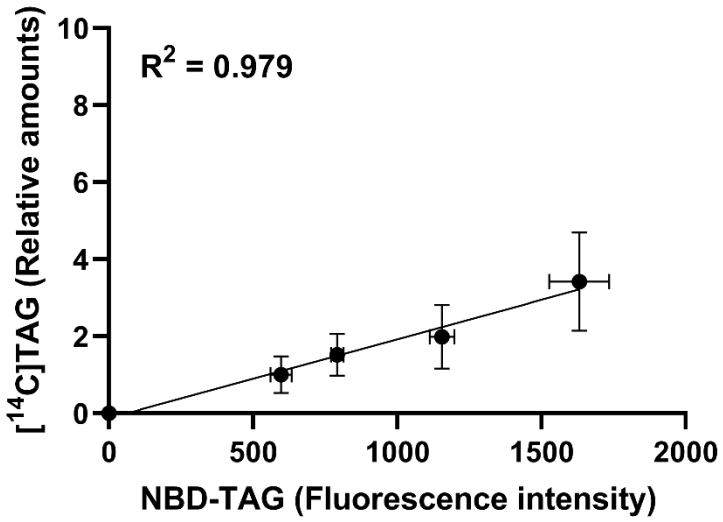


Figure 2.7. Comparison between the fluorescence-based PDAT assay and the conventional assay with radiolabeled chemicals for a wide range of microsomal protein amounts (20 – 160 μg) containing *Arabidopsis thaliana* PDAT1 (AtPDAT1) shows a good correlation (R^2 = 0.974). The conventional assay reactions were carried out with 2.5 nmol 16:0,[^{14}C]18:1-PtdCho, 1.5 nmol 16:0,18:1-DAG, 100 μl of phosphate buffer (pH 7.2) at 30ºC for 60min. The values of the first reaction were used to normalize the results of the conventional assay. n = 3. Values represent mean ± standard error. The data from the standard assay was generated by Dr. Yang (Sarena) Xu.

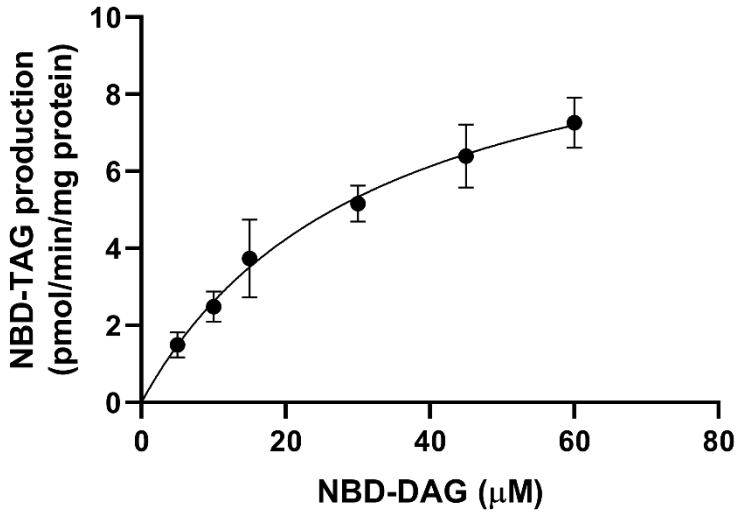


Figure 2.8. Rate of nitrobenzoxadiazole-labeled triacylglycerol (NBD-TAG) production by *Arabidopsis thaliana* PDAT1 (AtPDAT1) reveals Michaelis-Menten kinetics (R^2= 0.934) to increasing NBD-diacylglycerol (NBD-DAG) concentrations. Plots were generated using GraphPad Prism 8. The final concentration of NBD-DAG in reactions was increased from 5 μM to 60 μM. The final concentration of 18:1,18:1-PtdCho was kept at 25 μM in all reactions. n = 3. Reactions were carried out at 30ºC for 60 min. Values represent mean ± standard deviation

2.3.3. Substitution of benzene in PDAT assays

Benzene is used in PDAT assays to dissolve and deliver substrates to the microsomal fractions. After being added to the reaction tube, benzene must be immediately evaporated to prevent loss of protein activity. Although benzene works well in PDAT assays, this solvent is highly hazardous. As a known carcinogenic substance, benzene is also harmful to renal, cardiovascular, respiratory, and reproductive systems (Bahadar et al., 2014). Even with safety measures such as fume hoods and special gloves to avoid or minimize exposure to this compound, benzene poses a risk to the health of laboratory workers. In order to identify an alternative solvent, diethyl ether

and ethanol were tested as substitutes in our PDAT assays, with all other conditions remaining the same. As shown in Figure 2.9, the formation of NBD-TAG with diethyl ether as the solvent was statistically equivalent to results obtained using benzene, whereas NBD-TAG production was significantly lower with ethanol than with benzene as the solvent. Diethyl ether was used as a safe anesthetic for medical purposes for many years though it has some undesirable side effects such as post-anesthetic nausea and vomiting (Bovill, 2008). Therefore, this low-toxic solvent can provide an excellent substitution for carcinogenic benzene in PDAT assays.

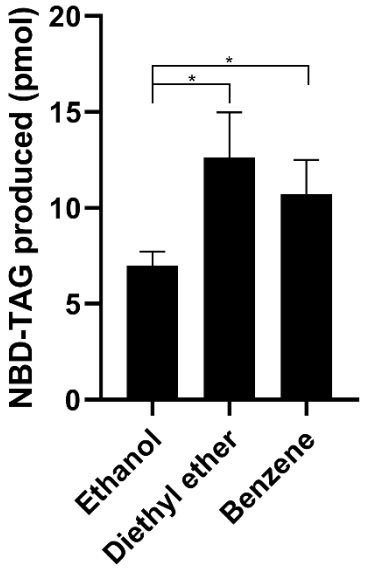


Figure 2.9. The effects of different solvents on fluorescence-based PDAT assays. Ethanol, diethyl ether or benzene was used as solvents for the addition of substrates to microsomal fractions. Student's t-tests indicate that diethyl ether and benzene are equivalent in terms of NBD-TAG production, and both solvents are significantly superior to ethanol in terms of triacylglycerol production (*$p<0.05$). All reactions were carried out with 1.5 nmol NBD-DAG, 2.5 nmol 18:1,18:1-PtdCho, 100 µl of phosphate buffer (pH 7.2) at 30ºC for 60 min. n = 3. Values represent mean ± standard deviation

2.4. Conclusions

In summary, a fluorescence-based assay for the quantitative analysis of PDAT activity was established in this study, which works well with a broad range of microsomal protein contents and reaction times, and thus should provide a means to assay the activity and substrate specificity of PDATs. Our results demonstrated that NBD-DAG is a suitable substrate for *in vitro* assays of PDATs, which is a safer, less costly and more convenient alternative to radioactive chemicals. The sole use of NBD-DAG is a limitation of the method regarding the test of different DAG molecular species. The fluorescence-based and the conventional PDAT assays have a good correlation, which validates the fluorescence-based method. I also showed that the highly toxic solvent benzene could be substituted with the low-toxic diethyl ether without affecting the formation rate of NBD-TAG. In addition, this fluorescence-based PDAT method also set up the foundation for further adaptation and development for various research purposes, such as comparing various substrates, using different NBD-labelled chemicals, and combining with HPLC to develop a robust medium throughput protocol for PDAT assay. Moreover, this method could be further improved to assay purified PDAT, which would reduce the influence of microsomal phospholipids in the reactions to test acyl specificity.

CHAPTER 3. The evolution of PDAT in photosynthetic organisms was marked by functional divergence and positive selection which were essential to higher plants but not green microalgae

Abstract

Phospholipid:diacylglycerol acyltransferases (PDATs) belong to the lecithin:cholesterol acyltransferase (LCAT) family and catalyze the biosynthesis of triacylglycerol in plants, algae and yeasts. PDATs are important for plant development and stress response and contribute to the incorporation of unusual fatty acids in some species. Therefore, they have been regarded as key targets for improving and manipulating plant performance and oil production. However, (i) the evolutionary and structural properties of algal PDATs are not fully characterized across algal species, (ii) their relationship with plant PDATs is unexplored, and (iii) their relationship with other LCAT-like is also unknown. In this study, combined approaches including *in silico* analysis, site-directed mutagenesis and enzymatic characterization were carried out to explore these questions. The results show that algal enzymes from LCAT family are usually longer than their plant homologs. In addition, the analysis of conserved motifs reveals the region between the "lid" domain and the catalytic serine in PDATs as the most conserved between algal and plant PDATs. In terms of evolution, the presence of functional divergence and positive selection between plant and green algal PDATs was confirmed. I also selected the first residue downstream the catalytic serine (methionine 255) for site-directed mutagenesis. This residue evolved under positive selection in green algae and is conserved as an aromatic amino acid, whereas a conserved methionine is present in plant PDATs. The mutation of methionine 255 to either tryptophan or alanine in *A. thaliana* PDAT1 destroyed its enzymatic activity, confirming

the relevance of conserved amino acids in this position for the enzyme conformation and activity. Overall, the results demonstrated that PDATs evolved to be essential to plants. In green algae, they evolved under more relaxed constraints likely because of the presence of multiple enzymes with a similar function.

3.1. Introduction

Although fossil fuels are a primary source of energy in many countries, their usage brings several disadvantages. Petroleum reservoirs are restricted to only a few regions and its prices fluctuate due to many factors, such as geopolitical crisis and limited supply (Baumeister and Kilian, 2016). It creates instability and threats the national security of many countries, especially the poorer ones. In addition, the use of fossil fuels releases large amounts of carbon dioxide (CO_2), carbon monoxide (CO), particulate matter and other air pollutants. They are believed to contribute to high pollution levels in the cities, and some of them, on a larger scale, contribute to global warming. Therefore, alternative fuel sources are necessary to meet future energy consumption and create energy security for the countries (Ang et al., 2015; Sovacool and Saunders, 2014). In a similar manner, there are many fossil fuel-based products that require feedstocks from biomass to replace them (Harmsen et al., 2014; Nikolau et al., 2008). These alternative fuels and feedstocks should be available for most countries and contribute to a cleaner environment.

A handful of plant species are able to produce large amounts of oils in the seeds or fruits (Durrett et al., 2008) and thus have been widely used for the production of biodiesel and other bioproducts (Rico and Sauer, 2015; Souza et al., 2018). Currently, soybean, palm fruit, rapeseed and sunflower are the most used feedstock in biodiesel production worldwide. In addition,

certain plant species can also produce special oils, such as ricinoleic acid, which are of important industrial value (Dyer et al., 2008; McKeon, 2016). Algal species also have the potential to supply the demand for liquid fuels and other compounds because of their high content of lipids (Hu et al., 2008). For instance, a *C. zofingiensis* strain and a genetically-modified *Phaeodactylum tricornutum* strain could accumulate 65% and 58% of their dry cell weight in the form of lipids, respectively (Feng et al., 2012; Xue et al., 2015). The use of algae and plants for biofuel production also contributes to a cleaner environment, because they are able to capture CO_2 during photosynthesis. In this case, the net emissions of CO_2 are reduced. Algae are additionally able to use residues from agriculture or industries for growth and their cultivation can release arable land for food crops (Benedetti et al., 2018; de Carvalho Lopes et al., 2018). They could substitute or complement the production of lipids by plants because of their capacity to grow faster and adapt to a diversity of environments (Hu et al., 2008). Therefore, understanding the evolution and adaptation of plants and algae to environments, as well as how lipid production occurs in algae and plants, are necessary to choose the right species and improve their lipid production.

Algae and plants are taxonomically assigned to several groups of species (Cavalier-Smith, 2018; Leliaert et al., 2012; Morris et al., 2018). Algae are ubiquitous organisms that can be found in a diversity of environments such as snow, high-salt ponds, and marine water and freshwater bodies (Hu et al., 2008; Leliaert et al., 2012). Indeed, they belong to diverse taxonomic groups, which helps explain why these organisms are so adaptable. Algae have undergone an endosymbiosis in their evolutionary history, which distinguishes these organisms. The primary endosymbionts comprise the green algae (Chlorophyta), red algae (Rhodophyta) and Glaucophyta. The plastid in these organisms is an ancient cyanobacterium that was engulfed

and established an endosymbiotic relation with the host (Brodie et al., 2017; Clerck et al., 2012). The green algae belong to the Viridiplantae, which also contains the land plants. The secondary endosymbionts comprise the heterotrophic organisms that engulfed a green alga or a red alga, which provided them their photosynthetic machinery (Archibald, 2012). These organisms can be classified as chromists (belong to the Kingdom Chromista; Cavalier-Smith, 2018), which comprise relevant algal species in the Stramenopiles, Haptophytes, and Cryptophytes. For instance, diatom, which is a dominant group in the phytoplankton, belongs to the Stramenopiles (Zulu et al., 2018).

Algae and plants accumulate neutral lipids mostly in the form of triacylglycerol (TAG), which is formed by a glycerol backbone acylated with three fatty acids (Xu et al., 2018a). In algae, TAG accumulation is often triggered by adverse environmental conditions (e.g. nutrient limitation and high light intensity), while in many plant species neutral lipids are accumulated in seeds as an energy reservoir (Li-Beisson et al., 2015; Xu et al., 2018a; Zulu et al., 2018). The biosynthesis of TAGs in these organisms has two mechanisms: the acyl-Co-enzyme A (CoA)-dependent pathway (also known as the Kennedy pathway) and the acyl-CoA-independent pathway. The first mechanism involves the sequential acylation of a G3P molecule by different enzymes (Figure 1.2). In the final step of the Kennedy pathway, DGAT transfers an acyl group from an acyl-CoA to the *sn*-3 position of a DAG to form a TAG molecule. The acyl-CoA-independent pathway differs solely in this step, which is catalyzed by PDAT. This enzyme uses a phospholipid as the acyl-donor in the TAG biosynthesis instead of acyl-CoA (Li-Beisson et al., 2015; Zulu et al., 2018). PDAT and DGAT complement each other for the TAG biosynthesis in plants and algae (Yoon et al., 2012; Zhang et al., 2009). Therefore, both enzymes are targets for

potential biotechnological approaches to increase the accumulation of lipids in plants and microalgae.

Although DGAT and PDAT both catalyze the production of TAG, these enzymes belong to different enzyme families. There are at least three types of enzymes with DGAT activity that share no homology and belong to different families, designated DGAT1, 2 and 3, respectively (Cases et al., 1998; Lardizabal et al., 2001; Saha et al., 2006). PDAT belongs to the LCAT family and shares a structural similarity with the human LCAT (Pan et al., 2015). This family of enzymes also contains a few phospholipases and a phospholipid:sterol O-acyltransferase (PSAT).

The function of PDAT in lipid metabolism has been studied in different plant species after the first identification of this enzyme. Interestingly, PDAT can use various substrates in the biosynthesis of TAGs (Dahlqvist et al., 2000). For instance, the microsomes of *Crepis palaestina* developing seeds preferentially incorporate vernoloyl groups, whereas those from *R. communis* developing seeds can transfer both ricinoleoyl and vernoloyl groups from phospholipids to DAG. In addition, the sunflower (*Helianthus annuus)* microsomes catalyzed the acylation of DAG with ricinoleoyl, vernoloyl, and oleoyl groups. Similarly, flax PDATs were observed to prefer linolenoyl groups (Pan et al., 2013). All these results indicated the importance of PDAT in the incorporation of unusual fatty acids to TAG molecules . Moreover, the head group in the phospholipid also affects PDAT activity (Dahlqvist et al., 2000). Considering their mechanism of action and their flexibility for different substrates, PDATs could be the target enzyme to increase the production of unusual fatty acids via biotechnology.

Due to the importance of PDATs in lipid biosynthesis, a better understanding of these enzymes in microalgae and plants is necessary. While plant PDATs have been well characterized

in a recent study, PDATs have been only characterized in two algal species (*C. reinhardtii* and *M. incisa*), which are both green algae (Liu et al., 2016; Pan et al., 2015; Yoon et al., 2012). Therefore, it is necessary to characterize algal PDAT broadly, especially from groups other than green algae. In addition, although algae (especially green algae) and plants are relatively close in taxonomy, the relationship between algal and plant PDATs is largely unexplored. The evolution of plant PDATs after the divergence from green algae could provide more information about the roles of these enzymes in the organisms. Moreover, the role of other LCAT-like enzymes in algae is largely unknown. Specifically, there are three major questions remaining unanswered: (i) Are algal and plant PDATs similar? (ii) Have they evolved to contribute equally to TAG production and fitness in algae and plants? (iii) What is their relationship with other enzymes from the LCAT family?

To perform a comprehensive analysis of algal and plant PDATs to address the above questions, 39 reported algal genomes were searched for PDAT homologs in this study. The properties of PDATs from green algae, red algae, and chromists were analyzed and compared with PDATs from 17 plant species. Conserved regions, important domains and the effect of putative mutations were analyzed and discussed. The relationship of PDATs with other LCAT-family members in algae is also analyzed. In particular, functional divergence and positive selection during the evolution of algal and plant PDATs were present. Among the identified residues under positive selection, the first site after the catalytic serine was chosen as a representative amino acid residue for additional investigation. The presence of aromatic amino acids at this position is shown to be unique to green algae. Moreover, site-directed mutagenesis in this position could destroy the activity of *A. thaliana* PDAT, indicating the critical role of this amino acid residue. Overall, the results show that PDATs are more important for plants than for

green algae and that conserved sites that evolved under positive selection may be essential for the activity of these enzymes.

3.2. Methodology

3.2.1. Identification of PDAT homologs in algae and plants

PDAT homologs were searched and retrieved from the available plant and algal genomes from Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html, last accessed at 28/11/2017), Genome Portal (https://genome.jgi.doe.gov/portal/, last accessed at 16/03/2017), Ensembl Genomes (http://ensemblgenomes.org/, 17/10/18), Genome - NCBI (https://www.ncbi.nlm.nih.gov/genome/, 17/10/18), and websites from individual projects. The genomes and their sources used in this study are listed in Supplementary table A.1. Algal genomes and *B. napus*, *Brassica rapa*, *Brassica oleracea capitata*, *Marchantia polymorpha*, and *R. communis* genomes were searched locally with TBLASTN v2.6.0 and BLASTP v2.6.0 from the Blast+ suite (Camacho et al., 2009). Two *A. thaliana* PDAT homologs (AT5G13640, AT3G44830) and the *C. reinhardtii* PDAT (Cre02.g106400) were used as queries in the BLAST searches. A cut-off of $1x10^{-9}$ was applied to the e-value to select PDAT homologs for further analyses. Additional PDAT homologs from other plant genomes used in this study were directly retrieved from the Phytozome database based on the findings of a previous study (Pan et al., 2015).

To improve the visualization of the results, each sequence was given a new identifier. The acronym was composed by the first character of the genus name followed by the first two characters of the species name. These three letters are followed by PDAT, PSAT, PLA or LCAT (this one represents members of LCAT family not covered by the PDAT, PSAT and PLA

subfamilies). The classification of the sequences follows the results from InterProScan as shown in section 3.2.2. In addition, a suffix is used to distinguish similar genes from the same organism. For example, *Symbiochloris reticulata* PDATs are called SrePDATa and SrePDATb. The identification of PDAT from *A. thaliana* and *C. reinhardtii* follows previous publications. The complete list of genes used in this study and their acronyms are found in Supplementary table A.2.

3.2.2. Characterization and classification of LCAT-like sequences

Protein sequences were classified with PANTHER (Mi et al., 2017) as implemented in InterProScan v5.22.61 (Jones et al., 2014). In addition, the length, isoelectric point (pI), molecular weight (MW), and amino acid composition of the protein sequences were estimated with CLC Sequence Viewer v7.0 (Qiagen Bioinformatics, 2017). The TMDs were predicted with TMHMM v2.0 program server (http://www.cbs.dtu.dk/services/TMHMM/, (Krogh et al., 2001). Conserved motifs were searched with MEME (Multiple Expectation Maximization for Motif Elicitation) v4.12.0 (Bailey and Elkan, 1994). The parameters implemented in MEME were "anr" mode, 100 motifs (nmotifs), minimum width of 3 (minw), a maximum of 300 (maxw), minimum of 3 sites (minsites), and a maximum of 110 (maxsites). The identification of exons and introns was carried out with the Gene Structure Display Server (GSDS) v2.0 (Hu et al., 2015).

3.2.3. Phylogenetic analysis

Multiple sequence alignment of the putative PDAT homologs was carried out with the L-INS-i method implemented in MAFFT v7.271 (Katoh and Standley, 2013). The aligned dataset was

used by TranslatorX (Abascal et al., 2010) (http://translatorx.co.uk/) to generate the nucleotide alignment of the sequences. Then, poorly aligned regions were removed with trimAl v1.2 (Capella-Gutiérrez et al., 2009). The best-fit model of nucleotide substitution was identified by jModelTest 2 (Darriba et al., 2012).

The phylogenetic tree construction was carried out with two different methods: ML and BI. MrBayes v3.2.6 (Ronquist et al., 2012) generated the BI trees with the following parameters implemented in all analyses: 10 million generations, two runs, eight Markov chains, and temperature of 0.1. Branch lengths are unconstrained with the priors set to be an exponential distribution with the value of 10.0. In addition, 25% of the samples from the beginning of the chain were discarded as burnin. The convergence was reached if the average standard deviation of split frequencies was lower than 0.01 by the end of the analysis. In addition, the estimated sample size (ESS) should always be higher than 200 and the potential scale reduction factor (PSRF) should be close to 1.000 for all parameters. The ML tree was generated with PhyML v3.0 (Guindon et al., 2010). The obtained tree was the best one among those generated with the Nearest Neighbor Interchange (NNI) method and the Subtree Prune and Regraft (SPR) method. Branch supports were estimated by the Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT) method (Anisimova et al., 2011). Finally, the trees were visualized with iTOL v3 (Letunic and Bork, 2016).

The topological hypothesis test was carried out with the stepping-stone sampling method implemented in MrBayes v3.2.6 (Xie et al., 2011). Two independent runs with 8 chains each were used in the analysis. The runs were performed with 50 steps with 2,000,000 generations each and an initial burnin of 2,000,000 generations (a total of 102,000,000 generations). Each analysis was carried out in triplicate.

3.2.4. Selection pressure analysis

Selection pressure analyses were performed by the CODEML program implemented in the package PAML v4.9 (Yang, 2007). Dataset alignments were generated as described in section 2.3, which outputs codon alignments. The phylogenetic trees were constructed using Phyml v3.0, after a jModelTest 2 search for the best-fit model of nucleotide substitution. The F3x4 model of codon frequencies was used in all analyses. The selection pressure analyses estimate a ω ratio (non-synonymous/synonymous ratio or dN/dS), which is the ratio between the rate of non-synonymous substitutions (dN) to the rate of synonymous substitutions (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. $\omega>1$, $\omega=1$, and $0<\omega<1$ indicates positive selection, neutral evolution, and purifying selection, respectively.

The hypothesis of functional divergence was tested using the Clade model C (model = 3, non-synonymous sites = 2; Bielawski and Yang, 2004; Yang et al., 2005). This model divides the phylogenetic tree into the foreground partition (the branches of interest) and the background partition (the remaining branches). Then, it estimates the ω ratio for 3 different site classes, while it calculates the proportion of sites into each site class. Site class 0 contains the sites under purifying selection ($0<\omega_0<1$) and site class 1 fixes $\omega_1=1$ (neutral evolution). While these ratios are equal in both partitions, site class 2 estimates divergent ω ratios for the background (ω_2) and foreground (ω_3) partitions. Clade model C is usually compared to M2a_rel model (null model; model = 0, non-synonymous sites = 22), where site class 0 and site class 1 are similar to Clade model C, but it estimates the same ω ratio for both partitions in the phylogenetic tree for site

class 2 (Weadick and Chang, 2012). The significance of Clade model C results is calculated by a likelihood ratio test (LRT), which compares the log-likelihood (ℓ) of both models (LRT = 2 x [$\ell_{modelC} - \ell_{M2a\_rel}$]) against a χ^2 distribution. The degree of freedom (df) is the difference between the number of parameters of each model and is usually equal to 1. Different initial ω ratios were used ($0<\omega<2$) for both Clade model C and M2a_rel to avoid local optima.

Positive selection was estimated using branch-site model A, which detects positive selection in only a few sites of the foreground branches (Yang et al., 2005; Yang and Nielsen, 2002; Zhang et al., 2005). Similar to the Clade model C, branch-site model A divides the tree into a foreground and background partitions, but the branch-site model has 4 site classes. Site class 0 detects the sites under purifying selection ($0<\omega_0<1$), while site class 1 estimates sites under neutral evolution ($\omega_1=1$) in both partitions. The sites under positive selection in the foreground are divided into site class 2a and 2b. The first detects purifying selection in the background, while site class 2b estimates the sites under neutral evolution in the background. The Bayes empirical Bayes (BEB) method is implemented in the branch-site model A to estimate the posterior probability of each site to be under positive selection (Yang et al., 2005). The null model fixes $\omega_2=1$. The LRT is equal to the one used in the analysis of functional divergence and the results are also compared against a χ^2 distribution (which reduces the chance of false-positives). Different initial κ values were used to avoid local optima.

3.2.5. Protein structure prediction

The models of AtPDAT1 and its variants were obtained through structural modeling using SWISS-MODEL software (Waterhouse et al., 2018). A phospholipase with 26.54% sequence homology with AtPDAT1 was used as a template (Glukhova et al., 2015). The structure of

HsaLCAT was previously obtained in complex with a Fab fragment and was used to determine the structure of the protein alone using SWISS-MODEL as well as those of HsaLCAT variants (Piper et al., 2015). For CrPDAT and its variants, an LCAT structure having 29% identity with CrPDAT was used as a template in the PHYRE2 software (Glukhova et al., 2015; Kelly et al., 2015).

3.2.6. Heterologous expression of *AtPDAT1* and its mutants

Site-directed mutagenesis of *AtPDAT1* was carried out with overlap extension PCR. The primers used in this study are listed in Supplementary table A.3. 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). The yeast transformation was performed using the lithium acetate/single-stranded carrier DNA/PEG method (Gietz and Schiestl, 2007). The recombinant strains were first grown in liquid minimal media lacking uracil (0.67% yeast nitrogen base, and 0.2 % SC-Ura) with 2% raffinose, followed by cultivation in liquid minimal media containing 2% galactose and 1% raffinose (Xu et al., 2018b). The yeast cultures producing the recombinant enzymes were harvested and resuspended in a lysis buffer containing 20 mM Tris-HCl (pH 7.9), 2 mM dithiothreitol, 10 mM magnesium chloride, 1mM EDTA, 5% glycerol (v/v), and 300 mM ammonium sulfate. The cells were then disrupted in a bead beater (Biospec, Bartlesville, OK, USA) with 0.5 mm glass beads. The lysate was recovered and centrifuged at 100,000 g for 70 min to obtain the microsomal fraction. The microsomes were resuspended with 0.1 M phosphate buffer (pH 7.2), and the protein content was quantified with the Bradford method with bovine

serum albumin to prepare the standard curve (Bradford, 1976). Then, the microsomes were flash-frozen, freeze-dried, and stored at -80ºC.

3.2.7. *In vitro* PDAT assay

PDAT assays were carried out using fluorescent substrates as described in Chapter 2. 1-palmitoyl-2-dodecanoyl-nitrobenzoxadiazole-*sn*-glycero-3-glycerol (NBD-DAG), 1,2-dihexadecanoyl-*sn*-glycero-3-phosphocholine (16:0,16:0-PtdCho), 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (18:1,18:1-PtdCho), and 1,2-dilinoleoyl-*sn*-glycero-3-phosphocholine (18:2,18:2-PtdCho) were used in the reactions. The PtdCho reagents were purchased from Avanti Polar Lipids Inc. (Alabaster, AL, USA). In brief**,** 1.5 nmol of NBD-DAG and 2.5 nmol of PtdCho dissolved in 14 µl of benzene were added to freeze-dried microsomes containing 40 µg of protein. Benzene was dried under an N_2 stream and 100 µl of 50 mM potassium phosphate buffer (pH 7.2) was added to initiate the reaction, which was carried out at 30ºC for 60 min. The reaction was quenched with 100 µl of chloroform/methanol (2:1, v/v). The chloroform phase was spotted on TLC plates and developed in diethyl ether/hexane/methanol/acetic acid (60:40:5:1, by volume). A Typhoon FLA 9500 (GE Healthcare, Mississauga, ON, Canada) was used for visualization of spots. The TLC plates were scanned with excitation and emission wavelengths of 495 nm and 519 nm, respectively, and the PMT voltage was set to 250 V. The fluorescence intensity of the products was quantified using ImageJ 1.52a (Schneider et al., 2012). The quantification of NBD-TAG was obtained from a standard curve generated with the fluorescence of known amounts of NBD-PtdCho. A t-test assuming unequal variance was carried out in Excel to compare the NBD-TAG production of microsomes with wild-type and mutant AtPDAT1.

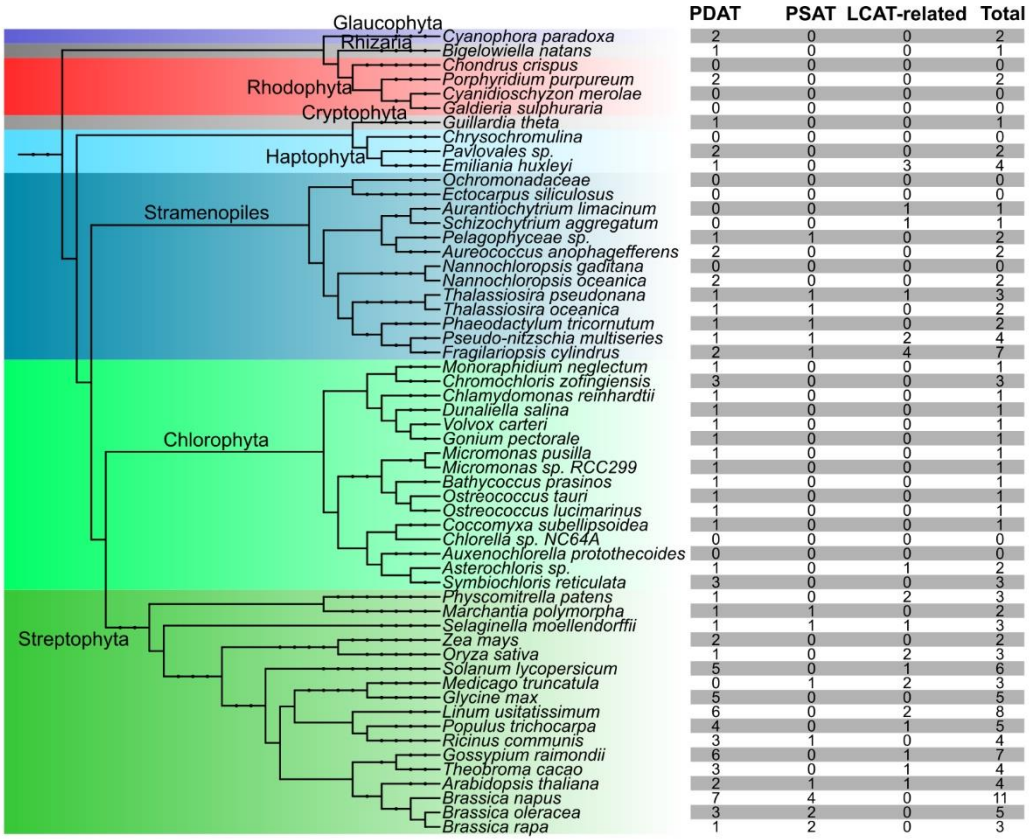
3.3. Results

3.3.1. PDAT enzymes are present in most plant and algal species

PDAT homologs were searched in 39 algal and 17 plant genomes available in public databases (Supplementary table A.1) and used in comprehensive characterization. The algal species include green algae, red algae, Glaucophyta, Rhizaria, Cryptophyta, Haptophyta and Stramenopiles. As for plant species, land plants, eudicots, monocots, the moss *Physcomitrella patens*, the liverwort *M. polymorpha* and the lycophyte *Selaginella moellendorffii* are all included in this study. A total of 59 and 81 putative PDAT homologs were found in algal and plant genomes, respectively. The distribution of PDAT homologs identified in the blast search varied among the genomes and between both plant and algal groups. On average, plants have more homologs per genome than algae (4.8 sequences/plant species versus 1.5 sequence/algal species). In addition, all studied plants have more than one putative homolog (Figure 3.1), which agrees with a previous publication that found at least one copy of *PDAT* is present in different plant genomes (Pan et al., 2015). Conversely, it was not possible to find LCAT-like enzymes in nine algal genomes. *Porphyridium purpureum* was the only red alga containing members of this family. Among green algae, *Chlorella sp.* NC64A and *Auxenochlorella protothecoides* do not have any similar sequences. Together, the higher number of LCAT-related sequences and the presence of homologs in all analyzed plants suggest the LCAT-family enzymes might be more important to plants than to algae.

Considering the conservation of important features in the LCAT family, some hits of the search for PDAT homologs might represent another member of the LCAT family, such as a PSAT, phospholipase A (PLA) or LCAT (Pan et al., 2015; Yoon et al., 2012). To clarify the correct function of the homologs, the sequences were classified by InterPro into different

PANTHER subfamilies (Jones et al., 2014; Mi et al., 2017). Thus, it was possible to identify 38 and 51 putative PDATs for algae and plants, respectively (Figure 3.1). PSAT, PLA and sequences lacking a PANTHER subfamily were also identified in the BLAST search. Interestingly, *PDAT* was not found in eleven algal genomes; most of them belong to the Stramenopiles. *Aurantiochytrium limacinum*, S*chizochytrium aggregatum*, and *Medicago truncatula* do not have a PDAT, but they contain copies of PSAT or LCAT-related enzymes. Most members of the Chlorophyta have only one copy of *PDAT*, while eudicots have as many as seven. Overall, *PDAT* is present in most eukaryotic photosynthetic organisms, but they might have a minor role in the algal evolution or metabolism since several genomes do not contain its homologs. Sequences not classified in the LCAT family were removed from further analyses. In addition, the candidate PDATs that could be a result of annotation errors (e.g. short length, truncated) were also discarded, as listed in Supplementary table A.4. Thus, a total of 37 and 69 algal and plant sequences from LCAT family, respectively, were used in further analyses. The plant sequences were further divided into two groups following the results of a previous study (Pan et al., 2015). Group A contains the basal land plants, monocots and eudicots from groups V and VI and Group B contains the eudicots from group VII.



PDAT	PSAT	LCAT-related	Total
2	0	0	2
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
2	0	0	2
1	0	3	4
0	0	0	0
0	0	0	0
0	0	1	1
0	0	1	1
1	1	0	2
2	0	0	2
0	0	0	0
2	0	0	2
1	1	1	3
1	1	0	2
1	1	0	2
1	1	2	4
2	1	4	7
1	0	0	1
3	0	0	3
1	0	0	1
1	0	0	1
1	0	0	1
1	0	0	1
1	0	0	1
1	0	0	1
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	1	2
3	0	0	3
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
4	0	1	5
3	1	0	4
6	0	1	7
3	0	1	4
2	1	1	4
7	4	0	11
3	2	0	5
1	2	0	3

Figure 3.1. Number of PDAT homologs found in the BLAST search using *Chlamydomonas reinhardtii* and *Arabidopsis thaliana* PDATs as the query. All sequences were further classified based on PANTHER implemented in InterProScan v5.22.61. PANTHER subfamilies PTHR11440:SF4 and PTHR11440:SF54 represent PDATs in algae and plants, respectively. Sequences classified in the PTHR11440:SF7 subfamily represent PSAT. The proteins classified in other LCAT subfamilies or not having a subfamily are shown as LCAT-related. PDAT, phospholipid:diacylglycerol acyltransferase; PSAT, phospholipid:sterol O-acyltransferase; LCAT, lecithin:cholesterol acyltransferase.

3.3.2. Algal LCAT-family enzymes are longer than the plant homologs

After showing the distribution of PDATs and other enzymes from the LCAT family in the algal and plant species, the main properties of these proteins were determined. The properties of LCAT-family enzymes show variation both among enzyme subfamilies and between algae and plants (Table 3.1 and Supplementary table A.5). Overall, algal LCAT-like enzymes are longer than their homologs in plants. During the evolution of chromists, long PSAT and LCAT-related enzymes may have acquired or retained features not present in plant enzymes. On average, algal PSAT and LCAT-related have 893, and 992 amino acids, respectively, while in plants they have 643 and 526 amino acids, respectively. There are 7 enzymes that are longer than 1,000 amino acids (TpsPSATa, PelPSATa, SagLCATa, PmuLCATb, FcyLCATd, EhuLCATa, EhuLCATb), while the longest enzyme among PSAT and LCAT-related in plants is an *M. polymorpha* PSAT (MpoPSATa) with 706 amino acids (Supplementary table A.5). *S. aggregatum* LCAT (SagLCATa), with 1589 amino acids and 176.55 kDa, is the longest enzyme identified in this study.

On average, PDATs from green algae, chromists/red algae, plants – group A, and plants – group B have 768, 641, 670, and 648 amino acids, respectively. In contrary to the PSAT and LCAT-related enzymes, which are found only in chromists, the lengths of chromist PDATs were similar to plants, suggesting that PDATs may have a different evolutionary pattern to other LCAT-like enzymes. In addition, green algal PDATs are longer than land plant PDATs (AtPDAT1 and CrPDAT have 671 and 1041 amino acids, respectively), despite that green algae and plants are evolutionarily closer (de Vries et al., 2016; Leliaert et al., 2012). The longer PDATs in green algae may have unique features in their structure.

In terms of pI, there are some differences among the LCAT-family enzymes in plants and algae (Table 3.1). PSAT in algae is more acidic than that in plants and it is lower than 7.0 for both groups, which seems to be a conserved characteristic in both plants and algae. In contrast, LCAT-related enzymes have a lower pI in plants. They also have a wider range of pIs in both plants and algae (pI of 5.30-8.88 and 5.36-9.77, respectively). Chromists have PDATs with an acidic pI, while, on average, the pI is 7.37 in similar enzymes from green algae.

Table 3.1. Properties of LCAT-family enzymes identified in the BLAST search from plants and algal sequences. Values represent the mean ± standard deviation for each subfamily. Sequences were classified into PDAT, PSAT, PLA or LCAT-related based on PANTHER implemented in InterProScan v5.22.61.

Group	**Length (aa)**	**MW[1] (kDa)**	**pI[2]**	**# TMD[3]**
Algal PSAT[4]	992±229	110.06±23.24	5.31±0.11	0.33±0.52
Plant PSAT[4]	643±23	72.76±2.67	6.04±0.27	0.40±0.52
Plant LCAT-related[5]	427±26	46.88±2.78	6.73±1.38	0.30±0.48
Plant PLA[6]	526±55	59.57±6.50	5.61±0.35	0.00±0.00
Algal LCAT-related[5]	893±334	98.81±36.81	7.41±1.66	0.00±0.00
Red algal/Chromist PDAT[7]	641±102	71.21±10.67	5.93±0.20	0.38±0.52
Green algal PDAT[7]	768±220	82.30±20.26	7.37±1.55	0.78±0.44
Plant PDAT[7] – Group A	670±42	74.44±4.81	7.07±0.97	0.85±0.36
Plant PDAT[7] – Group B	648±83	72.29±9.49	8.17±1.16	0.55±0.52

[1]MW: Molecular weight; [2]pI: Isoelectric point; [3]TMD: Transmembrane domains
[4]PSAT: Phospholipid:sterol O-acyltransferase; [5]LCAT: Lecithin:cholesterol acyltransferase
[6]PLA: Phospholipase A; [7]PDAT: Phospholipid:diacylglycerol acyltransferase

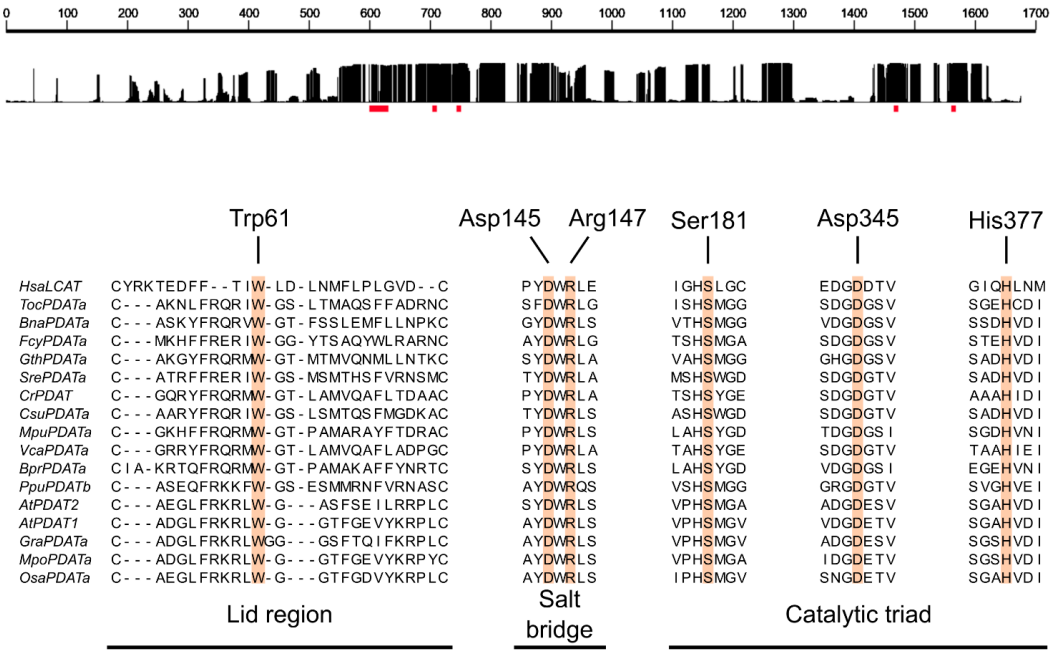
3.3.3. Relevant features of human LCAT are conserved in algal and plant PDATs

To better understand the structure of PDATs, which are still largely uncharacterized, I compared them with the human LCAT (Figure 3.2 and Supplementary figure B.1). The human LCAT is well studied and its crystal structure was already established (Glukhova et al., 2015; Piper et al., 2015). Parts of this enzyme follow an α/β hydrolase fold with the catalytic triad buried in the bottom of the active site. Relevant features of its structure include the "lid" region between two cysteines forming a disulfide bridge, the salt bridge, and the catalytic triad (Peelman et al., 1998; Piper et al., 2015). With different patterns of conservation, these elements are present in both algal and plant PDATs (Figure 3.2A). Specifically, the two cysteines that form the disulfide bridge and the tryptophan (Trp61 in the human LCAT) in the lid region, which are crucial for the LCAT activity likely via mediating substrate binding, are also conserved among PDATs from algae and plants (Peelman et al., 1998). PavPDATa, BraPDATa, and LusPDATd are some enzymes that lack part of the N-terminal region and, consequently, the "lid" region. LCAT enzymes lacking the entire "lid" or solely the residues 56-68 showed no activity (Peelman et al., 1999, 1998). Thus, it is possible these PDATs without the "lid" region identified in this study have no activity, but it is still unclear whether the truncation is caused by genome annotation errors. Of great importance for the LCAT structure stability is the salt bridge formed by aspartic acid and arginine (Asp145 and Arg147, respectively). The amino acids that compose this motif (Tyr144-Asp145-Trp146-Arg147-Leu148) are extremely conserved in both algae and plants PDATs. Similarly, the catalytic triad is very conserved among all the studied sequences.

The predicted structures of CrPDAT and AtPDAT1 were also analyzed and compared with the human LCAT (Figure 3.2B). The superposition of their structures shows that relevant motifs of HsaLCAT ("lid" region, salt bridge and the catalytic triad) and their three-dimensional

orientation in the structures are conserved in the PDATs. The “lid” region is at the surface of the protein as it is a putative contact region for protein activation. In addition, the catalytic triad is at the bottom of the catalytic pocket. The results suggest PDAT may have a similar structure and a likely analog mechanism of reaction to the human LCAT. When compared to HsaLCAT, there are some regions of algal PDATs that do not have a match, suggesting algal enzymes might have acquired new features not present in the LCAT.

A



B

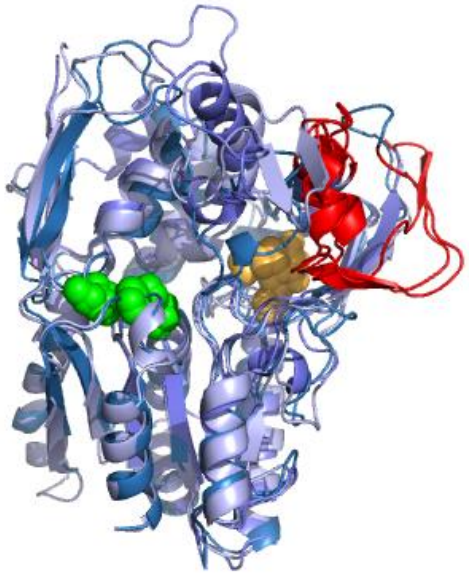


Figure 3.2. Legend on the next page.

Figure 3.2. Representative elements of PDAT structure. A) Sequence alignment of selected PDATs with human LCAT (HsaLCAT). The alignment conservation is illustrated on the top. The "lid" region, salt bridge and catalytic triad are relevant domains of HsaLCAT structure and are shown in the figure. The alignment demonstrates that these regions are also conserved in the algal and plant PDATs. The locations of these important domains are shown as red lines below the alignment conservation. B) Superposition of the structures of HsaLCAT, CrPDAT and AtPDAT1 shown in different tones of blue. The "lid" region, salt bridge, and catalytic triad are shown in red, orange and green, respectively. Figure 3.2B was generated by Dr. Kristian Caldo.

3.3.4. TMDs are a characteristic of plant and green algal PDATs

Considering the putative mechanism of action of PDATs, they might be localized in the ER membranes with an orientation towards the lumen (Pan et al., 2015; Stahl et al., 2004). However, PDATs seem to be very important for plastid remodeling in green algae and a chloroplast transit motif in the N-terminus of CrPDAT was found to be responsible to its chloroplast localization (Yoon et al., 2012). Compared to chromists, the presence of TMDs seems to be especially important among PDATs from green algae and plants from group A since most sequences of these groups contain at least one predicted TMD (Table 3.1 and Supplementary table A.5). The sequences among plant (group A) PDATs that do not have a TMD are BnpPDATa, BnpPDATc, BnpPDATd, BraPDATa, and TcaPDATb. In addition, AspPDATa, CsuPDATa, and GpePDATa are the sequences without a predicted TMD among green algae. Interestingly, the lack of a predicted TMD matched the absence of the initial moiety of the N-terminal region (Figure 3.3 and Supplementary figure B.2). Thus, the presence of a TMD may not be essential for PDAT activity and may be related to a specific role the enzyme has in each organism.

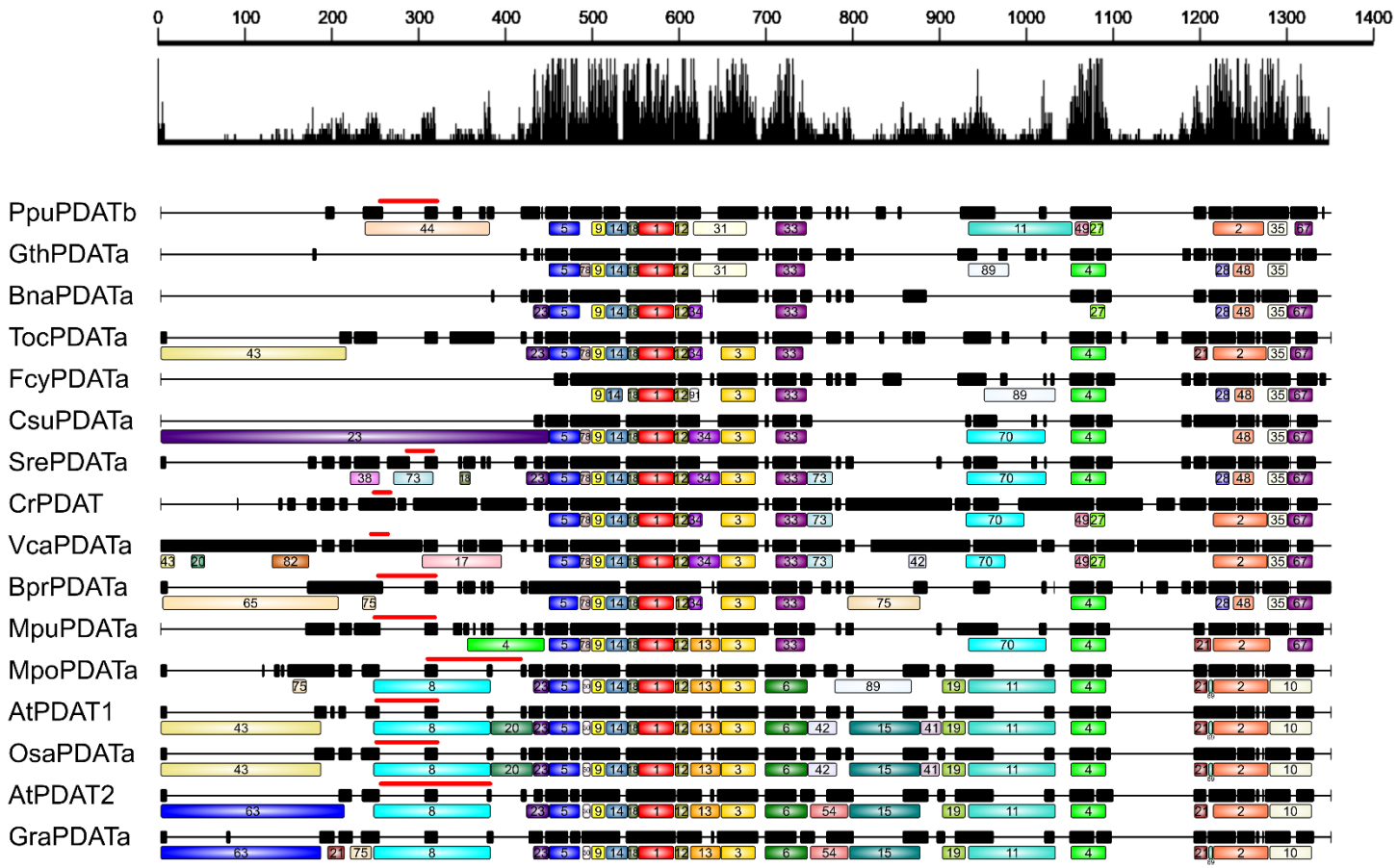


Figure 3.3. Predicted transmembrane domains (TMDs) and conserved motifs on selected algal and plant PDATs. Gaps and the aligned sequences are shown in black lines and black bars, respectively. The position of TMDs is shown as a thick red line above the aligned sequences. The position of motifs identified in MEME v4.12.0 is shown as bars with different colors below the aligned sequences. The motif number is indicated inside each bar.

Oppositely to PDATs in green algae and plants (group A), TMDs are not predicted in most of the other studied LCAT subfamilies. While no TMD was predicted among all algal LCAT-related sequences, only two enzymes have a single TMD in algal PSATs (TpsPSATa and PelPSATa). Similarly, only 3 and 4 sequences contain a TMD among LCAT-related and PSATs in plants, respectively, while no TMD was found in PLAs (Supplementary table A.5). Therefore, the presence of TMD in the N-terminal region of LCAT-family enzymes seems to be more important among PDATs but may not be essential to the catalysis. In addition, it is always present as a single helix. The exception is the AspLCATa enzyme, which has 5 predicted TMDs. Interestingly, four of them are located in a portion of the C-terminal moiety identified only in this sequence (Supplementary figure B.2).

3.3.5. Conserved motifs show that PDATs can be divided into three regions of different conservation

A deeper structural analysis of the conservation of members of LCAT family was obtained using MEME v4.12.0 analysis. Adopting an E-value cut-off of $1.0x10^{-5}$, 94 conserved motifs were obtained for the 106 submitted sequences (Supplementary figure B.3). The only motif identified in all sequences is Motif 12 (TNGGKKVVVVPHSMG), which contains the serine of the catalytic triad (Ser254, Asp573, and His 626 in AtPDAT1; Figure 3.3 and Supplementary figure B.4). In special, the last five amino acids ([P/S/T/C/A/G]HS[M/L/Y/W]G) are very conserved in this motif and, thus, could be considered a signature of LCAT family in algae and plants. The motifs containing the other two residues in the catalytic triad are not conserved among all studied groups. Motif 2 (GVYSVDGDETVPVLSAGFMCAKGWRGKTRFNPSGIRTYIREYNHSPPANLLEGRGTQS)

, present in plant and some green algal PDATs (including CrPDAT), contains the asparagine of the catalytic triad (Asp573 in AtPDAT1). However, this site in plant LCAT and PLA is represented by Motif 28 (YGDGDGTVPLVSLGAL), which is the same for other algal PDAT, PSAT and LCAT-related sequences. This site is also covered by Motif 7 (TGDETVPYHSLSWCKNWLGPKVNITMAPQPEHDGSDVQIELNVEHEHGSDIIANMTKA PRVKYITFYEDSESIPGKRTAVWELDKANHRNIVRSPVLMRELWLQMWHDIQPGAKSK FVTKAKRGPLRDADCYWDYGKACCAWQEYCEYRYSFGDVHLGQSCRLRNT) in plant PSAT. The same motif, which is unique to plant PSATs, also contains the histidine of the catalytic triad (His626 in AtPDAT1). Motif 10 (AHVDIMGNFALIEDIMRVAAGATGEDLGGDRVYSDIFKWSERI) and Motif 35 (SADHVDILGNVEVJEDVJKIATG) contains the catalytic histidine in plant and algal PDATs, respectively. Among algal PSAT, algal LCAT-related and plant LCAT-related, the histidine-containing motif is the Motif 37 (EQEVKVIEJDGVSHREILRDERALAAILREV). Also, Motif 64 (NAEARVGIPGDHRGILKDQHLFRILKHWLKVGDPDPFYDPLNDYVIIP) has the catalytic-triad histidine in the plant PLA group. These results show that during evolution, enzymes from the LCAT family acquired new features regarding the motifs containing the catalytic triad residues. Despite the high conservation of these individual amino acids, the surrounding residues suffered mutations which may lead to the formation of different subfamilies of enzymes. In addition, these results also suggest a divergent evolution of similar subfamilies of enzymes between plants and algae.

There are six motifs that are in very conserved regions in both algae and plant PDATs (Motifs 1, 3, 5, 9, 14, 18). The salt bridge motif present immediately upstream of the catalytic triad serine motif is represented by Motif 1

(RIGYEEKNMYMAAYDWRLSFQNTEVRDQTLSRLKSNIELMV). Motif 5 (KHPVVFVPGIVTGGLELWEGHQCAEGLFRKRLWG) contains the first half of the "lid" region, while the other half is divergent between plants and algae. In this case, it is represented in distinct motifs (Motifs 30 – TFGEVYKRP and 78 – TLAMAQAFFADR in plants and algae, respectively). The region formed by the Motifs 5, 9, 14, 18, 1, and 12 (in this order between the "lid" domain and the serine in the catalytic triad) can be considered the most conserved region between the PDAT groups in plants and algae. Thus, mutations in this portion of the N-terminal region of PDATs will likely have a greater impact on the activity of these enzymes.

Despite these results, several differences in other regions of PDATs support the occurrence of divergent evolution between algal and plant PDATs. The predicted TMDs in plant PDATs are present in Motif 8, whereas the TMDs in algae have not been assigned a conserved motif. This can be expected because the consensus of the multiple sequence alignment shows the initial portion of the N-terminal as a region with high variability (Supplementary figure B.2A). The higher divergence in the initial portion of the PDAT sequences (and consequently lack of a conserved motif) is also expected because the algal species in this study are more evolutionarily distant than the plant species.

Considering the conservation of motifs, PDATs could be divided into three parts. The first is the variable N-terminal region upstream to the "lid" domain, which may contain a TMD. The second part is the region between the "lid" domain and the catalytic serine, which is the most conserved between algal and plant PDATs. The last part comprises the region downstream the catalytic serine, which may be very divergent between algae and plants. Out of 32 motifs that are present solely in algae or plants, 24 are located downstream from the catalytic serine. Both plant PDAT groups contain 13 exclusive motifs (Motifs 4, 6, 8, 10,11,13,15, 19, 21, 23, 30, 69,

and 75). In addition, there are some motifs present solely in group A (Motifs 20, 41, 42, 43, and 89) or group B (Motifs 54 and 63). On the other hand, there are 12 motifs present uniquely in algae (Motifs 27, 28, 33, 34, 35, 48, 49, 67, 70, 73, 78, and 91). As a comparison, the region downstream the catalytic serine is 73% covered by unique motifs in the AtPDAT1, while the CrPDAT is 31% covered by unique motifs in the same region. This difference is explained (i) by the evolutionary distance of algal species used in the study and (ii) by the distinct evolutionary patterns that plant and algal PDATs were subjected during evolution. It shows a higher variability between algal and plant PDATs, confirms the region downstream the catalytic serine is more divergent and suggests they have a different evolutionary pattern between the two groups.

3.3.6. The gene structure is highly variable among algal LCAT-family genes

In terms of the gene structure in the LCAT family, there is high conservation among plants and high variability among algae for all LCAT subfamilies (Supplementary figure B.5). As shown previously, the plant *PDATs* have a very conserved gene structure with 5 introns following the 2-0-2-0-2 intron phase pattern (Pan et al., 2015). Similarly, plant *PSAT* genes contain 14 introns with a conserved intron phase pattern (0-1-2-1-0-1-2-0-2-0-0-1-1-2) and plant *LCAT*-related genes contain one intron classified as phase 2. Oppositely to plants, algal LCAT subfamily genes in this study do not have an intron pattern, although some sequences contain several introns. *AliLCATa*, *EhuLCATa*, *BprPDATa*, *MpuPDATa*, *MspPDATa*, *TocPDATa*, *PpuPDATa*, *PpuPDATb* do not contains introns. The other algal sequences may contain up to 19 introns, but the intron phase is unconserved among algal homologs.

It is common for unicellular eukaryotes to have fewer introns due to a major loss from an intron-rich ancestral (Rogozin et al., 2012). However, some species are also subject to intron gains, which seems to be related to different factors, such as generation time, expression level, gene function, the need for alternative splicing, the intron position, the population size, and how it affects the fitness of the organism (Jeffares et al., 2006; Rogozin et al., 2012). Thus, the data shown for intron frequency in algal and plant LCAT-related genes can be explained by the current theories of intron gain and loss. The lack of a pattern for introns among algae reflects the evolutionary distance of algal species when compared to the plant species. In addition, it is necessary to observe the intron pattern for each algal species. For instance, *P. purpureum* and both *Micromonas* species (isolates RCC299 and CCMP1545) have a small intron-poor genome (less than 1 intron per gene; *P. purpureum* has only 235 introns for 8,355 genes; Bhattacharya et al., 2013; Worden et al., 2009). On the other hand, the *C. reinhardtii* genome is similar to higher eukaryotes, containing 8.3 introns per gene (Merchant et al., 2007).

3.3.7. Green algal PDAT evolution is different from other major algal groups

To understand the relationship among LCAT-family enzymes, phylogenetic analyses were performed. Initially, the phylogeny for the algal sequences was reconstructed to better figure out the evolution of this enzyme family solely in algae (vide Pan et al., 2015, for a larger and complete description of the phylogeny of plant PDATs). Avoiding the use of the larger plant dataset in this initial stage may minimize the occurrence of phylogenetic incongruences as the plant dataset may be overrepresented (Menardo et al., 2018). In addition, more sites can be included in the phylogenetic analysis. Thus, an ML phylogenetic tree was reconstructed with Phyml v3.0 and a BI tree was also constructed to confirm the results of the ML analysis.

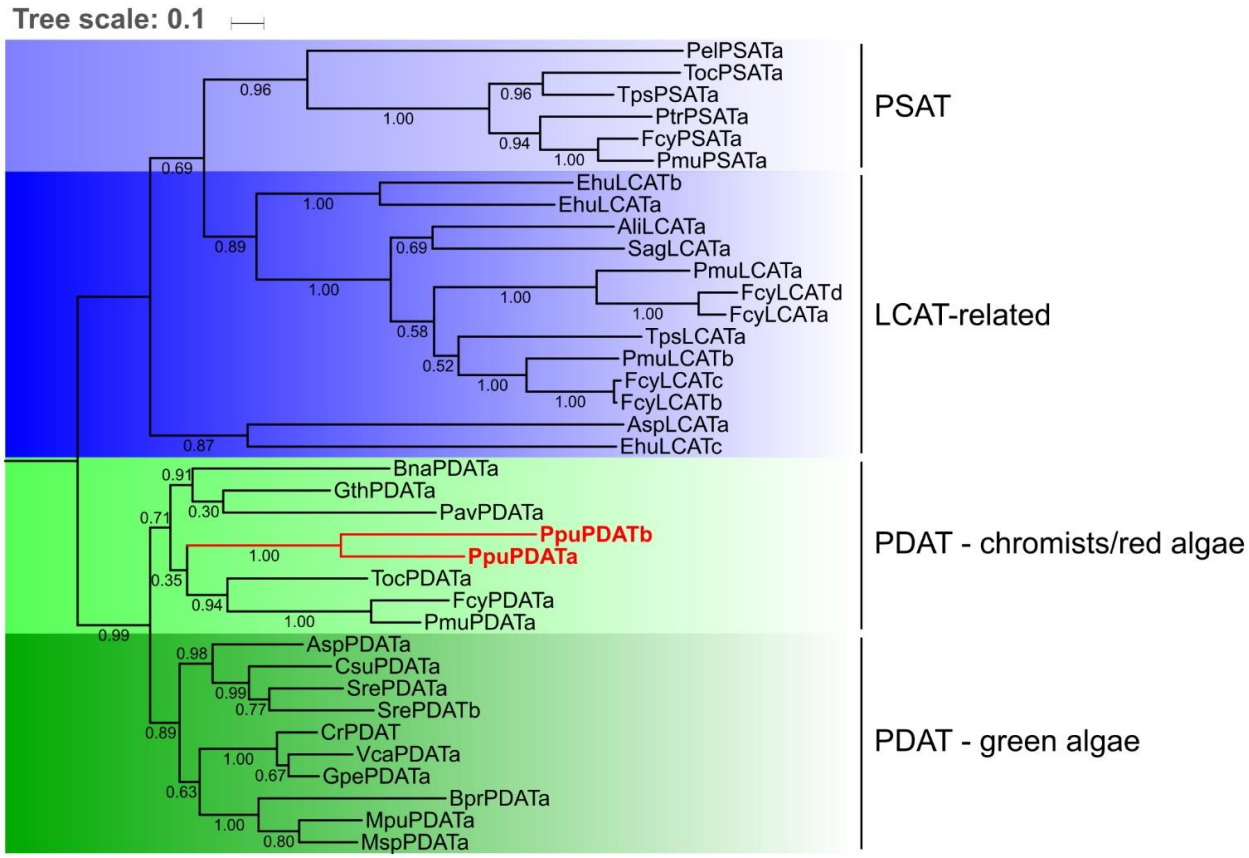


Figure 3.4. Phylogenetic analysis of algal LCAT-family enzymes. *Porphyridium purpureum* PDATs (PpuPDATa and PpuPDATb; shown in red) forms a monophyletic group with chromist PDATs in the maximum likelihood (ML) tree. The ML phylogenetic tree was reconstructed with Phyml v3.0. Branch-supporting values are shown on the branches. Sequences were classified into PDAT, PSAT, PLA or LCAT-related based on PANTHER implemented in InterProScan v5.22.61. LCAT, lecithin-cholesterol acyltransferase; PDAT, phospholipid:diacylglycerol acyltransferase; PSAT, phospholipid:sterol O-acyltransferase.

As expected, the major LCAT subfamilies form monophyletic groups (Figure 3.4 and Supplementary figure B.6). PSAT and LCAT-related enzymes share a common ancestor, while PDATs form a separate clade. In addition, the evolution of the sequences within each clade is

consistent with the species evolution (Figure 3.1), which discards the possibility of horizontal gene transfer. In terms of gene duplication, the algal phylogenetic tree shows that these events are rare in the evolutionary history of this protein family in algae. In most cases, the duplication is recent (LCAT-related of *Emiliania huxleyi* and *Fragilariopsis cylindrus* and PDATs from *S. reticulata* and *P. purpureum*), but there is evidence of ancestral duplication in the diatoms from the *Bacillariaceae* family. The LCAT-related enzymes from *F. cylindrus* and *Pseudo-nitzschia multiseries* are distributed into two sister clades, which suggest a duplication in their ancestor. Unfortunately, it is not possible to conclude about the function of LCAT-related enzymes with the information shown in Figure 3.4.

The PDATs from green algae and chromists are also classified into distinct monophyletic groups (Figure 3.4). However, the position of the red algal sequences was incongruent between the ML and BI trees. In the ML phylogenetic tree, the *P. purpureum* PDATs constitute monophyly with the Bacillariophyta PDATs, while they form the basal clade of algal PDATs in the BI tree. In this case, the green algal PDATs form a monophyletic group and the chromist PDATs compose a paraphyletic group. To clarify the position of red algal PDATs, a topological hypothesis test with the stepping-stone sampling method implemented in MrBayes v3.2.6 was carried out, which is more accurate than the harmonic mean (Xie et al., 2011). In the first hypothesis test, the null hypothesis (H0) does not include the *P. purpureum* sequences with the chromist sequences. The alternative hypothesis (H1) defines the constraint whereby the red algal and chromist PDATs form a monophyletic group in all sampled trees. In the second test, H0 does not allow red algal and Bacillariophyta PDATs to form a monophyletic group, while H1 constrains PpuPDATa and PpuPDATb in the same clade as TocPDATa, FcyPDATa and PmuPDATa. The third test compares whether red algal sequences are a basal group to chromists

(H0) or they form a monophyletic group with GthPDATa, PavPDATa and BnaPDATa. The difference between the log of marginal likelihoods (ℓ) of H0 and H1 multiplied by two shows how strong the evidence is in favor of the best ℓ (Kass and Raftery, 1995). The result for the first topological hypothesis test is 31.44 and for the second is 21.02 in favor of the alternative hypothesis (Table 3.2). Because of a number higher than 10 is considered very strong evidence, the red algal PDATs form a monophyletic group with the chromists. When comparing the alternative hypotheses of the second and third tests, there is strong evidence (7.30) that PpuPDATa and PpuPDATb will clade preferentially with the Bacillariophyta PDATs. Although still unclear, there are plenty of evidence grouping the Chlorophyta and Rhodophyta species in the same clade (Clerck et al., 2012). Thus, these findings show that PDAT evolution did not follow the evolution of major algal groups.

Table 3.2. Topological hypothesis test shows that *Porphyridium purpureum* PDATs are closer to chromists. The red algal sequences forming a monophyletic group with (1) all the chromists, with (2) the Bacillariophyta only or with (3) GthPDATa, BnaPDATa and PavPDATa representing the constraint of the alternative hypothesis (H1), respectively. The null hypothesis does not allow the red algal sequences to cluster with any of those groups. The test was carried out with the stepping-stone sampling method implemented in MrBayes v3.2.6.

Topological constraint	**1-Log-likelihood (ℓ) All chromists**	**2- Log-likelihood (ℓ) All Bacillariophyta**	**3- Log-likelihood (ℓ) Other chromists**
Red algal PDATs as a basal group (H0)	-34,894.80	-34,894.42	-34,894.39
Red algal PDATs with chromists (H1)	-34,879.08	-34,883.91	-34,887.56
2*(H1-H0)	31.44	21.02	13.66
Support	**Very strong***	**Very strong***	**Very strong***

*If twice the log difference of marginal likelihoods (ℓ) is higher than 10, the support in favor of the hypothesis with the lower ℓ can be considered very strong (Kass and Raftery 1995).

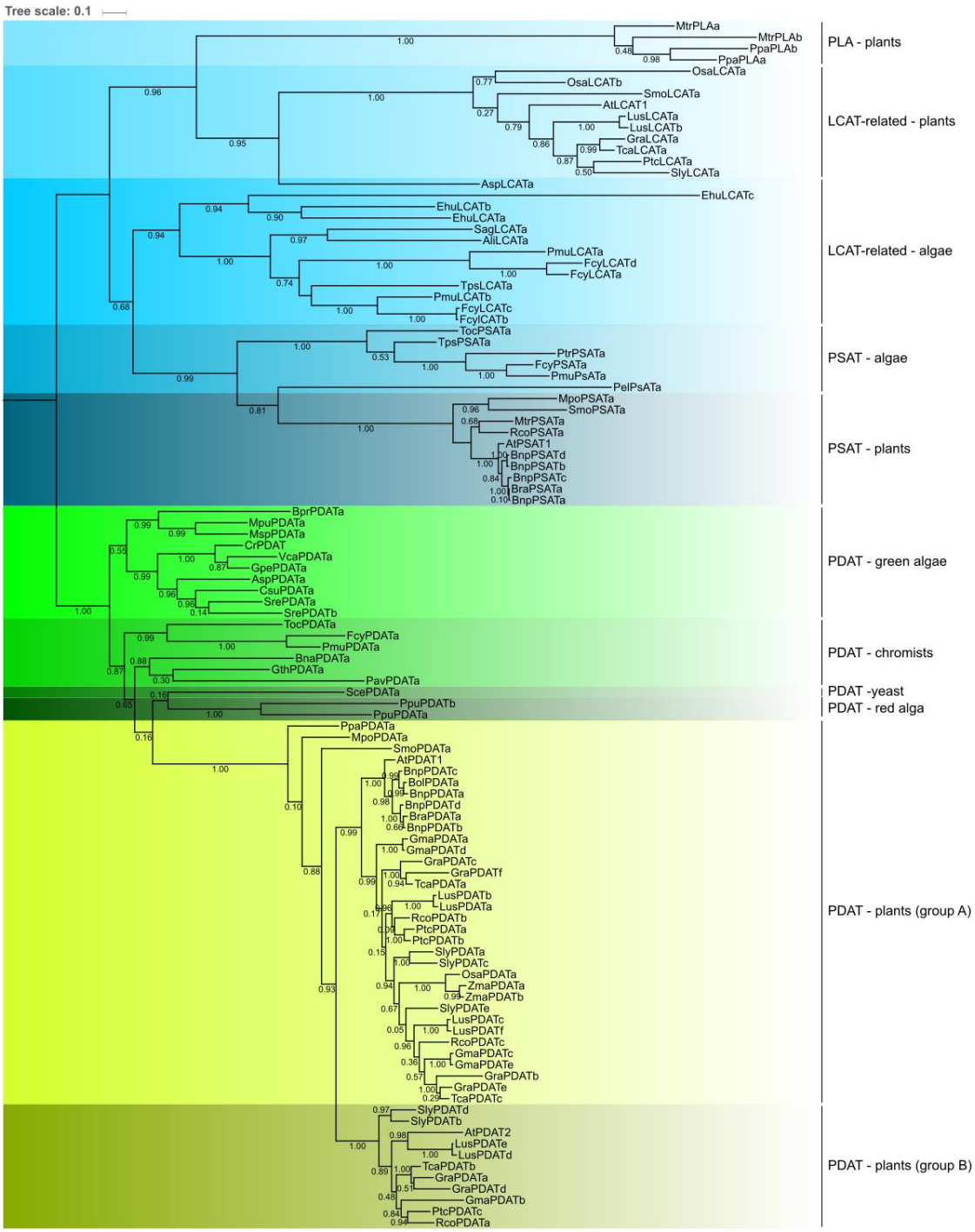


Figure 3.5. Phylogenetic comparison of algal and plant LCAT-family enzymes. Plant PDATs are divided into two groups. The maximum likelihood (ML) phylogenetic tree was reconstructed with Phyml v3.0. Branch-supporting values are shown on the branches. Sequences were classified into PDAT, PSAT, PLA or LCAT-related based on PANTHER implemented in

InterProScan v5.22.61. LCAT, lecithin-cholesterol acyltransferase; PDAT, phospholipid:diacylglycerol acyltransferase; PLA, phospholipase; PSAT, phospholipid:sterol O-acyltransferase.

3.3.8. Phylogenetic analysis reveals a possible functional divergence between algal and plant PDATs

Following the analysis of the phylogenetic relationship among algal LCAT-like enzymes, the plant sequences were included to compare their evolution with algal homologs. The ML phylogenetic tree is shown in Figure 3.5 and the BI tree is in Supplementary figure B.7. Both share a very similar topology with good branch-supporting values. Similar to the algal phylogenetic tree, the PDATs can form a monophyletic group and can be placed at the base of the tree (Figure 3.5 and Supplementary figure B.7). However, the ML tree shows the green algal sequences at the base of PDAT sequences, whereas it is not possible to conclude which group is basal due to the presence of polytomies in the BI tree. The PSAT enzymes from algae and plants are forming a monophyletic group, but the *Pelagophyceae sp.* PSAT is closer to the plant PSATs. As shown previously, there are 12 motifs found in the MEME analysis that are exclusive of algal PSATs, but none of them is present in the PelPSATa. The other motifs present both in algal and plant PSATs might explain why this sequence is closer to the plant group.

The LCAT-related enzymes in plants and algae are still largely uncharacterized, and their function is unknown. The only LCAT-related enzyme found in green algae (AspLCATa) is closer to the plant LCAT-related. The plant PLA enzymes form a monophyletic group with plant LCAT-related, but algal LCAT-related enzymes are closer to the PSAT clade. Considering these results, it is possible LCAT-related enzymes do not have similar functions in algae and plants

because they are evolutionarily divergent. In addition, there are conserved motifs (27, 28, 32, and 37) among algal PSAT, algal LCAT-related and plant LCAT-related, which hinders any clear conclusions about the functions of these enzymes.

PDATs form a monophyletic group, but it is not possible to conclude which subgroup is basal to all PDATs because of the topological differences encountered in the ML and BI trees. The plant PDATs can be grouped into a single clade, which can be subdivided into several clades as shown in a previous study (Pan et al., 2015). They showed it is possible to separate the basal land plants, the monocots, and eudicots into distinct clades. In addition, the eudicots can be separated into three subgroups (groups V, VI, and VII). They also concluded that groups V and VI were subjected to a stronger purifying selection than group VII. Considering the findings of the previous study and because I used fewer plant species in this study, I decided to split the PDATs into two groups. Group A contains the basal land plant PDATs (PpaPDATa, MpoPDATa and SmoPDATa), the monocot PDATs and the equivalent eudicots from group V and VI of the previous study. Group B contains solely the eudicots equivalent to group VII because this clade is basal to other eudicot and monocot PDATs, they have the weakest selection pressure, and they have unique conserved motifs as shown in the MEME analysis.

The phylogenetic comparison between algal and plant PDATs shows a divergence in the plant and green algal PDATs evolution because it is not similar to the species evolution. Modern green algae and plants share a common ancestor, which is an ancient green algal organism (Leliaert et al., 2012). Thus, one would expect a closer evolution of their genes, including the PDATs, which would resemble the species evolution. However, the green algal and plant PDATs do not show this pattern in the phylogenetic analysis. In lieu of Chlorophyta PDATs, the chromist sequences are closer to the plant homologs. On the previous section, the

characterization of the PDAT subfamily and the analysis of conserved motifs showed that both algal and plant PDATs accumulated mutations that led to different characteristics (e.g. length) and to divergent regions/motifs in the sequences. Together with the phylogenetic analysis, these results suggest the presence of functional divergence in the evolutionary history of PDATs from Viridiplantae. Does PDAT have the same relevance for the metabolism of plants and algae? Do the new features acquired by both groups of PDATs affect the structure and activity of these enzymes? To answer these questions, the following sections will focus solely on the evolutionary relationship between green algal and plant PDATs. Therefore, the datasets will include only PDATs from these groups.

3.3.9. Did PDATs evolve to be more important to plants than to green algae?

To investigate the hypothesis of functional divergence in the evolution of green algae and plant PDATs, a selection pressure analysis using Clade model C (alternative model) and M2a_rel (null model) was carried out. The phylogenetic trees showing the groups used in both analyses is shown in Supplementary figure B.8. The results show that the Clade model C fits the data better than the null model. In both cases there was a high statistical significance (Table 3.3A, p-value: 3.39×10^{-26}; Table 3.3B, p-value: 2.18×10^{-6}). The comparison of green algal PDATs with sequences from the plant groups A and B reveals that 58.7% and 53.5% of the analyzed sites, respectively, evolved under different selection pressures, which is a strong sign of functional divergence. In addition, some of these sites may be under positive selection, which could explain what has driven functional divergence. Both comparisons revealed that green algal PDATs have a higher ω ratio for site class 2. Thus, the evolution of green algal PDATs was more relaxed when compared to plant sequences, which indicates that green algal genes were allowed to

accumulate a higher number of mutations. In addition, the ω ratios for the green algal clade are 3.2x and 1.8x higher than the ω ratios for plant groups A and B, respectively. These results show that the degree of divergence of green algal PDATs is higher when compared with the sequences from group A. The stronger purifying selection acting on the plant PDATs (especially from group A) suggests higher importance of these enzymes for the plants. Overall, the results confirm that green algal and plant PDATs evolved under divergent selective pressures.

3.3.10. Positive selection supports the presence of functional divergence in the evolution of green algal and plant PDATs

Considering the high content of amino acids under divergent evolution on the PDAT sequences from green algae and plants, it might be possible some of them evolved under positive selection. To test this hypothesis, I used a branch-site model analysis, which estimates sites under positive selection in the foreground clades of a phylogenetic tree (Yang and Nielsen, 2002; Zhang et al., 2005). The branch-site model A analysis also estimates the posterior probability of sites to be under positive selection using the BEB method (Yang et al., 2005).

All the analyses show the alternative model with a significantly higher ℓ, which shows strong statistical support to accept the branch-site model A and reject the null model (Table 3.4). Considering the proportion of sites classified in the site class 2 and the $\omega > 1$ in that site class, it is possible to conclude that positive selection was present in the evolution of PDATs in plants and green algae. Although the majority of sites are classified under purifying selection, the number of sites under positive selection is also relevant, especially among plant PDATs. When they are in the foreground group, site class 2 comprises 27.1% and 22.2% of the analyzed sites for plant group A and group B, respectively. Green algal PDATs as the foreground clade have a

lower content of amino acids under positive selection; 17.4% and 11.6%, when compared to plant PDATs from group A and group B, respectively. These results corroborate the hypothesis of functional divergence among the green algal and plant PDATs. Such a high number of sites under positive selection may have introduced new elements to the PDAT structure in these organisms. They also support the conclusion that PDATs evolved to be more important for the plant metabolism than to the green algal because positive selection is advantageous to the organism.

Table 3.3. Parameter estimates of Clade model C analyses to understand the selection pressures that acted on the evolution of green algal and plant PDAT sequences. A) Comparison between green algal and plant (group A) PDAT sequences. B) Comparison between green algal and plant (group B) PDAT sequences. Clade model C represents the alternative model (H1). In this model, site classes 0 and 1 represent the amino acids under purifying selection and neutral evolution, respectively. Site class 2 represents the analyzed residues with a divergent evolution. M2a_rel represents the null model (H0). In this model, site classes 0 and 1 are the same as Clade model C. In site class 2, ω ratio is the same for both analyzed partitions.

		Variable	Site class			ℓ[1]	LRT[2]	np[3]	p-value
			0	1	2				
A	H0 – M2a_rel	Proportion[4]	0.43004	0.01365	0.55630	-31,023.68		90	
		ω	0.01905	1.00000	0.13106				
	H1 – Clade model C	Proportion[4]	0.38857	0.02420	0.58723	-30,967.63	112.10	91	3.39×10^{-26}
		Plant branch ω	0.02736	1.00000	0.10890				
		Algal branch ω	0.02736	1.00000	0.35169				
B	H0 – M2a_rel	Proportion[4]	0.40956	0.03699	0.55345	-21,708.77		44	
		ω	0.02856	1.00000	0.18592				

Table 3.3. Continued

	H1 – Clade model C	Proportion[4]	0.41913	0.04626	0.53461	-21,697.55	22.43	45	2.18×10^{-6}
		Plant branch ω	0.03807	1.00000	0.15379				
		Algal branch ω	0.03807	1.00000	0.27179				

[1]ℓ: Log-likelihood value of the analysis
[2]LRT: Likelihood ratio test to compare the models
[3]np: Number of parameters in each analysis
[4]Proportion: Fraction of analyzed sites under each site class

Table 3.4. Parameter estimates of branch-site models to verify the presence of positive selection in the evolution of green algal and plant PDATs. A) Comparison between green algal and plant (group A) PDAT sequences. B) Comparison between green algal and plant (group B) PDAT sequences. Branch-site model A represents the alternative model (H1). In this model, site classes 0 and 1 represent the amino acids under purifying selection and neutral evolution, respectively. Site classes 2a and 2b represent the amino acids that evolved under positive selection in the foreground, but in the background, they are under purifying selection and neutral evolution, respectively. The branch-site model A1 represents the null model (H0). In this model, site classes 0 and 1 are the same as branch-site model A. Site class 2 has $\omega_2 = 1$.

	Foreground group	Model	Variable	0	1	2a	2b	ℓ	LRT	np	p-value
A	Green algal PDAT	H0 – Model A1	Proportion (p)	0.84280	0.03680	0.11536	0.00504	-36,027.91		91	
			Background (ω)	0.08087	1.00000	0.08087	1.00000				
			Foreground (ω)	0.08087	1.00000	1.00000	1.00000				
		H1 – Model A	Proportion (p)	0.83742	0.03660	0.12070	0.05280	-36,017.77	20.29	92	6.65×10^{-6}
			Background (ω)	0.08160	1.00000	0.08160	1.00000				
			Foreground (ω)	0.08160	1.00000	999.00000	999.00000				
	Plant PDAT (group A)	H0 – Model A1	Proportion (p)	0.69935	0.03022	0.25923	0.01120	-36,015.24		91	
			Background (ω)	0.08047	1.00000	0.08047	1.00000				
			Foreground (ω)	0.08047	1.00000	1.00000	1.00000				
		H1 – Model A	Proportion (p)	0.69808	0.03044	0.26013	0.01134	-36,007.71	15.06	92	1.04×10^{-4}
			Background (ω)	0.08170	1.00000	0.08170	1.00000				
			Foreground (ω)	0.08170	1.00000	29.60476	29.60476				

Table 3.4. Continued

B	Green algal PDAT	H0 – Model A1	Proportion (p)	0.78607	0.11037	0.09083	0.01275	-22,502.32		45	
			Background (ω)	0.11318	1.00000	0.11318	1.00000				
			Foreground (ω)	0.11318	1.00000	1.00000	1.00000				
		H1 – Model A	Proportion (p)	0.77571	0.10793	0.10215	0.01421	-22,494.28	16.08	46	6.07×10^{-5}
			Background (ω)	0.11366	1.00000	0.11366	1.00000				
			Foreground (ω)	0.11366	1.00000	465.60906	465.60906				
	Plant PDAT (group B)	H0 – Model A1	Proportion (p)	0.68872	0.09530	0.18972	0.02625	-22,493.88		45	
			Background (ω)	0.11177	1.00000	0.11177	1.00000				
			Foreground (ω)	0.11177	1.00000	1.00000	1.00000				
		H1 – Model A	Proportion (p)	0.68339	0.09500	0.19457	0.02705	-22,485.32	17.13	46	3.49×10^{-5}
			Background (ω)	0.11397	1.00000	0.11397	1.00000				
			Foreground (ω)	0.11397	1.00000	36.08285	36.08285				

[1]ℓ: Log-likelihood value of the analysis; [2]LRT: Likelihood ratio test to compare the models; [3]np: Number of parameters in each analysis; [4]Proportion: Fraction of analyzed sites under each site class

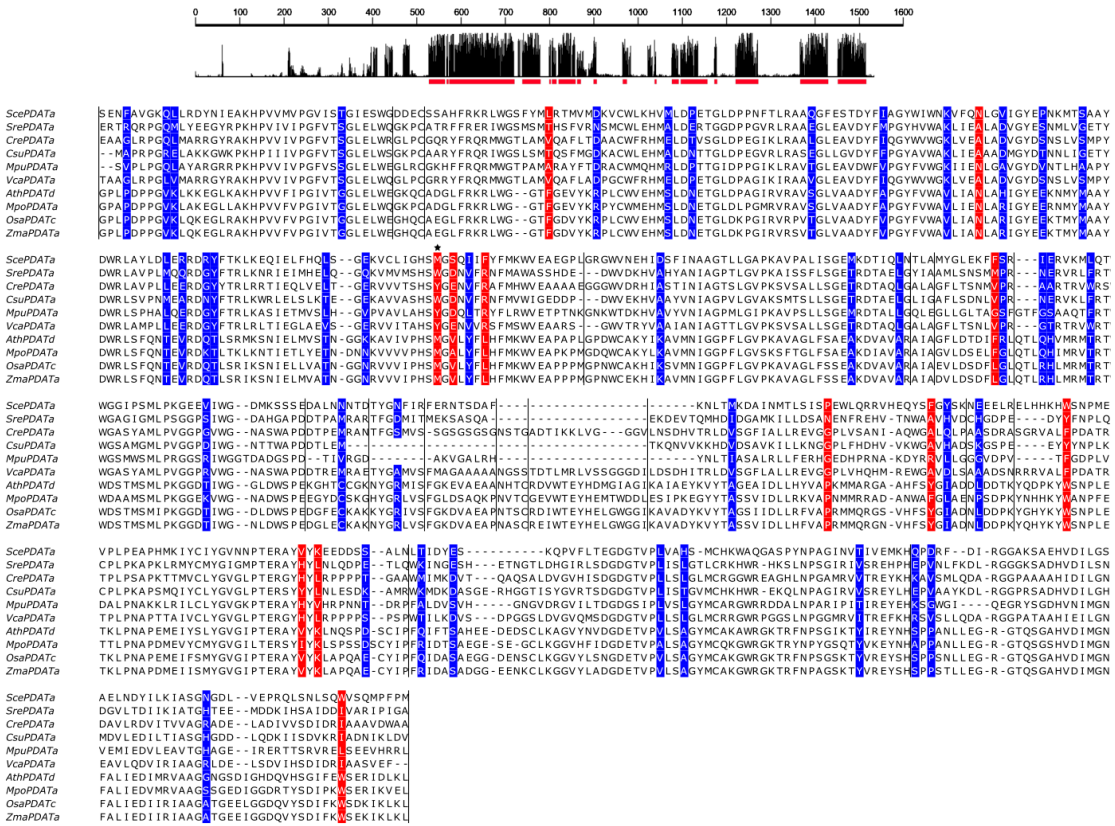


Figure 3.6. Alignment of selected green algal and plant (group A) PDATs which shows the sites under positive selection identified by the Bayes empirical Bayes method. The alignment consensus is shown on the top and the red bars below it represents the regions visualized in the figure. These regions are separated by black vertical lines in the alignment. Sites highlighted in red are under positive selection in green algal PDATs. Sites highlighted in blue are under positive selection in plant PDATs from group A. The amino acid position further studied in AtPDAT1 is marked with a star (★).

The BEB analysis results not only support the higher importance of PDATs for plants but also reveal that plant PDATs from group A might be more important than the enzymes from group B. The accumulation of more sites under positive selection after the divergence from the common ancestor of Chlorophyta and Streptophyta suggests a higher relevance for the plant enzymes. Plant (group A) PDATs have 42 sites under positive selection with a posterior probability higher than 0.95 when compared to green algae (Figure 3.6 and Supplementary figure B.10). They are well distributed in the sequences since 19 sites are upstream to the catalytic serine and 23 are downstream. These substitutions helped define the catalytic activity in plant PDATs from group A and differentiate them from other PDATs. They are especially important considering the higher conservation for the region upstream to the catalytic serine. Contrarily, from the 12 sites under positive selection identified in the algal group, only 3 are upstream to the catalytic serine motif (Figure 3.6 and Supplementary figure B.10). Plant PDATs from group B have only two sites under positive selection, while the green algae have one, according to the BEB analysis results (Supplementary figure B.11).

The analysis of the amino acids under positive selection reveals a large variability in terms of the properties of the residues that suffered a substitution (Supplementary table A.6). Interestingly, there are five sites under positive selection in green algae that are an aromatic amino acid in plants. Four of them represent a different type of side-chain in green algal PDATs. Among plant PDATs from group A, there are only two aromatic amino acids under positive selection, which are aliphatic amino acids in green algae. Amino acids with aliphatic side chains are the major group under positive selection in plants, nine of which are not aliphatic in green algae. In addition, the amino acids with a hydroxyl, sulfur, carboxamide, or cyclic group in the side-chain are the second major group. Basic and acid amino acids are also represented in the

results. Interestingly, there are three basic residues under positive selection which are acidic in the green algal PDATs. All of them are located upstream to the catalytic serine and one of them (Arg149 in AtPDAT1) is in the "lid" region, which might indicate a modified substrate preference (e.g. substrate with a different charge or functional group). Green algal PDATs also contain a site under positive selection in the "lid" (Val291 in CrPDAT). Similar to AtPDAT1, these sites can be related to substrate preference and catalytic activity. In this site, a threonine, alanine or valine (hydroxyl or aliphatic side-chains) is present in green algae and an aromatic amino acid is present in plants.

There are three amino acids under positive selection in the Chlorophyta that are located very close to the catalytic serine (Figure 3.6 and Supplementary figure B.10). Their proximity to the catalytic triad might impact the catalysis directly. The first positively selected site, in the position +1 from the catalytic serine, is conserved as methionine in plants, but it is an aromatic residue in green algal PDATs. Considering it is so close to the catalytic site, it might have a special role in the catalytic activity. Interestingly, this amino acid is also under positive selection in green algal PDATs compared to plant PDATs from group B (Supplementary figure B.11). The residues in this position are highly conserved in both plants and algae and thus may be very important for the protein structure and/or catalysis.

3.3.11. Positively selected sites are important for the proper structure and enzymatic activity of plant PDATs

The sites under positive selection in the homology model of AtPDAT1, CrPDAT and HsLCAT were further mapped to understand how those amino acids impact the structure of the enzymes. Initially, it is possible to observe the sites under positive selection are not enriched in a specific

place or motif of the enzyme (Supplementary figure B.13A). In the AtPDAT1 structure, they are distributed in the α/β domain and on the other relevant subdomains of this enzyme. There are five amino acids under positive selection in the motif between the catalytic aspartate and histidine. Their proximity to the catalytic triad could affect catalysis, which makes them candidates for future studies to investigate catalysis in PDATs. There are also several residues in the structures that form the active site, which could affect substrate selectivity. Furthermore, the side chain of a few AtPDAT1 positive selection sites, specifically Ser158, Thr222, Thr228, and Thr242, are shown to form polar contacts with the main chain atoms (Supplementary figure B.13B) and may be involved in stabilizing the secondary structures. For example, Thr242 is located in an α-helix and its hydroxyl group forms polar contacts with the amide groups of the other residues in the same helix, specifically Leu238 and Met239. On the other hand, the side chains of other sites under positive selection such as Arg149, Ser176, Gln336, and Ser610 (Supplementary figure B.13C) seem to be involved in polar contacts that influence tertiary structures. For example, Arg149 is located at the C-terminus of a helix and its guanidinium group forms polar contact with the carbonyl oxygen of Phe143. Phe143 is part of a preceding helix and the guanidinium-carboxyl interaction may be involved in stabilizing a helix-loop-helix motif in the structure. Altogether, the interactions formed by these sites may play important roles in influencing secondary and tertiary structures. As for CrPDAT, the position of the amino acids under positive selection are also very distributed throughout the structure. The three sites close to the catalytic triad may have a huge impact on the catalysis. In special, Tyr402 appears to occupy a large volume at the bottom of the catalytic pocket, having the potential to be directly involved in the catalysis. It is unclear what is the role of Glu404 and Arg408, but their proximity suggests the formation of a salt bridge in the helix.

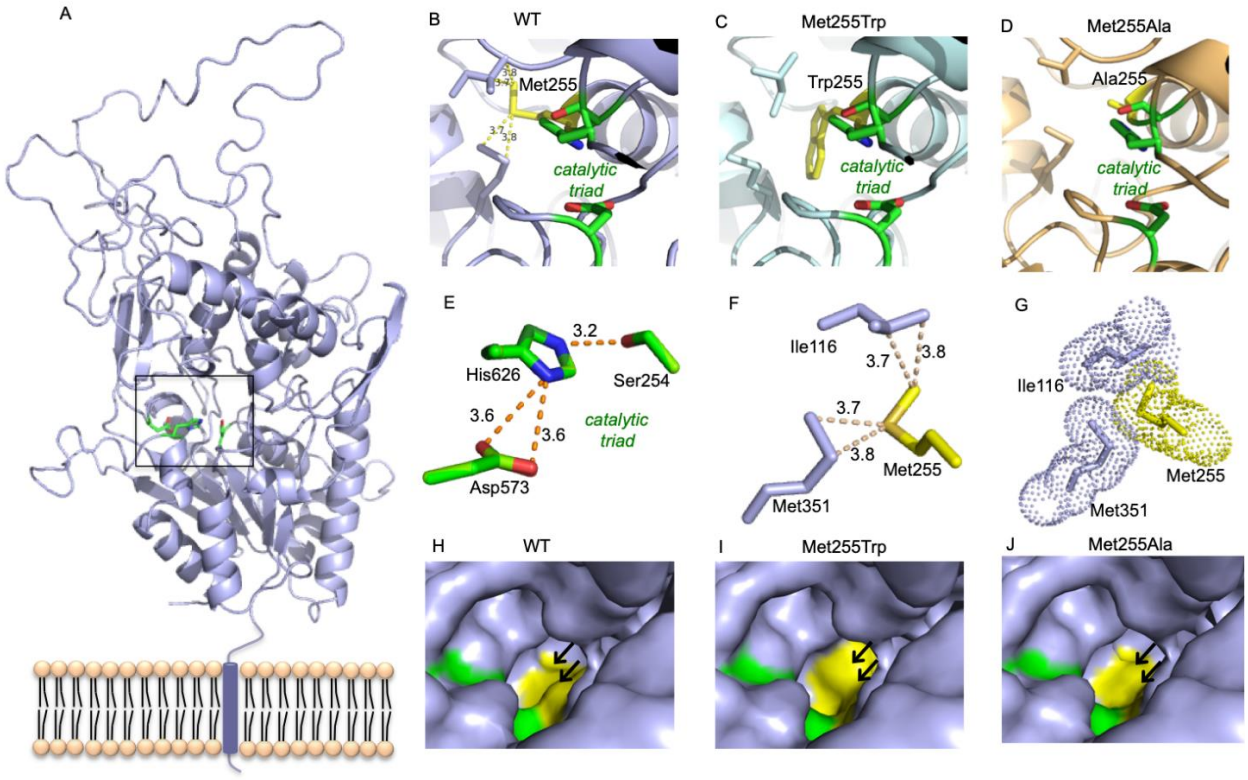


Figure 3.7. Analysis of Met255 in *Arabidopsis thaliana* PDAT1. A) AtPDAT1 homology model obtained using SWISS-MODEL software. B) Close-up view of the active site pocket wherein the side chains for the catalytic triad (Ser254, Asp573, and His626) are shown in different colors with green, red and blue corresponding to carbon, oxygen and nitrogen, respectively. The positive selection site (Met255) is also shown in yellow. C and D) Close-up view of active site pocket for the AtPDAT1 (WT) and the mutant Met255Trp and Met255Ala. E) Alignment of the three models (WT, Met255Trp, and Met255Ala) shows that the position and polar contacts made by the catalytic triad is unchanged during the mutation of Met255. 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 angstrom. F) Met255 appears to form a network of hydrophobic contacts (brown dash in angstrom) with non-polar residues (Ile116 and Met351) lining the active site. The three

non-polar residues form a hydrophobic patch in the active site pocket. G) The van der Waals surface of the non-polar residues are shown in dots. H-J) Surface view of the active site pocket wherein the Met255, 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. Mutation of Met255 may result in disruption of the non-polar interaction leading to changes in the landscape of the hydrophobic patch. This hydrophobic patch may be involved in interacting with the hydrophobic substrates and as such, changes in its landscape may directly affect catalysis. This figure was generated by Dr. Kristian Caldo.

Met255 in AtPDAT1, which is homologous to a site under positive selection in green algae, was chosen to further investigation due to its proximity to the active site pocket (Figure 3.7). In order to understand the relevance of this site for the catalytic activity of AtPDAT1, its structure was initially predicted and compared it with the model structures of Met255Trp and Met255Ala mutants (Figure 3.7B-D). The effect of the mutation of this site on the catalytic triad (Ser254, Asp573, and His 626) was checked by overlaying the 3 structures. A close-up view of the overlaid catalytic triads of the three structures revealed that the position and orientation of these residues within the active site pocket were not affected upon mutation of Met255 to either Trp or Ala (Figure 3.7E). Furthermore, these catalytic residues form polar contacts with one another as expected for a catalytic triad and mutation of Met255 did not affect the distances among the atoms involved in polar contacts. The possible interactions formed by Met255 was then determined to further probe its role in the active site pocket. It was found that Met255 forms a network of hydrophobic contacts with nearby hydrophobic residues namely Met351 and Ile116 (Figure 3.7F). The interactions were verified by measuring the distance of the atoms involved in

the interactions, which are all less than 4 angstroms. The hydrophobic interactions between Met255 and the other 2 hydrophobic residues were further confirmed when the van der Waals radii of the three residues were illustrated using dot representation (Figure 3.7G). In addition, a surface view of the active site pocket shows that Met255, Ile116, and Met351 (shown in yellow) form a hydrophobic patch on the wall of the active site pocket (Figure 3.7H). Since the two substrates of AtPDAT1 are both hydrophobic compounds, this hydrophobic patch may be involved in substrate interaction. Mutation of Met255 to either Trp of Ala appears to disrupt the network of hydrophobic contacts in the active site pocket (Figure 3.7C-D). The hydrophobic patch appears to form two ridges in AtPDAT1 native enzyme (Figure 3.7H). Mutation of Met255 seems to alter the landscape of the hydrophobic patch (Figure 3.7F-G) and as such, the resulting mutation may negatively affect enzyme activity.

In order to test this hypothesis and the relevance of the site under positive selection next to the catalytic serine in PDATs, I prepared AtPDAT1 mutants. Then, their enzymatic activity was determined with NBD-labeled DAG and PtdCho with different acyl groups. The product, NBD-TAG, was quantified by the fluorescence intensity. To obtain the modified enzymes, I introduced a single-point mutation at position 255 to include tryptophan (AtPDAT1-M255W) or alanine (AtPDAT1-M255A). Then, I produced these enzymes in the quadruple-mutant *S. cerevisiae* strain H1246, which is devoid of neutral lipids (Sandager et al., 2002). The results of the enzymatic activity in the microsomes containing the wild-type, the mutants and *LacZ* showed that the activity in the mutants was destroyed (Figure 3.8). Conversely, the wild-type PDAT is able to synthesize NBD-TAG showing a higher preference for C18:2,C18:2-PtdCho, which is consistent with previous findings (Stahl et al., 2004). Therefore, this conserved site is essential for the enzymatic activity of PDATs. A mutation may have a huge impact on the active site,

affecting the catalysis itself. Furthermore, the presence of aromatic amino acids in that position makes sense only in green algal PDATs, because the mutant AtPDAT1-M255W also did not present activity.

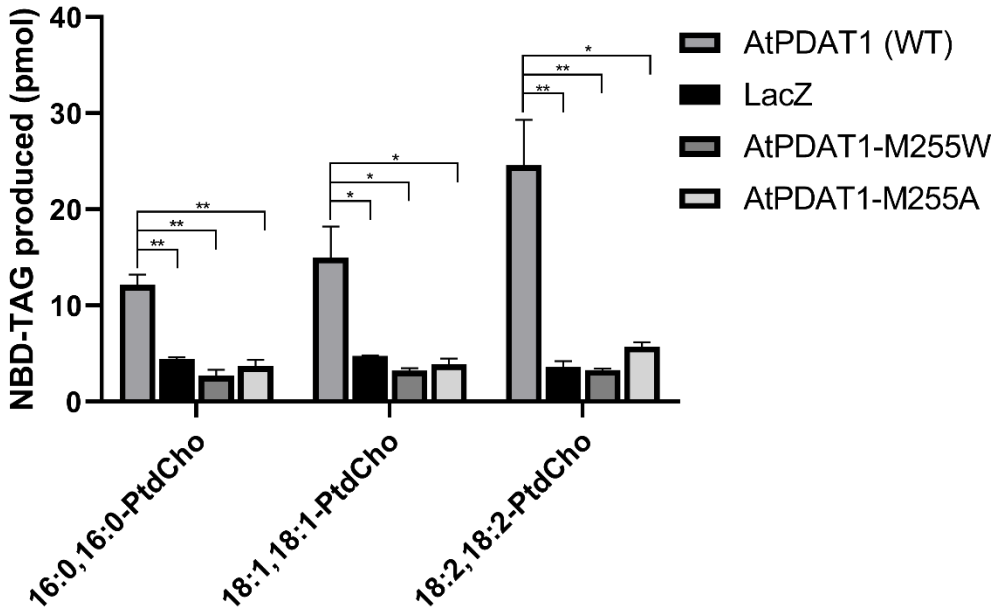


Figure 3.8. Assessment of the catalytic activity of microsomes containing *LacZ*, wild-type (WT) *Arabidopsis thaliana* PDAT1 (AtPDAT1), AtPDAT1-M255W and AtPDAT1-M255A. The last two enzymes are an AtPDAT1 subjected to single-point mutagenesis. The reaction was carried out with NBD-DAG and different PtdCho species (16:0,16:0-PtdCho, 18:1,18:1-PtdCho, and 18:2,18:2-PtdCho) as the substrates. A t-test assuming unequal variances shows that the production of NBD-TAG for AtPDAT1 (WT) microsomes is higher than the mutants or LacZ (*$p < 0.05$; **$p< 0.01$).

The effects of the amino acid next to the catalytic serine in the structure of the catalytic pocket of the human LCAT and CrPDAT were also observed (Supplementary figure B.14-B.15).

As the conformation of the pocket is essential to the enzyme reaction mechanism, replacing the leucine in HsaLCAT for methionine or tyrosine will likely affect the activity of the enzyme. Furthermore, why green algae have an aromatic amino acid in that position is unclear, but it might be related to the enzymatic structure (e.g. larger pocket for larger substrates) or catalytic activity (e.g. substrate binding). Overall, the first residue downstream the catalytic serine is important for maintaining a specific landscape of the hydrophobic patch lining of the catalytic pocket. A mutation in that position is very likely to affect the catalytic activity of PDATs.

3.3.12. The presence of aromatic amino acids next to the serine in the catalytic triad is unique to green algae

To understand the conservation of the residue in the position +1 to the catalytic serine in the LCAT family in other organisms, I built a phylogenetic tree using LCAT-like homologs found in a sponge (*Amphimedon queenslandica)*, a fish (*Danio rerio)*, a model worm (*Caenorhabditis elegans)*, a fly (*Drosophila melanogaster*), the horse (*Equus caballus*), the domestic chicken (*Gallus gallus*), a frog (*Xenopus tropicalis*), the mouse (*Mus musculus*), the human (*Homo sapiens*), a filamentous fungi (*Aspergillus nidulans*), ScePDATa, AtPDAT1 and CrPDAT (Figure 3.9). As expected, no PDATs have been found among the animal genomes. However, they contain LCAT and PLA2, each one forming a monophyletic group. LCAT was not found in the genomes of *D. melanogaster* and *C. elegans.* The LCAT-related enzymes from the sponge form a basal group to the enzymes from other organisms. Interestingly, the LCAT enzymes form a sister clade to the PLA2, instead of forming a cluster with the other LCAT. This shows a considerable change in LCAT in higher organisms.

Regarding the residue of interest, we observe it is conserved as methionine among PLA2 and it can be either methionine or leucine in LCAT (Figure 3.9). The fungi PDATs also contain methionine in that position. PDATs in chromists and plants have a methionine and green algae contain an aromatic amino acid (Supplementary figure B.1). In addition, plant PSAT, algal PSAT and LCAT-related have methionine or leucine in that position. The plant LCAT-related have solely leucine. Thus, I conclude that PDATs contain preferentially methionine next to the catalytic serine, but the presence of aromatic amino acids is unique to Chlorophyta. Furthermore, other LCAT-related enzymes may have methionine or leucine in that position. Together with the enzymatic assay results, they show that this position is extremely conserved in LCAT family, accepting only a limited number of different substitutions. Therefore, a mutation in this site could compromise the catalytic activity by affecting the active site conformation.

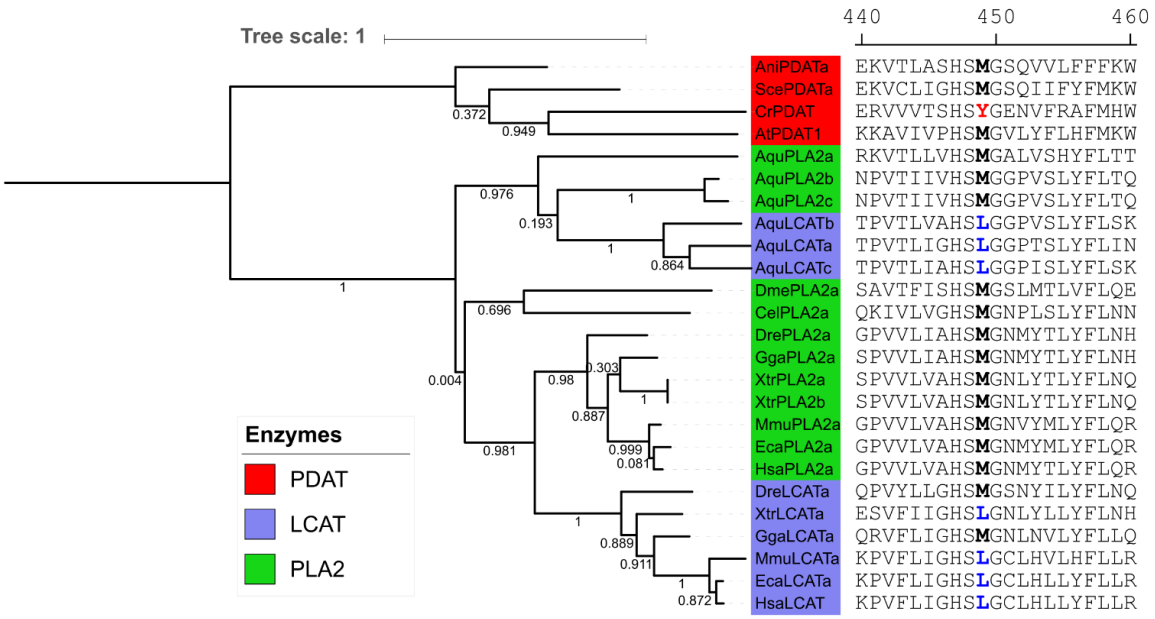


Figure 3.9. The conservation of amino acids in the position +1 downstream to the catalytic serine in animal, fungi, plant and green algae LCAT-related enzymes. Methionine is shown in black, leucine in blue and tyrosine in red in the alignment. LCAT, lecithin:cholesterol acyltransferase; PDAT, phospholipid:diacylglycerol acyltransferase; PLA2, phospholipase A2.

3.4. Discussion

3.4.1. Enzymes from LCAT family are not conserved between plants and algae

While a few PDATs have been characterized in plants, algae and yeast, there is not much information regarding other LCAT-family enzymes, especially in microalgae. In addition, the previous *in silico* characterization is limited to plant PDATs (Pan et al., 2015), whereas the algal PDATs are poorly studied yet. In this study, I explored the differences among the subfamilies and between algae and plants to explain their evolution and characterize the LCAT family in both groups.

As members of LCAT family, algal and plant PDATs have conserved domains to human LCAT (Figure 3.2A and Supplementary figure B.1). The "lid" region, the salt bridge, and the three catalytic residues are conserved and present in all groups of enzymes. Previously, the "lid" domain was not predicted between the cysteines that form the disulfide bridge in plant PDATs (Pan et al., 2015). Contrarily, I identified this region present between two cysteines, with tryptophan in the middle of the motif (Trp139 in AtPDAT1). Considering these residues are very conserved among the PDAT sequences, it is very likely they are essential for the proper catalytic activity of these enzymes in algae and plants (Peelman et al., 1998). In addition, the position of these motifs is also very conserved in the structure of PDATs (Figure 3.2B).

As shown previously, the LCAT-family enzymes may catalyze more than one type of reaction. For instance, human LCAT can catalyze both phospholipase and acyltransferase reactions (Peelman et al., 1999). The algal CrPDAT has lipase and phospholipase activities in addition to the acyltransferase activity (Yoon et al., 2012). However, the phospholipase activity was not clearly demonstrated in AtPDAT1 or any other plant PDAT (Dahlqvist et al., 2000;

Stahl et al., 2004). The presence of multiple functions in LCAT-like enzymes from algae rather than plant might be related to the difference in the length of these enzymes. Indeed, by comparing the properties of PDATs and other LCAT-family enzymes from algae and plants, I found the algal enzymes, in general, are longer than their plant homologs.

Several hydrolytic enzymes, such as lipases and serine proteases, have similar features, indicating they have a common origin. The α/β hydrolase fold, the catalytic triad, and the amino acid sequence pattern Gly-X-Ser-X-Gly (G-X-S-X-G, also known as "lipase" motif) are usually present in their structures (Derewenda and Derewenda, 1991; Ollis et al., 1992). Analogous motifs are also present in the LCAT-family enzymes and thus a similar catalytic activity is expected. The lipase motif, in special, has some differences among the enzymes studied in algae and plants. The first glycine is substituted by alanine among PSAT, serine or threonine in plant LCAT-related, and cysteine, glycine or alanine in algal LCAT-related (Motif 12 in Supplementary figure B.3). In addition, algal PDATs have a serine, threonine or alanine and plant PDATs have a very conserved proline in the same position. Thus, the first residue in this motif is crucial to distinguish LCAT-like enzymes from other hydrolases. Furthermore, it could be used to discriminate enzymes among LCAT subfamilies. Considering the first glycine is subject to several constraints in hydrolases (Derewenda and Derewenda, 1991), it is possible that this position was changed in LCAT to allow new substrates, and, consequently, new catalytic activities. In general, the "lipase" motif in plant and algal LCAT-like could be represented as ([P/S/T/C/A/G]HS[M/L/Y/W]G) and could be considered a motif for the characterization of LCAT enzymes in these organisms. In particular, the differences in this important pentapeptide corroborate to explain the functional divergence between plant and green algal PDATs.

The “lid” region can also be related to the functional divergence present in PDATs in algae and plants. Some very conserved residues in this motif were observed (e.g. cysteines of the disulfide bridge and tryptophan that binds the substrate), while other amino acids were changed during evolution. The deletion of the “lid” domain in human LCAT was shown to destroy the activity of this enzyme (Peelman et al., 1998). In addition, there is certain plasticity in this region, which could allow for the accommodation of different substrates, as shown in the crystal structure of the human lysosomal PLA2 (Glukhova et al., 2015). The “lid” domain is required for the interaction with the membranes, and, consequently, the substrates of the enzyme (Adimoolam and Jonas, 1997; Glukhova et al., 2015; Piper et al., 2015). Therefore, the differences in the “lid” between algal and plant PDATs, and the higher variability in PDATs may reflect the need to adapt to different substrates and functions in each organism. For instance, CrPDAT was predicted to act in the chloroplast and was shown to have activity towards galactolipids, which are enriched in that organelle (Yoon et al., 2012).

TMDs have been found in previous characterizations of PDATs (Pan et al., 2015; Stahl et al., 2004; Yoon et al., 2012). In this study, I could not predict the presence of these structures in several LCAT-like sequences. Instead, it seems that the TMDs are more important for PDATs because they were predicted mostly among plant and algal sequences (Table 3.1, Figure 3.3 and Supplementary figure B.2). In particular, most of the chromists lack a TMD in that region. The alignment of PDATs shows that the sequences lacking an initial portion of the N-terminal region (upstream to the “lid” domain) also lack a TMD. Thus, the presence of TMD in PDATs seems to be important among Viridiplantae organisms. The results also show that the presence of TMD may not be essential for the activity of PDATs, considering some sequences do not contain this motif. In *C. reinhardtii*, it has been shown that CrPDAT with removed TMD retained its activity

but was not able to use monogalactosyldiacylglycerol (MGDG) to produce TAG any longer (Yoon et al., 2012). Therefore, the presence of a TMD in PDAT, when necessary, may be associated with the recognition of different membrane compositions.

The region between the initial portion of the "lid" domain and the catalytic serine seems to be very important for the catalytic activity and structural conformation of PDATs (formed by the Motifs 5, 9, 14, 18, 1, and 12). This region also contains the salt bridge motif, which is the most conserved sequence in the LCAT family, and part of the α/β hydrolase domain (Glukhova et al., 2015; Piper et al., 2015). In the human LCAT, subjects containing homozygous mutations in the same region presented familial LCAT deficiency and fish-eye disease, which are caused by the total or partial loss of LCAT activity in the plasma (Calabresi et al., 2005; Charlton-Menys et al., 2007; Contacos et al., 1996; Guerin et al., 1997; Holleboom et al., 2011; Steyrer et al., 2004). Considering this part of the α/β hydrolase domain is predicted to have a variable conformation, the shape of this region is also important to determine the substrates that can be accepted by the LCAT-family enzymes (Glukhova et al., 2015).

Although there is a region with higher conservation between algal and plant PDATs, there are several motifs present solely in one of these two groups. They are mostly concentrated in the region downstream to the catalytic serine (Motif 12; Figure 3.3 and Supplementary figure B.4). From the 32 motifs that are unique to algal or plant PDATs, 24 are located downstream the Motif 12, which covers 73% and 31% of the residues located in the same portion of AtPDAT1 and CrPDAT, respectively. These motifs contain several substitutions that were accumulated during the evolution of algal and plant PDATs and represent strong evidence of divergent evolution among PDATs from these groups. The presence of the conserved Motif 8 representing the TMD in plants also supports the functional divergence among PDATs because this region is

not conserved in algae. Our results are consistent with a previous study that showed plant PDATs follow a similar pattern, with the region between the "lid" domain and catalytic serine being the most conserved and the region downstream containing several motifs that are not shared between PDATs from group VII and from group V and VI (Pan et al., 2015). Therefore, it is very likely they have evolved under different selective pressures, which caused such a divergence in their conserved motifs. In addition, the evolution seems to be more relaxed in the region downstream the catalytic serine. The following sections will show that plant and algal PDATs evolved under functional divergence, which explains the results discussed here.

3.4.2. PDATs are more important for plant metabolism

Similar to other LCAT-like sequences used in this study, PDATs form a monophyletic group. In addition, plant PDATs are clustered into a single clade. These results are consistent with those from previous publications (Pan et al., 2015; Yoon et al., 2012). To characterize the PDATs in other eukaryotic photosynthetic organisms, I also included sequences from chromists and a red alga. The results showed in the phylogenetic and evolutionary analyses confirm the presence of functional divergence between plant and algal PDATs and revealed a higher relevance of PDATs to land plants than algae.

The first evidence of lower relevance of PDATs to algae is their absence in several algal organisms, including green algae. LCAT-like enzymes were not found in nine genomes, whereas eleven did not contain a PDAT. It is possible that in future annotations of some of those genomes, enzymes from the LCAT family could be found. However, the PDAT loss is not surprising because their function could be replaced by other enzymes, such as DGAT1 or DGAT2 (Vieler et al., 2012; Xu et al., 2018a; Yoon et al., 2012). Conversely, all plant species in

this study contained LCAT-like enzymes and their mean copy number for each organism is higher than in algae. It suggests higher importance of LCAT-like enzymes for the land plants, especially considering that there might be a tissue-specific regulation for each copy. For instance, the expression of *AtPDAT1* is higher in leaves, flower and roots than in seeds; *AtPDAT2* has a higher expression in seeds (Pan et al., 2015).

Ancient genetic duplication is not present in algal PDATs, which may support their lower relevance for algae when compared to land plants. It is possible to observe two recent duplications in the genomes of *S. reticulata* and *P. purpureum*. In contrast, plants have a high copy number of PDATs (Figure 3.1). In addition, there are at least two ancient duplication events among PDATs, especially among eudicots, followed by recent duplication in some species (Pan et al., 2015). The retention of extra copies of *PDATs* in the evolution of eudicots shows that they might have contributed to increased fitness in these plants unless some of these genes became a pseudogene (Panchy et al., 2016). Overall, with the higher copy number among plants that were kept during evolution and with the greater conservation of the sequences, we can infer that PDATs have a higher impact on the plant metabolism.

In general, the evolution of algal PDATs followed the species evolution, especially inside each monophyletic group (Figure 3.4 and Supplementary figure B.6). However, they may not have followed the evolution of major algal groups, because the red algal PDATs are preferentially clustering with chromists. I showed that the position of the *P. purpureum* PDATs was not matching in the ML and BI phylogenetic trees, but the topological hypothesis test supported the red algal PDATs forming a monophyletic group with the Bacillariophyta. Considering that several of the current theories place the chromists as a separate group and the Rhodophyta and Viridiplantae in the Archaeplastida (Clerck et al., 2012; Jackson and Reyes-

Prieto, 2014; Price et al., 2012), our results corroborate a different evolutionary pattern for red algal or green algal PDATs. It is possible that Rhodophyta and plant PDATs had a much similar evolution since they are evolutionarily closer than with the green algal enzymes (Figure 3.5 and Supplementary figure B.7). Thus, it is possible to suggest that the selective forces acting on the green algal sequences caused the divergence from the red algal PDATs.

In a similar manner, we observed that the evolution of PDATs in green algae and plants do not follow the species evolution. A closer evolution would be expected since both groups belong to the Viridiplantae and share a common ancestor (Clerck et al., 2012; Leliaert et al., 2012). Instead, chromist PDATs are closer to plants than to green algal sequences, suggesting the presence of functional divergence in PDATs for the major groups of Viridiplantae. In this case, it is possible that green algal enzymes are also divergent from the chromist PDATs, having a unique evolution. Results from the selection pressure analysis confirmed that algal and plant PDATs evolved under divergent selective pressures. In both comparisons, more than 50% of the analyzed residues were classified at site class 2 and the ω value for the algal group was higher. These results help to explain the divergence in the evolution of green algal and plant PDATs. Indeed, the substitutions accumulated in both groups may have led to new functions, such as gain or loss of catalytic activities and the ability to use different substrates. It has been shown that CrPDAT has lipase and phospholipase catalytic activities, while it was not found among studied plant PDATs (Dahlqvist et al., 2000; Pan et al., 2013; Stahl et al., 2004; van Erp et al., 2011; Yoon et al., 2012). It is not possible to conclude whether these catalytic activities were present in the ancestral PDAT. CrPDAT was also shown to be able to use several membrane lipids for TAG synthesis, with a lower preference for PtdCho (Yoon et al., 2012). Contrarily, PtdCho is preferred over phosphatidic acid for AtPDAT1 (Stahl et al., 2004). This difference can be

explained by the presence of functional divergence during PDAT evolution and by the absence of PtdCho in the composition of *C. reinhardtii* (Dembitsky, 1996).

The results from the selective pressure analysis also show that algal PDATs evolved under more relaxed constraints, which means that more mutations were allowed during evolution with minimal impact on fitness (Yang, 2007; Yang et al., 2000). These results could be inconsistent with the high oil accumulation by several green algal species (Hu et al., 2008). However, these species contain an elevated number of DGATs (Chen and Smith, 2012). *C. reinhardtii* and *Volvox carteri*, for instance, have one DGAT1 and five DGAT2 in their genomes. These enzymes catalyze the biosynthesis of TAG and are, thus, functionally redundant to PDATs, especially in unicellular organisms. Therefore, PDATs could have evolved under a more relaxed selective pressure, because DGATs would replace or complement PDAT activity in TAG formation. In addition, the higher conservation shown for the plant PDATs suggests they evolved to be more important for these organisms (to their metabolism and fitness) when compared to green algal species.

The difference in the ω values between green algal and plant (group A) PDATs is higher than when compared with the plant (group B). It shows that this group evolved under a different selective pressure when compared to PDATs from group A. Our results agree with a previous publication that showed PDATs from that clade are under relaxed evolution when compared to the other monophyletic groups (Pan et al., 2015). This could be explained because the PDAT activity was not demonstrated for sequences from group B (Pan et al., 2013; van Erp et al., 2011; Zhang et al., 2009), suggesting a different function or possible pseudogenization for these enzymes (Pan et al., 2015).

Positive selection is usually associated with a beneficial mutation that was retained by a population due to increased fitness (Hughes, 2007). In addition, they might be the substitutions leading to the functional divergence, such as the one in the evolution of mammalian heme peroxidases (MHPs), which was caused by positive selection (Loughran et al., 2008). In this study, I showed that positive selection is part of PDAT evolution. The comparison of the branch-site model A with the null model confirmed the presence of positive selection for both green algal and plant PDATs. In both analyses, the plant group had a higher number of residues at site class 2 when compared to algae. Therefore, these results support the hypothesis of divergent evolution and confirm the higher relevance of PDATs for plants. The identification of sites under positive selection with the BEB method also supports these conclusions. The 42 sites identified in plant PDATs from group A compared to 12 in green algal enzymes suggest more beneficial substitutions were retained in plant PDATs. A large number of amino acids under positive selection downstream the catalytic serine (9 green algal residues and 23 plant residues) also explains why this region contains several motifs present only in algae or plants. Finally, there are fewer sites under positive selection in the group B of plant PDATs when compared to group A. These results suggest a lower relevance of this group to plants, supporting the hypothesis of nonfunctionalization (Pan et al., 2015).

In this study, (i) PDATs from algae and plants were shown to have structural differences. In addition, (ii) some conserved motifs were not present in both groups. In terms of physiological relevance, (iii) some algae do not have a *PDAT* gene in their genome, while most plants contain at least one copy; (iv) eudicots contain several copies acquired through gene duplication. I also showed that green algae did not evolve closer to (v) red algal and (vi) plant PDATs. Furthermore, (vii) selective pressure analysis confirmed the evolution of green algal and plant

PDATs under different selective constraints, with (viii) the algal group having a higher ω value. Finally, the (ix) presence of positive selection was demonstrated for both algal and plant PDATs, with (x) the plants having more amino acids that evolved under positive selection. The redundancy with DGATs might have contributed to the relaxed evolution of PDATs in algae, suggesting a minor role in the biosynthesis of neutral lipids. Knockout and knockdown experiments in *A. thaliana* and *C. reinhardtii* support these results. The TAG accumulation is especially important in algae cultivated under stress conditions (Li-Beisson et al., 2015). In this case, it would be expected a contribution of PDAT for TAG biosynthesis and, consequently, the cell survival. The knockdown of *PDAT* did not impact oil accumulation in *C. reinhardtii* cells grown under nitrogen-starvation (Yoon et al., 2012). Another study showed that the reduction in the TAG content could be up to 25% in PDAT insertional mutants of *C. reinhardtii*, but algal cells lacking PDATs are still viable (Boyle et al., 2012). Conversely, the disruption of AtPDAT1 in *A. thaliana* showed a decrease higher than 50% in oil accumulation in growing tissues, which suggests that DGATs play a minor role in the biosynthesis of TAG in these tissues (Fan et al., 2013a, 2013b). Moreover, PDAT appears to be crucial for the correct development of plants since the mutant *A. thaliana* plant showed necrotic developing leaves, growth retardation and gametophytic defects (Fan et al., 2013a, 2013b; Zhang et al., 2009). Therefore, the *in vivo* studies corroborate our results in the conclusion that PDATs are essential for plants, including their correct development and survival, but are not fundamental for green algae.

3.4.3. Impact of amino acids under positive selection on the structure of PDATs

In this study, a large number of sites were predicted to have evolved under positive selection. It is more evident in the comparison between green algae and plant (group A) PDATs, whereby 12

and 42 sites were predicted, respectively. Considering positive selection increases fitness (Hughes, 2007), a mutation in these residues could be harmful to the enzymes and the organisms. For instance, three sites under positive selection were identified in myeloperoxidase (MPO) and subjected to site-directed mutagenesis (Loughran et al., 2012, 2008). The mutants failed to process MPO precursors into normal subunits. In addition, they showed a reduction in their peroxidation activity and lost chlorination activity. Similarly, it is possible that a mutation in the residues under positive selection causes an alteration on the PDAT structure and catalysis, but *in vivo* and *in vitro* work are necessary to confirm the effects. In this analysis, the sites under positive selection were shown to be evenly distributed in the structures of AtPDAT1 and CrPDAT (Supplementary figure B.12), suggesting they can impact from substrate recognition to catalysis.

As explained, part of the functional divergence might be related to gain or loss of catalytic activities and recognition of different substrates. In this study, two residues under positive selection were identified in the "lid" domain (Val291 in CrPDAT and Arg149 in AtPDAT1). Considering this region has high conformational plasticity, these sites might be important for substrate recognition and binding (Glukhova et al., 2015; Peelman et al., 1998). In addition, four sites under positive selection in green algae (Val291, Val489, Ala728, Ile1029 in *C. reinhardtii*) have aliphatic chains and corresponding aromatic amino acid in plants. Site-directed mutagenesis was used to replace four very conserved aromatic amino acids in the active site in xylanase A of *Streptomyces lividans* (Roberge et al., 1999). The results revealed the aromatic residues are involved in substrate-binding, and one of them also contributes to the catalysis. Aromatic amino acids were also shown to contribute to substrate discrimination in a xylose isomerase (Meng et al., 1993). In plant PDATs, I suspect similar roles for the aromatic

amino acids. Indeed, mutagenesis of Trp61 in human LCAT affected the specific activity of this enzyme (Peelman et al., 1998). Therefore, the gain of aliphatic amino acids in green algae to replace aromatic side chains may reflect an adaptation towards new substrates or the adjustment to a new enzymatic conformation. In a similar manner, the charge of the amino acids might be related to substrate specificity and catalytic activity. There are three basic amino acids under positive selection in plant PDATs with corresponding acidic residues in green algae. For instance, Glu149 was replaced by alanine in the human LCAT, which changed the substrate preference from *sn*-1-palmitoyl-2-oleoyl-PtdCho to *sn*-1-palmitoyl-2-arachidonoyl-PtdCho (Wang et al., 1997).

In PDATs, there are two regions with residues under positive selection that are close to the catalytic site and might have a huge impact on the catalytic activity. The first region contains three residues that are very close to the catalytic serine. They are conserved as tyrosine, glutamic acid, and arginine in green algae but methionine, valine and leucine in plants, respectively. These residues may affect the conformation of the active site to allow different substrates or allow the catalysis of other reactions. There is also a minor possibility of residues Glu404 and Arg408 form a salt bridge because they are close, but not less than 4.0 Å apart. Unless a high-resolution crystal structure of CrPDAT clarifies the real distance of these amino acids, it will not be possible to determine if their interaction qualifies as a salt bridge (Donald et al., 2011). These residues are unique to green algae as they are not present in chromist or *S. cerevisiae* PDATs.

The second region contains five residues under positive selection (Val578, Ala581, Tyr603, Ser610, and Pro612 in AtPDAT1), which are located between the catalytic aspartic acid and catalytic histidine in plant PDATs. Their position in the structure, close to the catalytic triad, is very likely to affect catalysis. For instance, Ser610 was shown to interact with Gly527 to

stabilize the tertiary structure. Mutations in this region are predicted to disrupt the catalytic activity of the catalytic amino acids (aspartate or histidine) or affect the substrate binding (Piper et al., 2015). Indeed, the mutations L372R and T347M, which are in the same region, in the human LCAT cause familial LCAT deficiency and fish-eye disease, respectively (Calabresi et al., 2005; Klein et al., 1993). Similar to Ser610, other sites were shown to participate in the formation and stabilization of secondary and tertiary structures. These sites could have introduced new features in the PDAT structure to improve or adapt PDATs into plants. In addition, they are very important for proper catalysis. The site under positive selection Thr228 in AtPDAT1 is equivalent to Tyr156 in HsLCAT. In this case, a mutation of Tyr156 to Asn156 reduced LCAT activity contributing to familial LCAT deficiency (Klein et al., 1993).

The residue in the position +1 to the catalytic serine in AtPDAT1 (Met255) was chosen to be further studied with structural homology, site-directed mutagenesis and *in vitro* PDAT assay. It is a very important site because it is close to the catalytic triad and it is the only amino acid under positive selection identified in algae (Tyr402 in *C. reinhardtii*) in the comparison with both plant PDAT groups. This residue is very conserved in plants as methionine and in green algae as an aromatic amino acid. According to the predicted AtPDAT1 structure, Met255 forms a hydrophobic interaction with Met351 and Ile116 (Figure 3.7). A methionine in this conserved site is essential to the catalytic activity of the enzyme because its change with tryptophan or alanine disrupts the hydrophobic patch formed by those amino acids. The conformation in this region of the active site is essential to the enzyme because a mutation completely destroys the activity (Figure 3.8) since only the wild-type AtPDAT1 catalyzed NBD-TAG synthesis *in vitro*. Indeed, a Met252 in HsLCAT aligns in the same position as Met351 in AtPDAT1 and a mutation to lysine also destroys the enzymatic activity causing familial LCAT deficiency (Skretting et al., 1992). In

a similar manner, a mutation in the position -2 of catalytic serine in human LCAT (Gly179) also lost activity (Wang et al., 2011). Therefore, mutations in the sites sequentially or structurally close to the catalytic serine tend to negatively affect the activity of the enzyme. These results also show that a positive selection residue in one group does not necessarily have a beneficial effect on the other. In green algae, other substitutions were necessary to establish the large aromatic amino acids as beneficial. For instance, the positions +3 and +7 to the catalytic serine are also under positive selection and are very conserved in green algae. They might adjust the conformation of the catalytic pocket to allow aromatic amino acids. I compared the motif of the catalytic serine among animals and fungi to understand how the position +1 to the catalytic serine is conserved in the LCAT family. Methionine is preferred for PLA2 and fungi PDAT. Leucine or methionine are present among LCAT. Such conservation shows that this residue is important for the catalysis. In addition, aromatic amino acids are present solely in green algae. Based on our results, it can be suggested the human LCAT could lose its activity with a mutation in the same position.

3.5. Conclusion

Algal and plant LCAT-family enzymes, specifically PDATs, were characterized in this work. Several differences between LCAT enzymes from plants and algae were observed. Algal LCAT enzymes are longer and may not have TMDs. In addition, there is a region between the “lid” domain and the catalytic serine that is more conserved and would be more affected by mutations. PDATs from green algae and plants evolved under functional divergence and positive selection. I also showed that PDATs are not essential to green algae, but they are necessary for plant metabolism. In addition, mutations in positively selected sites might negatively affect catalysis.

Mutagenesis in the methionine in the position +1 to the catalytic serine destroyed AtPDAT1 activity, showing the importance of these sites. Finally, I showed that the presence of aromatic amino acids in that position is unique to green algal PDATs. Overall, our findings could be used in future biotechnological approaches to improve or modulate the lipid biosynthesis in algae and plants.

CHAPTER 4. Conclusions and perspectives

Considering that PDAT catalyzes the last step in the biosynthesis of TAG, it is essential to understand its relevance with algal and plant metabolism. Understanding the lipid metabolism is especially important for using molecular biotechnology approaches to improve these organisms, which may guarantee the viability of lipid-based projects. In this study, I developed a fluorescence-based PDAT assay to facilitate the characterization of PDATs and assessed their evolution and structure in algae and plants.

I proposed the use of fluorescent compounds as an alternative to radiolabeled substrates in the standard PDAT assay (Chapter 2). Considering that the mechanism of reaction of PDATs may be similar to LCAT and phospholipase A2, the large fluorescent group should be located at the end of the acyl group. This configuration would minimize the impact of large fluorescent groups on catalysis since its position is far from the catalytic triad. NBD-DAG, obtained from the commercially available NBD-PtdCho, was shown as an adequate substrate in the characterization of PDATs. In this regard, the enzymatic activity and kinetics of AtPDAT1 produced in the microsomal fraction of yeast was described. This method was also validated with the conventional assay. I further showed that diethyl ether can substitute the toxic benzene. Overall, this novel methodology is reliable, safer and less expensive than the standard method.

This NBD-based method can also serve as the basis for easy modification for other applications. This method is ideal to compare the activity of different PDATs or the specificity of acyl groups. In the future, the method can be adjusted to test different DAG molecular species. In this case, NBD-PtdCho with the NBD group in the *sn*-2 acyl chain should be the acyl donor. Similarly, different NBD-labeled phospholipids are commercially available and their specificity could be determined with this assay. The method can also be used for the estimation of PDAT

activity directly in microsomal fractions of algae and plants. Moreover, the enzymatic activity of purified PDATs could be assayed with fluorescent-labeled substrates. Finally, the method can be modified to allow the quantification of NBD-TAG with other instruments, such as HPLC, which could provide a better separation from the unknown compound.

After establishing the PDAT assay method, PDATs from eukaryotic photosynthetic organisms were characterized in this project (Chapter 3). A dissimilar evolution for the PDATs from algal and plant groups was demonstrated. In special, green algal PDATs seem to have diverged from chromist and red algal PDATs. They are also functionally divergent from plants, which is unexpected since they have a close evolution. Positive selection is also part of the evolution of PDATs in green algae and plants. In this case, 12 and 42 sites are under positive selection, respectively. These findings show that PDATs are essential for plants and less important for green algal metabolism.

In terms of enzyme properties, enzymes from the algal LCAT family are usually longer than their plant homologs. However, important features of LCAT enzymes were demonstrated to be conserved in the studied PDATs, such as the "lid" region, the salt bridge, and the catalytic triad. The analysis of conserved motifs showed that PDATs have a region of higher conservation between the "lid" domain and the catalytic serine. The homology model showed that some sites under positive selection participate in the formation of secondary and tertiary structures. Moreover, the first residue downstream to the catalytic serine is under positive selection in green algae. A mutation of this residue in AtPDAT1 was predicted to negatively affect the enzymatic activity because the conserved methionine interacts with other amino acids. The assay of AtPDAT1-M255W and AtPDAT1-M255A confirmed that they completely lost the enzymatic

activity. Overall, the information provided in this study contributes to a better understanding of the PDATs in eukaryotic photosynthetic organisms.

Future analyses derived from this project should combine the novel methodology and site-directed mutagenesis to assay the impact of mutations in other sites identified as under positive selection. The *in vivo* effects of the modified PDATs should be assessed to further understand the relevance of PDATs for the organisms. In addition, it would be possible to correlate the *in vivo* and *in vitro* results. For instance, sites under positive selection in algae could initially be tested *in vitro* and those with confirmed activity could be subjected to *in vivo* assessments (in algae or plants). Further studies are also necessary to understand the role of TMDs, since some enzymes, especially in algae, are devoid of this structure.

Currently, we interpret the PDAT structure based on the established crystal structure of human LCAT and lysosomal PLA2. The determination of the structure of PDATs would confirm their mechanism of action and clarify the predicted regions of the structure. Moreover, knowing the structure of an algal PDAT would show why they are longer and the conformation of the structures not present in plants. Further experiments are necessary to confirm plant *PDATs* from group B are pseudogenes. New studies are also necessary to characterize the function of LCAT-related enzymes, which could also reveal their relevance for both plants and algae. In this regard, the information from this study can be a starting point for future studies related to enzymes from the LCAT family.

REFERENCES

Abascal, F., Zardoya, R., Telford, M.J., 2010. TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 38, W7–W13.

Adimoolam, S., Jin, L., Grabbe, E., Shieh, J.J., Jonas, A., 1998. Structural and functional properties of two mutants of lecithin-cholesterol acyltransferase (T123I and N228K). J. Biol. Chem. 273, 32561–32567.

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

Allen, M.R., Barros, V.R., Broome, J., Cramer, W., Christ, R., Church, J.A., Clarke, L., Dahe, Q., Dasgupta, P., Dubash, N.K., Edenhofer, O., Elgizouli, I., Field, C.B., Forster, P., Friedlingstein, P., Fuglestvedt, J., Gomez-Echeverri, L., Hallegatte, S., Hegerl, G., Howden, M., Jiang, K., Jimenez Cisneros, B., Kattsov, V., Lee, H., Mach, K.J., Marotzke, J., Mastrandrea, M.D., Meyer, L., Minx, J., Mulugetta, Y., O'Brien, K., Oppenheimer, M., Pachauri, R.K., Pereira, J.J., Pichs-Madruga, R., Plattner, G.-K., Pörtner, H.-O., Power, S.B., Preston, B., Ravindranath, N.H., Reisinger, A., Riahi, K., Rusticucci, M., Scholes, R., Seyboth, K., Sokona, Y., Stavins, R., Stocker, T.F., Tschakert, P., van Vuuren, D., van Ypersele, J.-P., Blanco, G., Eby, M., Edmonds, J., Fleurbaey, M., Gerlagh, R., Kartha, S., Kunreuther, H., Rogelj, J., Schaeffer, M., Sedláček, J., Sims, R., Ürge-Vorsatz, D., Victor, D., Yohe, G., 2014. IPCC fifth assessment synthesis report - Climate Change 2014 synthesis report. Intergovernmental Panel on Climate Change (IPCC).

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Amaro, M., Filipe, H.A.L., Prates Ramalho, J.P., Hof, M., Loura, L.M.S., 2016. Fluorescence of nitrobenzoxadiazole (NBD)-labeled lipids in model membranes is connected not to lipid mobility but to probe location. Phys. Chem. Chem. Phys. 18, 7042–7054.

Ambati, R.R., Phang, S.M., Ravi, S., Aswathanarayana, R.G., 2014. Astaxanthin: Sources, extraction, stability, biological activities and its commercial applications - A review. Mar. Drugs.

Ang, B.W., Choong, W.L., Ng, T.S., 2015. Energy security: Definitions, dimensions and indexes. Renew. Sustain. Energy Rev. 42, 1077–1093.

Anisimova, M., Gascuel, O., 2006. Approximate likelihood-ratio test for branches: A Fast, accurate, and powerful alternative. Syst. Biol. 55, 539–552.

Anisimova, M., Gil, M., Dufayard, J.F., Dessimoz, C., Gascuel, O., 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst. Biol. 60, 685–699.

Archibald, J.M., 2012. The evolution of algae by secondary and tertiary endosymbiosis, in: Piganeau, G. (Ed.), Genomic Insights into the Biology of Algae, Advances in Botanical Research. Academic Press, pp. 87–118.

Arenas, M., 2015. Trends in substitution models of molecular evolution. Front. Genet. 6, 319.

Arif, M., Chia, L.-S., Pauls, K.P., 2018. Protein-based bioproducts, in: Chen, G., Weselake, R.J., Singer, S.D. (Eds.), Plant Bioproducts. Springer New York, New York, pp. 143–175.

Arisz, S.A., Heo, J.-Y., Koevoets, I.T., Zhao, T., van Egmond, P., Meyer, J., Zeng, W., Niu, X., Wang, B., Mitchell-Olds, T., Schranz, M.E., Testerink, C., 2018. DIACYLGLYCEROL ACYLTRANSFERASE1 contributes to freezing tolerance. Plant Physiol. 177, 1410–1424.

Bahadar, H., Mostafalou, S., Abdollahi, M., 2014. Current understandings and perspectives on

non-cancer health effects of benzene: A global concern. Toxicol. Appl. Pharmacol. 276, 83–94.

Bailey, T.L., Elkan, C., 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc. Second Int. Conf. Intell. Syst. Mol. Biol. 28–36.

Banaś, W., Sanchez Garcia, A., Banaś, A., Stymne, S., 2013. Activities of acyl-CoA:diacylglycerol acyltransferase (DGAT) and phospholipid:diacylglycerol acyltransferase (PDAT) in microsomal preparations of developing sunflower and safflower seeds. Planta 237, 1627–1636.

Barsanti, L., Gualtieri, P., 2014. Algae - Anatomy, biochemistry, and biotechnology. CRC Press, Boca Raton.

Bates, P.D., Johnson, S.R., Cao, X., Li, J., Nam, J.-W., Jaworski, J.G., Ohlrogge, J.B., Browse, J., 2014. Fatty acid synthesis is inhibited by inefficient utilization of unusual fatty acids for glycerolipid assembly. Proc. Natl. Acad. Sci. 111, 1204–1209.

Baumeister, C., Kilian, L., 2016. Forty years of oil price fluctuations: Why the price of oil may still surprise us. J. Econ. Perspect. 30, 139–160.

Benedetti, M., Vecchi, V., Barera, S., Dall'osto, L., 2018. Biomass from microalgae: the potential of domestication towards sustainable biofactories. Microb Cell Fact 17, 173.

Bhattacharya, D., Price, D.C., Chan, C.X., Qiu, H., Rose, N., Ball, S., Weber, A.P., Arias, M.C., Henrissat, B., Coutinho, P.M., Krishnan, A., Zauner, S., Morath, S., Hilliou, F., Egizi, A., Perrineau, M.M., Yoon, H.S., 2013. Genome of the red alga *Porphyridium purpureum*. Nat Commun 4, 1941.

Bielawski, J.P., 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.

Bligh, E.G., Dyer, W.J., 1959. A rapid method of total lipid extraction and purification. Can. J. Biochem. Physiol. 37, 911–917.

Bottu, G., 2009. Sequence databases and database searching, in: Vandamme, A.-M., Salemi, M., Lemey, P. (Eds.), The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. Cambridge University Press, Cambridge, pp. 33–67.

Bovill, J.G., 2008. Inhalation Anaesthesia: From Diethyl Ether to Xenon, in: Schüttler, J., Schwilden, H. (Eds.), Modern Anesthetics. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 121–142.

Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., Berger, F., Adam, C., Aki, S.S., Althoff, F., Araki, T., Arteaga-Vazquez, M.A., Balasubrmanian, S., Barry, K., Bauer, D., Boehm, C.R., Briginshaw, L., Caballero-Perez, J., Catarino, B., Chen, F., Chiyoda, S., Chovatia, M., Davies, K.M., Delmans, M., Demura, T., Dierschke, T., Dolan, L., Dorantes-Acosta, A.E., Eklund, D.M., Florent, S.N., Flores-Sandoval, E., Fujiyama, A., Fukuzawa, H., Galik, B., Grimanelli, D., Grimwood, J., Grossniklaus, U., Hamada, T., Haseloff, J., Hetherington, A.J., Higo, A., Hirakawa, Y., Hundley, H.N., Ikeda, Y., Inoue, K., Inoue, S. ichiro, Ishida, S., Jia, Q., Kakita, M., Kanazawa, T., Kawai, Y., Kawashima, T., Kennedy, M., Kinose, K., Kinoshita, T., Kohara, Y., Koide, E., Komatsu, K., Kopischke, S., Kubo, M., Kyozuka, J., Lagercrantz, U., Lin, S.-S.S., Lindquist, E., Lipzen, A.M., Lu, C.-W.W., Luna, E. De, Martienssen, R.A., Minamino, N., Mizutani, M.M., Mizutani, M.M., Mochizuki, N., Monte, I., Mosher, R., Nagasaki, H., Nakagami, H., Naramoto, S., Nishitani, K., Ohtani, M., Okamoto, T., Okumura, M., Phillips, J., Pollak, B., Reinders, A., Rövekamp, M., Sano, R.,

Sawa, S., Schmid, M.W., Shirakawa, M., Solano, R., Spunde, A., Suetsugu, N., Sugano, S., Sugiyama, A., Sun, R., Suzuki, Y., Takenaka, M., Takezawa, D., Tomogane, H., Tsuzuki, M., Ueda, T., Umeda, M., Ward, J.M., Watanabe, Y., Yazaki, K., Yokoyama, R., Yoshitake, Y., Yotsui, I., Zachgo, S., Schmutz, J., De Luna, E., Martienssen, R.A., Minamino, N., Mizutani, M.M., Mizutani, M.M., Mochizuki, N., Monte, I., Mosher, R., Nagasaki, H., Nakagami, H., Naramoto, S., Nishitani, K., Ohtani, M., Okamoto, T., Okumura, M., Phillips, J., Pollak, B., Reinders, A., Rövekamp, M., Sano, R., Sawa, S., Schmid, M.W., Shirakawa, M., Solano, R., Spunde, A., Suetsugu, N., Sugano, S., Sugiyama, A., Sun, R., Suzuki, Y., Takenaka, M., Takezawa, D., Tomogane, H., Tsuzuki, M., Ueda, T., Umeda, M., Ward, J.M., Watanabe, Y., Yazaki, K., Yokoyama, R., Yoshitake, Y., Yotsui, I., Zachgo, S., Schmutz, J., 2017. Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. Cell 171, 287–304.

Boyle, N.R., Page, M.D., Liu, B., Blaby, I.K., Casero, D., Kropat, J., Cokus, S.J., Hong-Hermesdorf, A., Shaw, J., Karpowicz, S.J., Gallaher, S.D., Johnson, S., Benning, C., Pellegrini, M., Grossman, A., Merchant, S.S., 2012. Three acyltransferases and nitrogen-responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in *Chlamydomonas*. J. Biol. Chem. 287, 15811–15825.

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248–254.

Brodie, J., Chan, C.X., De Clerck, O., Cock, J.M., Coelho, S.M., Gachon, C., Grossman, A.R., Mock, T., Raven, J.A., Smith, A.G., Yoon, H.S., Bhattacharya, D., 2017. The algal revolution. Trends Plant Sci. 22, 726–738.

Calabresi, L., Pisciotta, L., Costantin, A., Frigerio, I., Eberini, I., Alessandrini, P., Arca, M., Bon,

G.B., Boscutti, G., Busnach, G., Frascà, G., Gesualdo, L., Gigante, M., Lupattelli, G., Montali, A., Pizzolitto, S., Rabbone, I., Rolleri, M., Ruotolo, G., Sampietro, T., Sessa, A., Vaudo, G., Cantafora, A., Veglia, F., Calandra, S., Bertolini, S., Franceschini, G., 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.

Caldo, K.M.P., Shen, W., Xu, Y., Hanley-Bowdoin, L., Chen, G., Weselake, R., Lemieux, M.J., 2018. Diacylglycerol acyltransferase 1 is activated by phosphatidate and inhibited by SnRK1-catalyzed phosphorylation. Plant J. 96, 287–299.

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

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973.

Cases, S., Smith, S.J., Zheng, Y.-W., Myers, H.M., Lear, S.R., Sande, E., Novak, S., Collins, C., Welch, C.B., Lusis, A.J., Erickson, S.K., Farese, R. V, 1998. Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. Proc. Natl. Acad. Sci. U. S. A. 95, 13018–13023.

Cases, S., Stone, S.J., Zhou, P., Yen, E., Tow, B., Lardizabal, K.D., Voelker, T., Farese, R. V, 2001. Cloning of DGAT2, a second mammalian diacylglycerol acyltransferase, and related family members. J. Biol. Chem. 276, 38870–38876.

Cavalier-Smith, T., 2018. Kingdom Chromista and its eight phyla: a new synthesis emphasising periplastid protein targeting, cytoskeletal and periplastid evolution, and ancient divergences. Protoplasma 255, 297–357.

Charlton-Menys, V., Pisciotta, L., Durrington, P.N., Neary, R., Short, C.D., Calabresi, L., Calandra, S., Bertolini, S., 2007. Molecular characterization of two patients with severe LCAT deficiency. Nephrol. Dial. Transplant. 22, 2379–2382.

Chen, G., Greer, M.S., Lager, I., Lindberg Yilmaz, J., Mietkiewska, E., Carlsson, A.S., Stymne, S., Weselake, R.J., 2012. Identification and characterization of an LCAT-like *Arabidopsis thaliana* gene encoding a novel phospholipase A. FEBS Lett. 586, 373–377.

Chen, G., Xu, Y., Siloto, R.M.P., Caldo, K.M.P., Vanhercke, T., Tahchy, A. El, Niesner, N., Chen, Y., Mietkiewska, E., Weselake, R.J., 2017. High-performance variants of plant diacylglycerol acyltransferase 1 generated by directed evolution provide insights into structure function. Plant J. 92, 167–177.

Chen, H., Li, T., Wang, Q., 2019. Ten years of algal biofuel and bioproducts: gains and pains. Planta 249, 195–219.

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

Clerck, O. De, Bogaert, K.A., Leliaert, F., 2012. Chapter Two - Diversity and Evolution of Algae: Primary Endosymbiosis, in: Piganeau, G. (Ed.), Genomic Insights into the Biology of Algae, Advances in Botanical Research. Academic Press, pp. 55–86.

Contacos, C., Sullivan, D.R., Rye, K.A., Funke, H., Assmann, G., 1996. A new molecular defect in the lecithin:cholesterol acyltransferase (LCAT) gene associated with fish eye disease. J Lipid Res 37, 35–44.

Dahlqvist, A., Stahl, U., Lenman, M., Banas, A., Lee, M., Sandager, L., Ronne, H., 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. 97,

6487–6492.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat Meth 9, 772.

de Carvalho Lopes, A.P., dos Santos, F.M.L.F., Vilar, V.J.P., Pires, J.C.M., 2018. Process integration applied to microalgal biofuels production, in: Jacob-Lopes, E., Queiroz Zepka, L., Queiroz, M.I. (Eds.), Energy from Microalgae. Springer International Publishing, Cham, pp. 35–57.

De Clerck, O., Bogaert, K.A., Leliaert, F., 2012. Diversity and evolution of algae: Primary endosymbiosis.

de Vries, J., Stanton, A., Archibald, J.M., Gould, S.B., 2016. Streptophyte terrestrialization in light of plastid evolution. Trends Plant Sci. 21, 467–476.

Dembitsky, V.M., 1996. Betaine ether-linked glycerolipids: Chemistry and biology. Prog. Lipid Res. 35, 1–51.

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

Doğan, H., Otu, H.H., 2014. Objective functions, in: Russell, D.J. (Ed.), Multiple Sequence Alignment Methods. Humana Press, Totowa, NJ, pp. 45–58.

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

Du, Z.-Y., Benning, C., 2016. Triacylglycerol accumulation in photosynthetic cells in plants and algae, in: Nakamura, Y., Li-Beisson, Y. (Eds.), Lipids in Plant and Algae Development. Springer International Publishing, Cham, pp. 179–205.

Durrett, T.P., Benning, C., Ohlrogge, J., 2008. Plant triacylglycerols as feedstocks for the production of biofuels. Plant J. 54, 593–607.

Dyer, J.M., Chapital, D.C., Kuan, J.W., Mullen, R.T., Turner, C., McKeon, T.A., Pepperman, A.B., 2002. Molecular analysis of a bifunctional fatty acid conjugase/desaturase from tung. Implications for the evolution of plant fatty acid diversity. Plant Physiol. 130, 2027–2038.

Dyer, J.M., Stymne, S., Green, A.G., Carlsson, A.S., 2008. High-value oils from plants. Plant J. 54, 640–655.

Empresa de Pesquisa Energética - EPE, 2018. Balanço energético nacional 2018: Ano base 2017. Rio de Janeiro.

Fakhri, S., Abbaszadeh, F., Dargahi, L., Jorjani, M., 2018. Astaxanthin: A mechanistic review on its biological activities and health benefits. Pharmacol. Res. 136, 1–20.

Falarz, L.J., Deyholos, M.K., Chen, G., 2018. Materials and related bioproducts from plant carbohydrates, in: Chen, G., Weselake, R.J., Singer, S.D. (Eds.), Plant Bioproducts. Springer New York, New York, pp. 109–120.

Fan, J., Andre, C., Xu, C., 2011. A chloroplast pathway for the de novo biosynthesis of triacylglycerol in *Chlamydomonas reinhardtii*. FEBS Lett. 585, 1985–1991.

Fan, J., Yan, C., Xu, C., 2013a. 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., Xu, C., 2013b. 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, 3506–3518.

Feng, P., Deng, Z., Fan, L., Hu, Z., 2012. Lipid accumulation and growth characteristics of

Chlorella zofingiensis under different nitrate and phosphate concentrations. J. Biosci. Bioeng. 114, 405–410.

Gietz, R.D., Schiestl, R.H., 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. 2, 31–34.

Giroud, C., Gerber, A., Eichenberger, W., 1988. Lipids of *Chlamydomonas reinhardtii*. Analysis of molecular species and intracellular site(s) of biosynthesis. Plant Cell Physiol. 29, 587–595.

Glukhova, A., Hinkovska-Galcheva, V., Kelly, R., Abe, A., Shayman, J.A., Tesmer, J.J., 2015. Structure and function of lysosomal phospholipase A2 and lecithin:cholesterol acyltransferase. Nat. Commun. 6, 6250.

Goold, H., Beisson, F., Peltier, G., Li-Beisson, Y., 2015. Microalgal lipid droplets: composition, diversity, biogenesis and functions. Plant Cell Rep.

Greer, M.S., Pan, X., Weselake, R.J., 2016. Two clades of type-1 *Brassica napus* diacylglycerol acyltransferase exhibit differences in acyl-CoA preference. Lipids 51, 781–786.

Guerin, M., Dachet, C., Goulinet, S., Chevet, D., Dolphin, P.J., Chapman, M.J., Rouis, M., 1997. Familial lecithin:cholesterol acyltransferase deficiency: Molecular analysis of a compound heterozygote. LCAT (Arg^{147} → Trp) and LCAT (Tyr^{171} → Stop). Atherosclerosis 131, 85–95.

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

Guo, X., Fan, C., Chen, Y., Wang, J., Yin, W., Wang, R.R.C., Hu, Z., 2017. Identification and characterization of an efficient acyl-CoA: diacylglycerol acyltransferase 1 (*DGAT1*) gene

from the microalga *Chlorella ellipsoidea*. BMC Plant Biol. 17, 48.

Han, D., Li, Y., Hu, Q., 2013. Astaxanthin in microalgae: Pathways, functions and biotechnological implications. Algae 28, 131–147.

Harmsen, P.F.H., Hackmann, M.M., Bos, H.L., 2014. Green building blocks for bio-based plastics. Biofuels, Bioprod. Biorefining 8, 306–324.

Higgins, D., Lemey, P., 2009. Multiple sequence alignment, in: Vandamme, A.-M., Salemi, M., Lemey, P. (Eds.), The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. Cambridge University Press, Cambridge, pp. 68–108.

Higuera-Ciapara, I., Félix-Valenzuela, L., Goycoolea, F.M., 2006. Astaxanthin: A review of its chemistry and applications. Crit. Rev. Food Sci. Nutr. 46, 185–196.

Holleboom, A.G., Kuivenhoven, J.A., Peelman, F., Schimmel, A.W., Peter, J., Defesche, J.C., Kastelein, J.J.P., Hovingh, G.K., Stroes, E.S., Motazacker, M.M., 2011. High prevalence of mutations in *LCAT* in patients with low HDL cholesterol levels in the Netherlands: Identification and characterization of eight novel mutations. Hum. Mutat. 32, 1290–1298.

Hu, B., Jin, J., Guo, A.-Y., Zhang, H., Luo, J., Gao, G., 2015. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics 31, 1296–1297.

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

Huang, W., Wang, X., Endo-Streeter, S., Barrett, M., Waybright, J., Wohlfeld, C., Hajicek, N., Harden, T.K., Sondek, J., Zhang, Q., 2018. A membrane-associated, fluorogenic reporter for mammalian phospholipase C isozymes. J. Biol. Chem. 293, 1728–1735.

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.

Jackson, C.J., Reyes-Prieto, A., 2014. The mitochondrial genomes of the Glaucophytes *Gloeochaete wittrockiana* and *Cyanoptyche gloeocystis*: Multilocus phylogenetics suggests amonophyletic archaeplastida. Genome Biol. Evol. 6, 2774–2785.

Jeffares, D.C., Mourier, T., Penny, D., 2006. The biology of intron gain and loss. Trends Genet. 22, 16–22.

Johnson, K.A., Goody, R.S., 2011. The original Michaelis constant: Translation of the 1913 Michaelis-Menten paper. Biochemistry 50, 8264–8269.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R., Hunter, S., 2014. InterProScan 5: Genome-scale protein function classification. Bioinformatics 30, 1236–1240.

Jukes, T.H., Cantor, C.R., 1969. Evolution of protein molecules, in: MUNRO, H.N.B.T.-M.P.M. (Ed.), Mammalian Protein Metabolism. Academic Press, New York, pp. 21–132.

Jung, S.H., Kim, R.J., Kim, K.J., Lee, D.H., Suh, M.C., 2019. Plastidial and mitochondrial malonyl CoA-ACP malonyltransferase is essential for cell division and its overexpression increases storage oil content. Plant Cell Physiol. 0, 1–11.

Kass, R., Raftery, A., 1995. Bayes factors. J. Am. Stat. Assoc.

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Kelly, L.A., Mezulis, S., Yates, C., Wass, M., Sternberg, M., 2015. The Phyre2 web portal for

protein modelling, prediction, and analysis. Nat. Protoc. 10, 845–858.

Kim, H.U., Lee, K.R., Go, Y.S., Jung, J.H., Suh, M.C., Kim, J.B., 2011. Endoplasmic reticulum-located PDAT1-2 from castor bean enhances hydroxy fatty acid accumulation in transgenic plants. Plant Cell Physiol. 52, 983–993.

Klein, H.G., Lohse, P., Duverger, N., Albers, J.J., Rader, D.J., Zech, L.A., Santamarina-Fojo, S., 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.

Knothe, G., 2005. Dependence of biodiesel fuel properties on the structure of fatty acid alkyl esters. Fuel Process. Technol. 86, 1059–1070.

Koshland Jr., D.E., 1953. Stereochemistry and the mechanism of enzymatic reactions. Biol. Rev. 28, 416–436.

Krogh, A., Larsson, B., Von Heijne, G., Sonnhammer, E.L.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J. Mol. Biol. 305, 567–580.

Lardizabal, K.D., Mai, J.T., Wagner, N.W., Wyrick, A., Voelker, T., Hawkins, D.J., 2001. DGAT2 is a new diacylglycerol acyltransferase gene family. Purification, cloning, and expression in insect cells of two polypeptides from *Mortierella ramanniana* with diacylglycerol acyltransferase activity. J. Biol. Chem. 276, 38862–38869.

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

Letunic, I., Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and

annotation of phylogenetic and other trees. Nucleic Acids Res. 44, W242–W245.

Li-Beisson, Y., Beisson, F., Riekhof, W., 2015. Metabolism of acyl-lipids in *Chlamydomonas reinhardtii*. Plant J. 82, 504–522.

Li-Beisson, Y., Nakamura, Y., Harwood, J., 2016. Lipids: From Chemical Structures, Biosynthesis, and Analyses to Industrial Applications, in: Nakamura, Y., Li-Beisson, Y. (Eds.), Lipids in Plant and Algae Development. Springer International Publishing, Cham, pp. 1–18.

Li-Beisson, Y., Thelen, J.J., Fedosejevs, E., Harwood, J.L., 2019. The lipid biochemistry of eukaryotic algae. Prog. Lipid Res. 74, 31–68.

Li, N., Xu, C., Li-Beisson, Y., Philippar, K., 2016. Fatty acid and lipid transport in plant cells. Trends Plant Sci. 21, 145–158.

Lineweaver, H., Burk, D., 1934. The determination of enzyme dissociation constants. J. Am. Chem. Soc. 56, 658–666.

Liu, J., Ma, X.-N., Guo, B.-B., Yang, B., Chen, F., Liu, B., 2018. Physiochemical and gene expression analyses reveal differential responses of the marine oleaginous alga *Nannochloropsis salina* under different lipid-induction conditions. J. Appl. Phycol. 30, 909–919.

Liu, M., Subramanian, V.S., Subbaiah, P. V., 1998. Modulation of the positional specificity of lecithin-cholesterol acyltransferase by the acyl group composition of its phosphatidylcholine substrate: Role of the *sn*-1-acyl group. Biochemistry 37, 13626–13633.

Liu, Q., Siloto, R.M.P., Lehner, R., Stone, S.J., Weselake, R.J., 2012. Acyl-CoA:diacylglycerol acyltransferase: Molecular biology, biochemistry and biotechnology. Prog. Lipid Res. 51, 350–377.

Liu, X.Y., Ouyang, L.L., 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.

Loughran, N.B., Hinde, S., McCormick-Hill, S., Leidal, K.G., Bloomberg, S., Loughran, S.T., O'Connor, B., Ó'Fágáin, C., Nauseef, W.M., O'Connell, M.J., 2012. Functional consequence of positive selection revealed through rational mutagenesis of human myeloperoxidase. Mol. Biol. Evol. 29, 2039–2046.

Loughran, N.B., O'Connor, B., Ó'Fágáin, C., O'Connell, M.J., 2008. The phylogeny of the mammalian heme peroxidases and the evolution of their diverse functions. BMC Evol. Biol. 8, 101.

Markou, G., Nerantzis, E., 2013. Microalgae for high-value compounds and biofuels production: A review with focus on cultivation under stress conditions. Biotechnol. Adv.

McFie, P.J., Stone, S.J., 2011. A fluorescent assay to quantitatively measure in vitro acyl CoA:diacylglycerol acyltransferase activity. J. Lipid Res. 52, 1760–1764.

McKeon, T.A., 2016. Castor (*Ricinus communis* L.), in: McKeon, T.A., Hayes, D.G., Hildebrand, D.F., Weselake, R.J. (Eds.), Industrial Oil Crops. Elsevier/AOCS Press, New York/Urbana, pp. 75–112.

Menardo, F., Loiseau, C., Brites, D., Coscolla, M., Gygli, S.M., Rutaihwa, L.K., Trauner, A., Beisel, C., Borrell, S., Gagneux, S., 2018. Treemmer: A tool to reduce large phylogenetic datasets with minimal loss of diversity. BMC Bioinformatics 19, 164.

Meng, M., Bagdasarian, M., 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.

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Maréchal-Drouard, L., Marshall, W.F., Qu, L.H., Nelson, D.R., Sanderfoot, A.A., Spalding, M.H., Kapitonov, V. V, Ren, Q., Ferris, P., Lindquist, E., Shapiro, H., Lucas, S.M., Grimwood, J., Schmutz, J., Grigoriev, I. V, Rokhsar, D.S., Grossman, A.R., Cardol, P., Cerutti, H., Chanfreau, G., Chen, C.L., Cognat, V., Croft, M.T., Dent, R., Dutcher, S., Fernández, E., Fukuzawa, H., González-Ballester, D., González-Halphen, D., Hallmann, A., Hanikenne, M., Hippler, M., Inwood, W., Jabbari, K., Kalanon, M., Kuras, R., Lefebvre, P.A., Lemaire, S.D., Lobanov, A. V, Lohr, M., Manuell, A., Meier, I., Mets, L., Mittag, M., Mittelmeier, T., Moroney, J. V, Moseley, J., Napoli, C., Nedelcu, A.M., Niyogi, K., Novoselov, S. V, Paulsen, I.T., Pazour, G., Purton, S., Ral, J.P., Riaño-Pachón, D.M., Riekhof, W., Rymarquis, L., Schroda, M., Stern, D., Umen, J., Willows, R., Wilson, N., Zimmer, S.L., Allmer, J., Balk, J., Bisova, K., Chen, C.J., Elias, M., Gendler, K., Hauser, C., Lamb, M.R., Ledford, H., Long, J.C., Minagawa, J., Page, M.D., Pan, J., Pootakham, W., Roje, S., Rose, A., Stahlberg, E., Terauchi, A.M., Yang, P., Ball, S., Bowler, C., Dieckmann, C.L., Gladyshev, V.N., Green, P., Jorgensen, R., Mayfield, S., Mueller-Roeber, B., Rajamani, S., Sayre, R.T., Brokstein, P., Dubchak, I., Goodstein, D., Hornick, L., Huang, Y.W., Jhaveri, J., Luo, Y., Martínez, D., Ngau, W.C.A., Otillar, B., Poliakov, A., Porter, A., Szajkowski, L., Werner, G., Zhou, K., 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science (80-.). 318, 245–251.

Mhaske, V., Beldjilali, K., Ohlrogge, J., Pollard, M., 2005. Isolation and characterization of an *Arabidopsis thaliana* knockout line for phospholipid: diacylglycerol transacylase gene (At5g13640). Plant Physiol. Biochem. 43, 413–417.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., 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.

Moreau, R.A., 1989. An evaluation of NBD-phospholipids as substrates for the measurement of phospholipase and lipase activities. Lipids 24, 691–699.

Morris, J.L., Puttick, M.N., Clark, J.W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Yang, Z., Schneider, H., Donoghue, P.C.J., 2018. The timescale of early land plant evolution. Proc. Natl. Acad. Sci. 115, E2274–E2283.

Mutlu, H., Meier, M.A.R., 2010. Castor oil as a renewable resource for the chemical industry. Eur. J. Lipid Sci. Technol. 112, 10–30.

Nalbantoglu, Ö.U., 2014. Dynamic programming, in: Russell, D.J. (Ed.), Multiple Sequence Alignment Methods. Humana Press, Totowa, NJ, pp. 3–27.

Needleman, S.B., Wunsch, C.D., 1970. A general method applicable to search for similarities in amino acid sequence of two proteins. J. Mol. Biol. 48, 443–453.

Nikolau, B.J., Perera, M.A.D.N., Brachova, L., Shanks, B., 2008. Platform biochemicals for a biorenewable chemical industry. Plant J. 54, 536–545.

Notredame, C., Higgins, D.G., Heringa, J., 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302, 205–217.

Ohlrogge, J., Browse, J., 1995. Lipid biosynthesis. Plant Cell 7, 957–970.

Ollis, D.L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S.M., Harel, M., Remington, S.J., Silman, I., Schrag, J., Sussman, J.L., Verschueren, K.H.G., Goldman, A., 1992. The α/β hydrolase fold. Protein Eng. 5, 197–211.

Pan, X., Peng, F.Y., 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., 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., Shiu, S.-H., 2016. Evolution of gene duplication in plants. Plant Physiol. 171, 2294–2316.

Peelman, F., Vanloo, B., Perez-Mendez, O., Decout, A., Verschelde, J.-L., Labeur, C., Vinaimont, N., Verhee, A., Duverger, N., Brasseur, R., Vandekerckhove, J., Tavernier, J., Rosseneu, M., 1999. Characterization of functional residues in the interfacial recognition domain of lecithin cholesterol acyltransferase (LCAT). Protein Eng. 12, 71–78.

Peelman, F., Vinaimont, N., Verhee, A., Vanloo, B., Verschelde, J.L., Labeur, C., Seguret-Mace, S., Duverger, N., Hutchinson, G., Vandekerckhove, J., Tavernier, J., Rosseneu, M., 1998. A proposed architecture for lecithin cholesterol acyl transferase (LCAT): identification of the catalytic triad and molecular modeling. Protein Sci. 7, 587–99.

Piper, D.E., Romanow, W.G., Gunawardane, R.N., Fordstrom, P., Masterman, S., Pan, O., Thibault, S.T., Zhang, R., Meininger, D., Schwarz, M., Wang, Z., King, C., Zhou, M., Walker, N.P.C., 2015. The high-resolution crystal structure of human LCAT. J. Lipid Res. 56, 1711–1719.

Posada, D., 2009. Selecting models of evolution, in: Vandamme, A.-M., Salemi, M., Lemey, P. (Eds.), The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and

Hypothesis Testing. Cambridge University Press, Cambridge, pp. 345–361.

Price, D.C., Chan, C.X., Yoon, H.S., Yang, E.C., Qiu, H., Weber, A.P.M., Schwacke, R., Gross, J., Blouin, N.A., Lane, C., Reyes-Prieto, A., Durnford, D.G., Neilson, J.A.D., Lang, B.F., Burger, G., Steiner, J.M., Löffelhardt, W., Meuser, J.E., Posewitz, M.C., Ball, S., Arias, M.C., Henrissat, B., Coutinho, P.M., Rensing, S.A., Symeonidi, A., Doddapaneni, H., Green, B.R., Rajah, V.D., Boore, J., Bhattacharya, D., 2012. *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. Science (80-.). 335, 843–847.

Punekar, N.S., 2018a. Enzymes: Their place in biology, in: ENZYMES: Catalysis, Kinetics and Mechanisms. Springer Singapore, Singapore, pp. 3–4.

Punekar, N.S., 2018b. Principles of enzyme assays, in: ENZYMES: Catalysis, Kinetics and Mechanisms. Springer Singapore, Singapore, pp. 115–129.

Punekar, N.S., 2018c. Henri--Michaelis--Menten equation, in: ENZYMES: Catalysis, Kinetics and Mechanisms. Springer Singapore, Singapore, pp. 155–176.

Qiagen Bioinformatics, 2017. CLC Sequence Viewer [WWW Document].

Rico, J.A.P., Sauer, I.L., 2015. A review of Brazilian biodiesel experiences. Renew. Sustain. Energy Rev. 45, 513–529.

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

Rogozin, I.B., Carmel, L., Csuros, M., Koonin, E. V., 2012. Origin and evolution of spliceosomal introns. Biol. Direct 7, 11.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: Efficient Bayesian

phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.

Ronquist, F., van der Mark, P., Huelsenbeck, J.P., 2009. Bayesian phylogenetic analysis using MRBAYES, in: Vandamme, A.-M., Salemi, M., Lemey, P. (Eds.), The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. Cambridge University Press, Cambridge, pp. 210–266.

Roth, M.S., Cokus, S.J., Gallaher, S.D., Walter, A., Lopez, D., Erickson, E., Endelman, B., Westcott, D., Larabell, C.A., Merchant, S.S., Pellegrini, M., Niyogi, K.K., 2017. Chromosome-level genome assembly and transcriptome of the green alga Chromochloris zofingiensis illuminates astaxanthin production. Proc. Natl. Acad. Sci. 114, E4296–E4305.

Saha, S., Enugutti, B., Rajakumari, S., Rajasekharan, R., 2006. Cytosolic triacylglycerol biosynthetic pathway in oilseeds. Molecular cloning and expression of peanut cytosolic diacylglycerol acyltransferase. Plant Physiol. 141, 1533–1543.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–25.

Sakurai, T., Sakurai, A., Vaisman, B.L., Nishida, T., Neufeld, E.B., Demosky, S.J., Sampson, M.L., Shamburek, R.D., Freeman, L.A., Remaley, A.T., 2018. Development of a novel fluorescent activity assay for lecithin:cholesterol acyltransferase. Ann. Clin. Biochem. 55, 414–421.

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

Sanderson, M.C., Venable, M.E., 2012. A novel assay of acyl-coa: diacylglycerol acyltransferase

activity utilizing fluorescent substrate. J. Phycol. 48, 580–584.

Schmidt, H.A., 2009. Testing tree topologies, in: Vandamme, A.-M., Salemi, M., Lemey, P. (Eds.), The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. Cambridge University Press, Cambridge, pp. 381–404.

Schmidt, H.A., von Haeseler, A., 2009. Phylogenetic inference using maximum likelihood methods, in: Vandamme, A.-M., Salemi, M., Lemey, P. (Eds.), The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. Cambridge University Press, Cambridge, pp. 181–209.

Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9, 671–675.

Shockey, J.M., Gidda, S.K., Chapital, D.C., Kuan, J.-C., Dhanoa, P.K., Bland, J.M., Rothstein, S.J., Mullen, R.T., Dyer, J.M., 2006. Tung tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum. Plant Cell 18, 2294–2313.

Shoemaker, J.S., Fitch, W.M., 1989. Evidence from nuclear sequences that invariable sites should be considered when sequence divergence is calculated. Mol. Biol. Evol. 6, 270–289.

Show, K.-Y., Yan, Y.-G., Lee, D.-J., 2019. Algal biomass harvesting and drying, in: Pandey, A., Chang, J.-S., Soccol, C.R., Lee, D.-J., Chisti, Y.B.T.-B. from A. (Second E. (Eds.), Biomass, Biofuels, Biochemicals. Elsevier, pp. 135–166.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539.

Siloto, R.M.P., Truksa, M., He, X., McKeon, T., Weselake, R.J., 2009. Simple methods to detect triacylglycerol biosynthesis in a yeast-based recombinant system. Lipids 44, 963–973.

Singer, S.D., Weselake, R.J., 2018a. Production of other bioproducts from plant oils, in: Chen, G., Weselake, R.J., Singer, S.D. (Eds.), Plant Bioproducts. Springer New York, New York, NY, pp. 59–85.

Singer, S.D., Weselake, R.J., 2018b. Production of biodiesel from plant oils, in: Chen, G., Weselake, R.J., Singer, S.D. (Eds.), Plant Bioproducts. Springer New York, New York, NY, pp. 41–58.

Skretting, G., Blomhoff, J.P., Solheim, J., Prydz, H., 1992. The genetic defect of the original Norwegian lecithin:cholesterol acyltransferase deficiency families. FEBS Lett. 309, 307–310.

Smith, T.F., Waterman, M.S., 1981. Identification of common molecular subsequences. J. Mol. Biol. 147, 195–197.

Souza, S.P., Seabra, J.E.A., Nogueira, L.A.H., 2018. Feedstocks for biodiesel production: Brazilian and global perspectives. Biofuels 9, 455–478.

Sovacool, B.K., Saunders, H., 2014. Competing policy packages and the complexity of energy security. Energy 67, 641–651.

Spolaore, P., Joannis-Cassan, C., Duran, E., Isambert, A., 2006. Commercial applications of microalgae. J. Biosci. Bioeng. 101, 87–96.

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

States, D.J., Gish, W., Altschul, S.F., 1991. Improved sensitivity of nucleic acid database

searches using application-specific scoring matrices. Methods 3, 66–70.

Steyrer, E., Haubenwallner, S., Hörl, G., Gießauf, W., Kostner, G.M., Zechner, R., 2004. A single G to A nucleotide transition in exon IV of the lecithin:cholesterol acyltransferase (LCAT) gene results in an Arg^{140} to His substitution and causes LCAT-deficiency. Hum. Genet. 96, 105–109.

Strausberg, R.L., Strausberg, S.L., 1995. Overview of protein expression in *Saccharomyces cerevisiae*. Curr. Protoc. Protein Sci. 5.6.1-5.6.7.

Strimmer, K., von Haeseler, A., 2009. Genetic distances and nucleotide substitution models, in: Vandamme, A.-M., Salemi, M., Lemey, P. (Eds.), The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. Cambridge University Press, Cambridge, pp. 111–141.

Sumner, J., Jarvis, P., Fernandez-Sanchez, J., Kaine, B., Woodhams, M., Holland, B., 2012. Is the general time-reversible model bad for molecular phylogenetics? Syst. Biol. 61, 1069–1074.

Tan, W.-J., Yang, Y.-C., Zhou, Y., Huang, L.-P., Xu, L., Chen, Q.-F., Yu, L.-J., Xiao, S., 2018. DIACYLGLYCEROL ACYLTRANSFERASE and DIACYLGLYCEROL KINASE modulate triacylglycerol and phosphatidic acid production in the plant response to freezing stress. Plant Physiol. 177, 1303–1318.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Tiefenauer, L., Demarche, S., 2012. Challenges in the development of functional assays of membrane proteins. Materials (Basel). 5, 2205–2242.

Vaisman, B.L., Remaley, A.T., 2013. Measurement of lecithin-cholesterol acyltransferase activity with the use of a peptide-proteoliposome substrate, in: Freeman, L.A. (Ed.), Lipoproteins and Cardiovascular Disease: Methods and Protocols. Humana Press, Totowa, NJ, pp. 343–352.

Van de Peer, Y., Salemi, M., 2009. Phylogenetic inference based on distance methods, in: Vandamme, A.-M., Salemi, M., Lemey, P. (Eds.), The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. Cambridge University Press, Cambridge, pp. 142–180.

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

van Erp, H., Shockey, J., Zhang, M., Adhikari, N.D., Browse, J., 2015. Reducing isozyme competition increases target fatty acid accumulation in seed triacylglycerols of transgenic Arabidopsis. Plant Physiol. 168, 36–46.

Vandamme, A.-M., 2009. Basic concepts of molecular evolution, in: Vandamme, A.-M., Salemi, M., Lemey, P. (Eds.), The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. Cambridge University Press, Cambridge, pp. 3–30.

Vieler, A., Wu, G., Tsai, C.H., Bullard, B., Cornish, A.J., Harvey, C., Reca, I.B., Thornburg, C., Achawanantakun, R., Buehl, C.J., Campbell, M.S., Cavalier, D., Childs, K.L., Clark, T.J., Deshpande, R., Erickson, E., Armenia Ferguson, A., Handee, W., Kong, Q., Li, X., Liu, B., Lundback, S., Peng, C., Roston, R.L., Sanjaya, Simpson, J.P., TerBush, A., Warakanont, J., Zäuner, S., Farre, E.M., Hegg, E.L., Jiang, N., Kuo, M.H., Lu, Y., Niyogi, K.K., Ohlrogge, J., Osteryoung, K.W., Shachar-Hill, Y., Sears, B.B., Sun, Y., Takahashi, H., Yandell, M.,

Shiu, S.H., Benning, C., 2012. Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga *Nannochloropsis oceanica* CCMP1779. PLoS Genet. 8, e1003064.

Wang, D., Ning, K., Li, J., Hu, J., Han, D., Wang, H., Zeng, X., Jing, X., Zhou, Q., Su, X., Chang, X., Wang, A., Wang, W., Jia, J., Wei, L., Xin, Y., Qiao, Y., Huang, R., Chen, J., Han, B., Yoon, K., Hill, R.T., Zohar, Y., Chen, F., Hu, Q., Xu, J., 2014. *Nannochloropsis* genomes reveal evolution of microalgal oleaginous traits. PLoS Genet. 10, e1004094.

Wang, J., Gebre, A.K., Anderson, R.A., 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.

Wang, X.L., Osuga, J., Tazoe, F., Okada, Kenta, Nagashima, S., Takahashi, M., Ohshiro, T., Bayasgalan, T., Yagyu, H., Okada, Koji, Ishibashi, S., 2011. Molecular Analysis of a Novel LCAT Mutation (Gly179 → Arg) Found in a Patient with Complete LCAT Deficiency. J. Atheroscler. Thromb. 18, 713–719.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46, W296–W303.

Weadick, C.J., 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.

Weiss, S.B., Kennedy, E.P., 1956. The Enzymatic Synthesis of Triglycerides. J. Am. Chem. Soc. 78, 3550.

Weiss, S.B., Kennedy, E.P., Kiyasu, J.Y., 1960. The Enzymatic Synthesis of Triglycerides. J. Biol. Chem. 235, 40–44.

Wernersson, R., Pedersen, A.G., 2003. RevTrans: Multiple alignment of coding DNA from aligned amino acid sequences. Nucleic Acids Res. 31, 3537–3539.

Weselake, R.J., Chen, G., Singer, S.D., 2018. Building a case for plant bioproducts, in: Chen, G., Weselake, R.J., Singer, S.D. (Eds.), Plant Bioproducts. Springer New York, New York, NY, NY, pp. 1–8.

Weselake, R.J., Shah, S., Tang, M., Quant, P.A., Snyder, C.L., Furukawa-Stoffer, T.L., Zhu, W., Taylor, D.C., Zou, J., Kumar, A., Hall, L., Laroche, A., Rakow, G., Raney, P., Moloney, M.M., Harwood, J.L., 2008. Metabolic control analysis is helpful for informed genetic manipulation of oilseed rape (*Brassica napus*) to increase seed oil content. J. Exp. Bot. 59, 3543–3549.

Weselake, R.J., Woodfield, H.K., Field, C.J., Harwood, J.L., 2017. Production of edible oils through metabolic engineering, in: Akoh, C.C. (Ed.), Food Lipids: Chemistry, Nutrition, and Biotechnology, Fourth Edition. CRC Press, Boca Raton, pp. 973–995.

Worden, A.Z., Lee, J., Mock, T., Rouzé, P., Simmons, M.P., Aerts, A.L., Allen, A.E., Cuvelier, M.L., Derelle, E., Everett, M. V, Foulon, E., Grimwood, J., Gundlach, H., Henrissat, B., Napoli, C., McDonald, S.M., Parker, M.S., Rombauts, S., Salamov, A., Dassow, P. Von, Badger, J.H., Coutinho, P.M., Demir, E., Dubchak, I., Gentemann, C., Eikrem, W., Gready, J.E., John, U., Lanier, W., Lindquist, E.A., Lucas, S., Mayer, K.F.X., Moreau, H., Not, F., Otillar, R., Panaud, O., Pangilinan, J., Paulsen, I., Piegu, B., Poliakov, A., Robbens, S., Schmutz, J., Toulza, E., Wyss, T., Zelensky, A., Zhou, K., Armbrust, E.V., Bhattacharya, D., Goodenough, U.W., Peer, Y. Van de, Grigoriev, I. V., 2009. Green evolution and

dynamic adaptations revealed by genomes of the marine picoeukaryotes Micromonas. Science (80-.). 324, 268–272.

Xie, W., Lewis, P.O., Fan, Y., Kuo, L., Chen, M.-H., 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Syst. Biol. 60, 150–160.

Xie, Y., Han, X., Miao, Y., 2018. An Effective Recombinant Protein Expression and Purification System in *Saccharomyces cerevisiae*. Curr. Protoc. Mol. Biol. 123, e62.

Xu, C., Andre, C., Fan, J., Shanklin, J., 2016. Cellular organization of triacylglycerol biosynthesis in microalgae, in: Nakamura, Y., Li-Beisson, Y. (Eds.), Lipids in Plant and Algae Development. Springer International Publishing, Cham, pp. 207–221.

Xu, C., Fan, J., Riekhof, W., Froehlich, J.E., Benning, C., 2003. A permease-like protein involved in ER to thylakoid lipid transfer in *Arabidopsis*. EMBO J. 22, 2370–2379.

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

Xu, Y., Chen, G., Greer, M.S., Caldo, K.M.P., Ramakrishnan, G., Shah, S., Wu, L., Lemieux, M.J., Ozga, J., Weselake, R.J., 2017. Multiple mechanisms contribute to increased neutral lipid accumulation in yeast producing recombinant variants of plant diacylglycerol acyltransferase 1. J. Biol. Chem. 292, 17819–17831.

Xu, Y., Holic, R., Li, D., Pan, X., Mietkiewska, E., Chen, G., Ozga, J., Weselake, R.J., 2018b. Substrate preferences of long-chain acyl-CoA synthetase and diacylglycerol acyltransferase contribute to enrichment of flax seed oil with α-linolenic acid. Biochem. J. 475, 1473–1489.

Xue, J., Niu, Y.F., Huang, T., Yang, W.D., Liu, J.S., Li, H.Y., 2015. Genetic improvement of the

microalga *Phaeodactylum tricornutum* for boosting neutral lipid accumulation. Metab. Eng. 27, 1–9.

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

Yang, Z., 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. J. Mol. Evol. 39, 306–314.

Yang, Z., 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., Pedersen, A.-M.K., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155, 431–449.

Yang, Z., Rannala, B., 2012. Molecular phylogenetics: principles and practice. Nat. Rev. Genet. 13, 303–314.

Yang, Z., Wong, W.S.W., 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., 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.

Yuan, L., Mao, X., Zhao, K., Ji, X., Ji, C., Xue, J., Li, R., 2017. Characterisation of phospholipid: diacylglycerol acyltransferases (PDATs) from *Camelina sativa* and their roles in stress responses. Biol. Open 6, 1024–1034.

Zhang, J., Nielsen, R., 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., Taylor, D.C., 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.

Zharkikh, A., Li, W.-H., 1992. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. I. Four taxa with a molecular clock. Mol. Biol. Evol. 9, 1119–1147.

Zulu, N.N., Zienkiewicz, K., Vollheyde, K., Feussner, I., 2018. Current trends to comprehend lipid metabolism in diatoms. Prog. Lipid Res. 70, 1–16.

APPENDIX A – Chapter 3 supplemental tables

Table A.1. List of genomes used in this study and their sources

Algal genomes		
Species	**Strain**	**Source**
Asterochloris sp.	Cgr/DA1pho	Genome Portal[1]
Aurantiochytrium limacinum	ATCC MYA-1381	Genome Portal
Aureococcus anophagefferens	CCMP1984	Genome Portal
Auxenochlorella protothecoides	0710	NCBI Genome[2]
Bathycoccus prasinos	RCC1105	https://bioinformatics.psb.ugent.be/gdb/bathycoccus/RELEASE_15jul2011/
Bigelowiella natans	CCMP2755	Genome Portal
Chlamydomonas reinhardtii	CC-503 MT+	Phytozome[3]
Chlorella sp.	NC64A	Genome Portal
Chondrus crispus	Stackhouse	Ensembl Plants[4]
Chromochloris zofingiensis	SAG 211-14	Phytozome
Chrysochromulina sp.	CCMP291	NCBI Genome
Coccomyxa subellipsoidea	C-169	Phytozome
Cyanidioschyzon merolae	10D	Ensembl Plants
Cyanophora paradoxa	CCMP329	http://cyanophora.rutgers.edu/cyanophora/
Dunaliella salina	CCAP19-18	Phytozome
Ectocarpus siliculosus	Ec 32	https://bioinformatics.psb.ugent.be/gdb/ectocarpus/
Emiliania huxleyi	CCMP1516	Genome Portal
Fragilariopsis cylindrus	CCMP1102	Genome Portal
Galdieria sulphuraria	074W	Ensembl Plants
Gonium pectorale	NIES-2863	NCBI Genome
Guillardia theta	CCMP2712	Genome Portal
Micromonas pusilla	CCMP1545	Phytozome
Micromonas sp.	RCC299	Phytozome
Monoraphidium neglectum	SAG 48.87	NCBI Genome
Nannochloropsis gaditana	CCMP526	Ensembl Protists[5]

Nannochloropsis oceanica	CCMP1779	Genome Portal
Ochromonadaceae sp.	CCMP2298	Genome Portal
Ostreococcus lucimarinus	CCMP2972	Phytozome
Ostreococcus tauri	RCC4221	Genome Portal
Pavlovales sp.	CCMP2436	Genome Portal
Pelagophyceae sp.	CCMP2097	Genome Portal
Phaeodactylum tricornutum	CCAP1055/1	Genome Portal
Porphyridium purpureum	CCMP1328	http://cyanophora.rutgers.edu/porphyridium/
Pseudo-nitzschia multiseries	CLN-47	Genome Portal
Schizochytrium aggregatum	ATCC 28209	Genome Portal
Symbiochloris reticulata	Spain	Genome Portal
Thalassiosira oceanica	(Hustedt) Hasle et Heimdal CCMP1005	Ensembl Protists
Thalassiosira pseudonana	CCMP1355	Genome Portal
Volvox carteri f. nagariensis, Eve	subclone of HK10	Phytozome
Plant genomes		
Species	**Cultivar**	**Source**
Arabidopsis thaliana	Columbia	Phytozome
Brassica napus	Darmor-bzh	Ensembl Plants
Brassica oleracea capitata	02–12	Phytozome
Brassica rapa	FPsc	Phytozome
Glycine max	Williams 82	Phytozome
Gossypium raimondii		Phytozome
Linum usitatissimum	CDC Bethune	Phytozome
Marchantia polymorpha	Tak-1 + Tak-2	Phytozome
Medicago truncatula	A17	Phytozome
Oryza sativa	Nipponbare	Phytozome
Physcomitrella patens	Gransden 2004	Phytozome
Populus trichocarpa	Nisqually 1	Phytozome
Ricinus communis	Hale	Phytozome
Selaginella moellendorffii		Phytozome
Solanum lycopersicum	Heinz 1706	Phytozome
Theobroma cacao	Matina 1-6	Phytozome

Zea mays	B73	Phytozome
Animal genomes		
Species	**Strain/breed**	**Source**
Amphimedon queenslandica		Ensembl Metazoa[6]
Apis mellifera	DH4	Ensembl Metazoa
Caenorhabidtis elegans	Bristol N2	Ensembl Metazoa
Danio rerio	Tuebingen	Ensembl US West[7]
Drosophila melanogaster	iso-1	Ensembl Metazoa
Equus caballus	Female thoroughbred	Ensembl US West
Gallus gallus	Female inbred line UCD001	Ensembl US West
Homo sapiens		https://www.ncbi.nlm.nih.gov/projects/genome/guide/human/
Mus musculus	C57BL/6J	Ensembl US West
Xenopus tropicalis	Nigerian, partially inbred F6	Ensembl US West
Fungi genomes		
Species	**Strain/breed**	**Source**
Aspergillus nidulans	FGSC A4	Ensembl Fungi[8]
Saccharomyces cerevisiae	S288C	Ensembl Fungi

[1] https://genome.jgi.doe.gov/portal/
[2] https://www.ncbi.nlm.nih.gov/genome
[3] https://phytozome.jgi.doe.gov/pz/portal.html
[4] http://plants.ensembl.org/index.html
[5] http://protists.ensembl.org/index.html
[6] http://metazoa.ensembl.org/index.html
[7] http://uswest.ensembl.org/index.html
[8] http://fungi.ensembl.org/index.html

Table A.2. List of the protein identifiers and acronyms used in this study. The protein identifiers are the same used in the original genome annotation.

Algal sequences		
Species	**Acronym**	**Protein identifier**
Astherochloris sp.	AspPDATa	Astpho2|2292
	AspLCATa	Astpho2|1746
Aurantiochytrium limacinum	AliLCATa	Aurli1|137802
Bathycoccus prasinos	BprPDATa	Bathy10g01290
Bigelowiella natans	BnaPDATa	Bigna1|56399
Chlamydomonas reinhardtii	CrPDAT	Cre02.g106400.t1.1
Coccomyxa subellipsoidea	CsuPDATa	37169
Emiliania huxleyi	EhuLCATa	EOD13534
	EhuLCATb	EOD16315
	EhuLCATc	EOD27528
Fragilariopsis cylindrus	FcyPSATa	Fracy1|189188
	FcyPDATa	Fracy1|166543
	FcyLCATa	Fracy1|244100
	FcyLCATb	Fracy1|245361
	FcyLCATc	Fracy1|235453
	FcyLCATd	Fracy1|238395
Gonium pectorale	GpePDATa	KXZ46616.1
Guillardia theta	GthPDATa	Guith1|159571
Micromonas pusilla	MpuPDATa	58904
Micromonas sp. RCC299	MspPDATa	58725
Pavlovales sp.	PavPDATa	Pavlov2436_1|1067693
Pelagophyceae sp.	PelPSATa	Pelago2097_1|586287
Phaeodactylum tricornutum	PtrPSATa	Phatr2|49702
Porphyridium purpureum	PpuPDATa	evm.model.contig_3587.9
	PpuPDATb	evm.model.contig_3587.12
Pseudo-nitzschia multiseries	PmuPDATa	Psemu1|239257
	PmuPSATa	Psemu1|258595
	PmuLCATa	Psemu1|321250

	PmuLCATb	Psemu1|288981
Schizochytrium aggregatum	SagLCATa	Schag1|101408
Symbiochloris reticulata	SrePDATa	Dicre1|415299
	SrePDATb	Dicre1|787468
Thalassiosira oceanica	TocPSATa	EJK45976
	TocPDATa	EJK56618
Thalassiosira pseudonana	TpsPSATa	Thaps3|24413
	TpsLCATa	Thaps3|8051
Volvox carteri	VcaPDATa	Vocar.0027s0118.1.p
Plant sequences		
Species	**Acronym**	**Protein identifier**
Arabidopsis thaliana	AtLCAT1	AT1G27480
	AtPDAT1	AT5G13640
	AtPDAT2	AT3G44830
	AtPSAT1	AT1G04010
Brassica napus	BnpPDATa	CDY09815
	BnpPDATb	CDX85636
	BnpPDATc	CDY65993
	BnpPDATd	CDX91106
	BnpPSATa	CDX89848
	BnpPSATb	CDY10085
	BnpPSATc	CDY18201
	BnpPSATd	CDY32523
Brassica oleracea capitata	BolPDATa	Bol043418
Brassica rapa	BraPSATa	Brara.J00227.1.p
	BraPDATa	Brara.K00401.1.p
Glycine max	GmaPDATa	Glyma.07G036400
	GmaPDATb	Glyma.11G190400
	GmaPDATc	Glyma.13G108100
	GmaPDATd	Glyma.16G005800
	GmaPDATe	Glyma.17G051300
Gossipium raimondii	GraLCATa	Gorai.009G221300
	GraPDATa	Gorai.001G118800

	GraPDATb	Gorai.004G284600
	GraPDATc	Gorai.006G095300
	GraPDATd	Gorai.013G258900
	GraPDATe	Gorai.010G086400
	GraPDATf	Gorai.010G234300
Linum usitatissimum	LusLCATa	Lus1007728.g
	LusLCATb	Lus10018664.g
	LusPDATa	Lus10037657.g
	LusPDATb	Lus10015639.g
	LusPDATc	Lus10017165.g
	LusPDATd	Lus10043360.g
	LusPDATe	Lus10019519.g
	LusPDATf	Lus10021564.g
Marchantia polymorpha	MpoPDATa	Mapoly0008s0119.1.p
	MpoPSATa	Mapoly0023s0070.1.p
Medicago truncatula	MtrPLAa	Medtr4g083980
	MtrLCATa	Medtr7g080450
	MtrPSATa	Medtr7g098010
Oryza sativa	OsaLCATa	LOC_Os01g71800
	OsaLCATb	LOC_Os03g52010
	OsaPDATa	LOC_OS09g27210
Physcomitrella patens	PpaPDATa	Pp3c22_13120
	PpaPLAa	Pp3c2_20030
	PpaPLAb	Pp3c5_28340
Populus trichocarpa	PtcPDATa	Potri.001G171000
	PtcPDATb	Potri.003G063100
	PtcPDATc	Potri.004G190100
	PtcLCATa	Potri.014G014400
Ricinus communis	RcoPDATa	29991.m000626
	RcoPDATb	29706.m001305
	RcoPDATc	29912.m005286
	RcoPSATa	29637.m000766
Solanum lycopersicum	SlyLCATa	Solyc05g050710.2

	SlyPDATa	Solyc03g121960.2
	SlyPDATb	Solyc11g066710.1
	SlyPDATc	Solyc06g062870.2
	SlyPDATd	Solyc06g074680.2
	SlyPDATe	Solyc07g041210.2
Selaginella moellendorffii	SmoPDATa	164869
	SmoPSATa	172410
	SmoLCATa	85956
Theobroma cacao	TcaLCATa	Thecc1EG034671
	TcaPDATa	Thecc1EG029105
	TcaPDATb	Thecc1EG041539
	TcaPDATc	Thecc1EG037574
Zea mays	ZmaPDATa	GRMZM2G061885
	ZmaPDATb	GRMZM2G095763
Animal sequences		
Species	**Acronym**	**Protein identifier**
Amphimedon queenslandica	AquLCATa	Aqu2.1.40890_001
	AquLCATb	Aqu2.1.40889_001
	AquLCATc	Aqu2.1.40891_001
	AquPLA2a	Aqu2.1.40892_001
	AquPLA2b	Aqu2.1.37964_001
	AquPLA2c	Aqu2.1.37965_001
Danio rerio	DrePLA2a	ENSDARP00000131824.1
	DreLCATa	ENSDARP00000132129.1
Drosophila melanogaster	DmePLA2a	FBpp0080897
Equus caballus	EcaPLA2a	ENSECAP00000012423.1
	EcaLCATa	ENSECAP00000020026.1
Gallus gallus	GgaPLA2a	ENSGALP00000056433.1
	GgaLCATa	ENSGALP00000043323.2
GRCh38 (*Homo sapiens*)	HsaPLA2a	NP_036452.1
	HsaLCAT	NP_000220.1
Caenorhabditis elegans	CelPLA2a	M05B5.4
Mus musculus	MmuLCATa	ENSMUSP00000038232.7

	MmuPLA2a	ENSMUSP00000034377.6
Xenopus tropicalis	XtrLCATa	ENSXETP00000006701.1
	XtrPLA2a	ENSXETP00000055403.2
	XtrPLA2b	ENSXETP00000012787.3
Fungi sequences		
Species	**Acronym**	**Protein identifier**
Aspergillus nidulans	AniPDATa	CBF76326
Saccharomyces cerevisiae	ScePDATa	YNR008W_mRNA

Table A.3. Primers used to perform the site-directed mutagenesis

ID	**Primer sequence**
AtPDAT1_M255W_forward	5' – TTC CGC ATT CCT GGG GAG TCT TGT ATT – 3'
AtPDAT1_M255W_reverse	5' – AAT ACA AGA CTC CCC AGG AAT GCG GAA – 3'
AtPDAT1_M255A_forward	5' – TCC GCA TTC AGC AGG AGT CTT GT – 3'
AtPDAT1_M255A_reverse	5' – ACA AGA CTC CTG CTG AAT GCG GA – 3'

Table A.4. Potential PDAT homologs discarded from further analyses due to putative annotation errors

Species	**Protein ID**	**Annotation error**
Aureococcus anophagefferens	Auran1|23177	Unusual start codon
	Auran1|1287	Unusual start codon
	Auran1|70709	Not classified in the LCAT family
Brassica napus	CDX77836	Shorter than 200 amino acids
	CDX77835	Truncated
	CDY71286	Truncated
Brassica oleracea capitata	Bol024121	Redundant
	Bol040863	Redundant
	Bol018368	Redundant
	Bol024120	Shorter than 200 amino acids
Brassica rapa	Brara.I05515.1.p	Redundant
Chromochloris zofingiensis	Cz10g07200.t1	Shorter than 200 amino acids
	Cz10g07180.t1	Truncated
	Cz10g07210.t1	Truncated
Cyanophora paradoxa	ConsensusfromContig55694-snap_masked-ConsensusfromContig55694-abinit-gene-0.0-mRNA-1:cds:11476_256-419:1:-	Shorter than 200 amino acids
	ConsensusfromContig10545-snap_masked-ConsensusfromContig10545-abinit-gene-0.5-mRNA-1:cds:3530_416-440:0:-	Truncated
Dunaliella salina	Dusal.0002s00083.1.p	Unusual start codon
Emiliania huxleyi	EOD03641	Shorter than 200 amino acids
Fragilariopsis cylindrus	Fracy1|203865	Unusual start codon
Glycine max	Glyma.12G083900	Not classified in the LCAT family
Medicago truncatula	Medtr8g104960	Shorter than 200 amino acids

	Medtr6g045280	Not classified in the LCAT family
Monoraphidium neglectum	XP_013903842.1	Truncated
Nannochloropsis oceanica	Nanoce1779|11464	Truncated
	Nanoce1779|1202	Truncated
Ostreococcus lucimarinus	1469	Unusual start codon
Ostreococcus tauri	Ostta4|994	Unusual start codon
Pavlovales sp.	Pavlov2436_1|865042	Shorter than 200 amino acids
Pelagophyceae sp.	Pelago2097_1|557834	Causing long branch attraction
Phaeodactylum tricornutum	Phatr2|8860	Unusual start codon
Populus trichocarpa	Potri.009G150800	Shorter than 200 amino acids
Symbiochloris reticulata	Dicre1|845548	Truncated
Thalassiosira pseudonana	Thaps3|261132	Unusual start codon
Volvox carteri	Vocar.0027s0117.1.p	Not classified in the LCAT family

Table A.5. Properties of individual sequences of LCAT-family proteins used in this study. Sequences were classified into PDAT, PSAT, PLA or LCAT-related based on PANTHER implemented in InterProScan v5.22.61.

Algal PSAT[1]				
Identifier	**Length (aa)**	**MW[2] (kDa)**	**pI[3]**	**# TMD[4]**
TocPSATa	714	79.67	5.20	0
TpsPSATa	1174	130.87	5.20	1
FcyPSATa	797	90.30	5.25	0
PmuPSATa	977	110.37	5.38	0
PtrPSATa	960	108.29	5.49	0
PelPSATa	1328	140.89	5.32	1
Plant PSAT[1]				
Identifier	**Length (aa)**	**MW[2] (kDa)**	**pI[3]**	**# TMD[4]**
AtPSAT1	633	71.69	6.17	1
BraPSATa	641	72.20	5.88	0
BnpPSATa	642	72.28	5.88	0
BnpPSATc	641	72.29	5.94	1
BnpPSATd	641	72.52	6.11	1
BnpPSATb	641	72.54	6.05	0
MtrPSATa	632	71.99	5.73	0
RcoPSATa	638	72.34	6.01	0
MpoPSATa	706	80.01	6.71	1
SmoPSATa	619	69.74	5.88	0
Algal LCAT[5]				
Identifier	**Length (aa)**	**MW[2] (kDa)**	**pI[3]**	**# TMD[4]**
AspLCATa	892	98.67	5.85	5
Plant PLA[6]				
Identifier	**Length (aa)**	**MW[2] (kDa)**	**pI[3]**	**# TMD[4]**
MtrPLAa	538	61.04	5.85	0
PpaPLAa	577	65.40	5.31	0
PpaPLAb	540	61.60	5.31	0
MtrPLAb	447	50.26	5.96	0

Algal LCAT-related[5]				
Identifier	**Length (aa)**	**MW[2] (kDa)**	**pI[3]**	**# TMD[4]**
AliLCATa	721	80.57	8.66	0
SagLCATa	1589	176.55	8.97	0
FcyLCATc	499	55.89	6.76	0
FcyLCATb	705	79.33	8.64	0
PmuLCATb	1113	122.14	6.33	0
TpsLCATa	777	85.85	9.17	0
FcyLCATd	1055	117.07	5.36	0
FcyLCATa	916	101.71	8.32	0
PmuLCATa	895	100.09	5.69	0
EhuLCATb	1085	118.38	5.36	0
EhuLCATa	1068	116.77	5.93	0
EhuLCATc	288	31.40	9.77	0
Plant LCAT-related[5]				
Identifier	**Length (aa)**	**MW[2] (kDa)**	**pI[3]**	**# TMD[4]**
AtLCAT1	432	48.31	8.88	1
GraLCATa	430	47.20	8.68	0
TcaLCATa	428	47.10	5.60	0
PtcLCATa	432	48.28	6.80	0
SlyLCATa	416	46.21	6.19	0
LusLCATa	439	47.76	5.68	1
LusLCATb	440	47.89	5.49	1
OsaLCATb	465	50.03	6.39	0
SmoLCATa	420	46.46	8.27	0
OsaLCATa	363	39.60	5.30	0
Red algal/Chromist PDAT[7]				
Identifier	**Length (aa)**	**MW[2] (kDa)**	**pI[3]**	**# TMD[4]**
PpuPDATb	643	71.80	5.83	1
PpuPDATa	768	86.00	5.96	1
BnaPDATa	524	58.41	5.69	0
GthPDATa	551	62.03	6.21	0
PavPDATa	772	82.49	5.72	1

TocPDATa	730	81.04	6.23	0
FcyPDATa	571	64.13	5.84	0
PmuPDATa	569	63.75	5.96	0
Green algal PDAT[7]				
Identifier	**Length (aa)**	**MW[2] (kDa)**	**pI[3]**	**# TMD[4]**
AspPDATa	485	54.30	6.19	0
SrePDATa	712	80.30	6.63	1
CsuPDATa	510	57.05	6.33	0
SrePDATb	631	69.87	6.55	1
CrPDAT	1041	104.65	6.20	1
VcaPDATa	1143	117.78	8.86	1
GpePDATa	982	99.68	5.35	0
MpuPDATa	686	75.61	9.61	1
MspPDATa	744	79.93	8.67	1
BprPDATa	748	83.83	9.27	1
Plant PDAT[7] – Group A				
Identifier	**Length (aa)**	**MW[2] (kDa)**	**pI[3]**	**# TMD[4]**
AtPDAT1	671	74.16	6.75	1
BolPDATa	667	73.83	6.32	1
BnpPDATa	597	65.95	6.17	0
BnpPDATc	597	65.81	6.20	0
BraPDATa	515	56.64	6.30	0
BnpPDATb	665	73.61	6.75	1
BnpPDATd	597	65.73	6.84	0
GmaPDATa	676	75.57	6.52	1
GmaPDATd	668	74.56	6.52	1
LusPDATb	715	78.92	7.02	1
LusPDATa	734	81.38	6.38	1
PtcPDATa	672	74.92	6.07	1
PtcPDATb	670	74.69	6.74	1
RcoPDATb	685	76.55	6.00	1
GraPDATc	697	77.38	6.10	1
GraPDATf	716	80.28	8.92	1

TcaPDATb	685	76.10	6.35	0
SlyPDATa	668	74.28	6.86	1
SlyPDATc	663	74.05	7.33	1
GmaPDATc	668	74.78	8.79	1
GmaPDATe	668	74.83	8.79	1
GraPDATb	690	76.89	8.50	1
GraPDATe	672	74.56	6.62	1
TcaPDATc	670	74.15	8.62	1
RcoPDATc	660	73.40	7.66	1
LusPDATc	695	76.82	8.56	1
LusPDATf	695	76.96	8.40	1
SlyPDATe	695	77.53	8.87	1
OsaPDATa	691	76.14	6.68	1
ZmaPDATa	678	75.15	6.43	1
ZmaPDATb	676	74.79	6.46	1
MpoPDATa	735	81.99	6.60	1
PpaPDATa	693	76.85	6.75	1
SmoPDATa	647	71.75	6.45	1
Plant PDAT[7] – Group B				
Identifier	**Length (aa)**	**MW[2] (kDa)**	**pI[3]**	**# TMD[4]**
AtPDAT2	665	73.65	8.90	1
LusPDATe	685	75.49	8.35	0
LusPDATd	429	46.80	5.73	0
GmaPDATb	582	65.78	6.15	1
GraPDATa	691	77.34	8.55	0
TcaPDATa	691	77.60	9.05	1
GraPDATd	677	75.72	9.09	0
PtcPDATc	680	75.95	7.94	1
RcoPDATa	612	68.17	8.63	1
SlyPDATd	709	79.28	8.57	1
SlyPDATb	710	79.44	8.96	0

[1]PSAT: Phospholipid:sterol O-acyltransferase; [2]MW: Molecular weight; [3]pI: Isoelectric point;
[4]TMD: Transmembrane domain; [5]LCAT: Lecithin:cholesterol acyltransferase;
[6]PLA: Phospholipase A_1; [7]PDAT: Phospholipid:diacylglycerol acyltransferase

Table A.6. Amino acids under positive selection identified in the Bayes empirical Bayes analyses comparing green algal and plant (group A) PDATs. The most conserved amino acids in each site are shown as well as the type of side-chain in that residue. The position of each site in the *Chlamydomonas reinhardtii* PDAT and *Arabidopsis thaliana* PDAT1 is also shown.

Amino acids under positive selection in green algae					
Amino acids in green algal PDAT		**Amino acids in plants (group A)**		**Sequence position**	
Amino acid	**Type of side-chain**	**Amino acid**	**Type of side-chain**	**CrPDAT**	**AtPDAT1**
T, A, V	Hydroxyl (T); Aliphatic (A,V)	F	Aromatic	291	143
A	Aliphatic	N	Carboxamide	345	197
W, Y	Aromatic	M	Sulfur	402	255
D, E	Acidic	V	Aliphatic	404	257
R	Basic	L	Aliphatic	408	261
V	Aliphatic	F	Aromatic	489	330
G	Aliphatic	P	Cyclic	716	475
A	Aliphatic	Y	Aromatic	728	487
F, Y	Aromatic	W	Aromatic	798	504
H	Basic	I, V	Aliphatic	829	535
L, V	Aliphatic	K	Basic	831	537
I	Aliphatic	W	Aromatic	1029	663
Amino acids under positive selection in plants (group A)					
Amino acids in green algal PDAT		**Amino acids in plants (group A)**		**Sequence position**	
Amino acid	**Type of side-chain**	**Amino acid**	**Type of side-chain**	**CrPDAT**	**AtPDAT1**
G	Aliphatic	P	Cyclic	238	92
Q	Carboxamide	V	Aliphatic	243	97
V	Aliphatic	K	Basic	244	98
S	Hydroxyl	G	Aliphatic	265	119

D, N	Acidic (D); Carboxamide (N)	R	Basic	297	149
G, A, E	Aliphatic (G, A); Acidic (E)	S	Hydroxyl	306	158
P	Cyclic	N	Carboxamide	309	161
E	Acidic	S	Hydroxyl	324	176
I, M	Aliphatic (I), Sulfur (M)	A	Aliphatic	333	185
K	Basic	V	Aliphatic	341	193
D	Acidic	R	Basic	348	200
S, T	Hydroxyl	E	Acidic	353	205
G, A, S	Aliphatic (G, A); Hydroxyl (S)	M	Sulfur	358	210
L, M	Aliphatic (L), Sulfur (M)	T	Hydroxyl	370	222
E, Q	Acidic (E), Carboxamide (Q)	V	Aliphatic	372	224
G	Aliphatic	Q	Carboxamide	375	227
Y	Aromatic	T	Hydroxyl	376	228
L	Aliphatic	T	Hydroxyl	390	242
E	Acidic	K, N	Basic (K), Carboxamide (N)	393	246
V, L	Aliphatic	Y	Aromatic	406	259
A	Aliphatic	K	Basic	447	287
T, M	Hydroxyl (T), Sulfur (M)	A	Aliphatic	471	311
L	Aliphatic	A	Aliphatic	477	317
P	Cyclic	G, R	Aliphatic (G), Basic (R)	490	331
A, G	Aliphatic	Q	Carboxamide	492	336
T, E	Hydroxyl (T), Acidic (E)	H	Basic	493	337
F, W	Aromatic	S, T	Hydroxyl	498	342

R, D, E	Basic (R), Acidic (D, E)	T	Hydroxyl	514	358
M	Sulfur	C, P	Sulfur (C), Cyclic (P)	529	373
A, D, E	Aliphatic (A), Acidic (D, E)	R	Basic	538	402
I, V	Aliphatic	A	Aliphatic	704	463
H	Basic	I	Aliphatic	730	489
A, S, K	Aliphatic (A), Hydroxyl (S), Basic (K)	L	Aliphatic	747	493
T, K, E	Hydroxyl (T), Basic (K), Acidic (E)	D, E	Acidic	837	543
A, M, K	Aliphatic (A), Sulfur (M), Basic (K)	Q	Carboxamide	918	549
S, D, E	Hydroxyl (S), Acidic (D, E)	S	Hydroxyl	922	553
L	Aliphatic	V	Aliphatic	945	578
L	Aliphatic	A	Aliphatic	948	581
V	Aliphatic	Y	Aromatic	970	603
R, K, E	Basic (R, K), Acidic (E)	A, S	Aliphatic (A), Hydroxyl (S)	977	610
V	Aliphatic	P	Cyclic	979	612
H, R	Basic	G, A	Aliphatic	1014	646

APPENDIX B – Chapter 3 supplemental figures

Figure B.1. Sequence alignment of selected PDATs with human LCAT (HsaLCAT). The alignment conservation is illustrated at the top. The "lid" region, salt bridge and catalytic triad are relevant domains of HsaLCAT structure shown in the figure. The location of these important domains is shown as red lines below the alignment conservation.

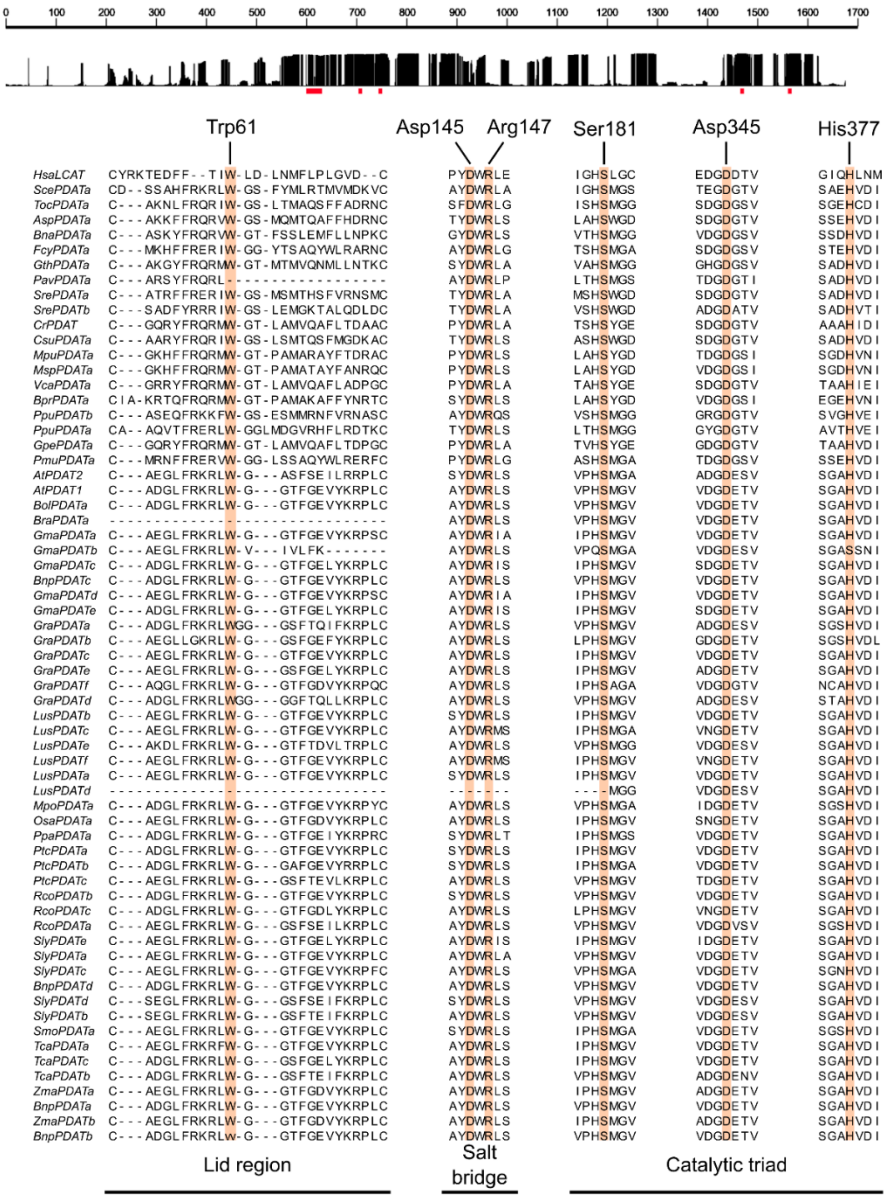
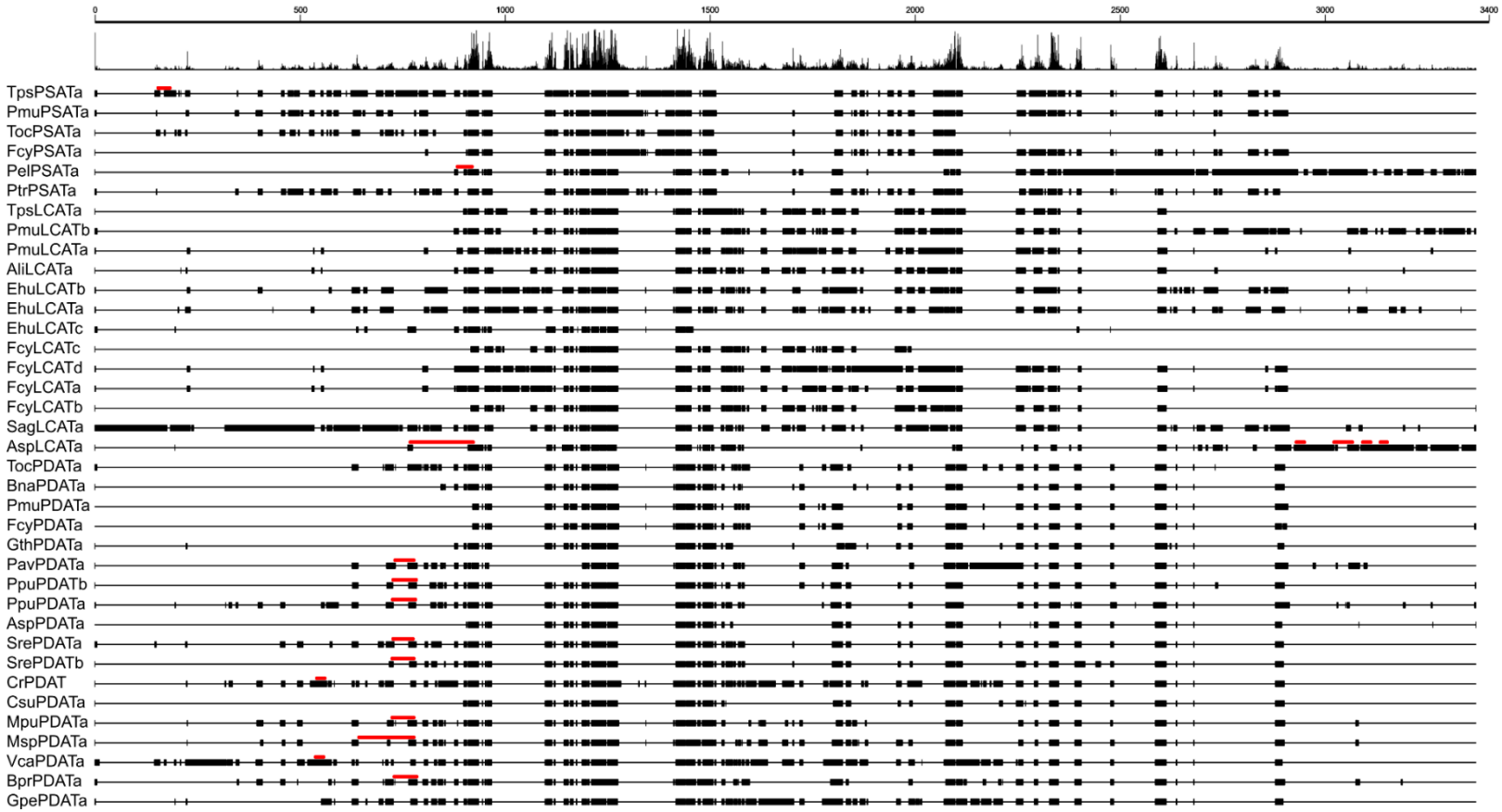


Figure B.2. Transmembrane domains (TMDs) of all sequences analyzed in this study. A) Multiple sequence alignment of all algal PDATs showing the TMDs as a thick red line above the aligned sequences. B) Multiple sequence alignment of all plant PDATs showing the TMDs as a thick red line above the aligned sequences. The alignment conservation is shown at the top of each figure. The thick black bars represent the sequences and the lines represent the gaps.

A



B

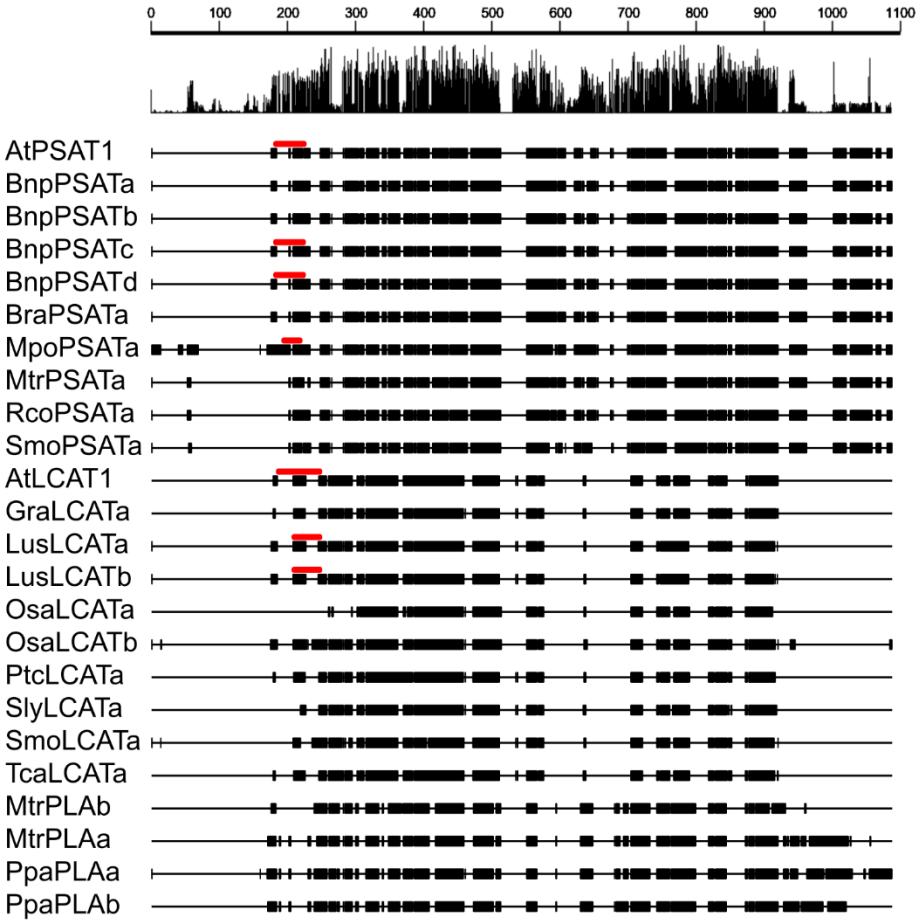


Figure B.2: B) (Continued)

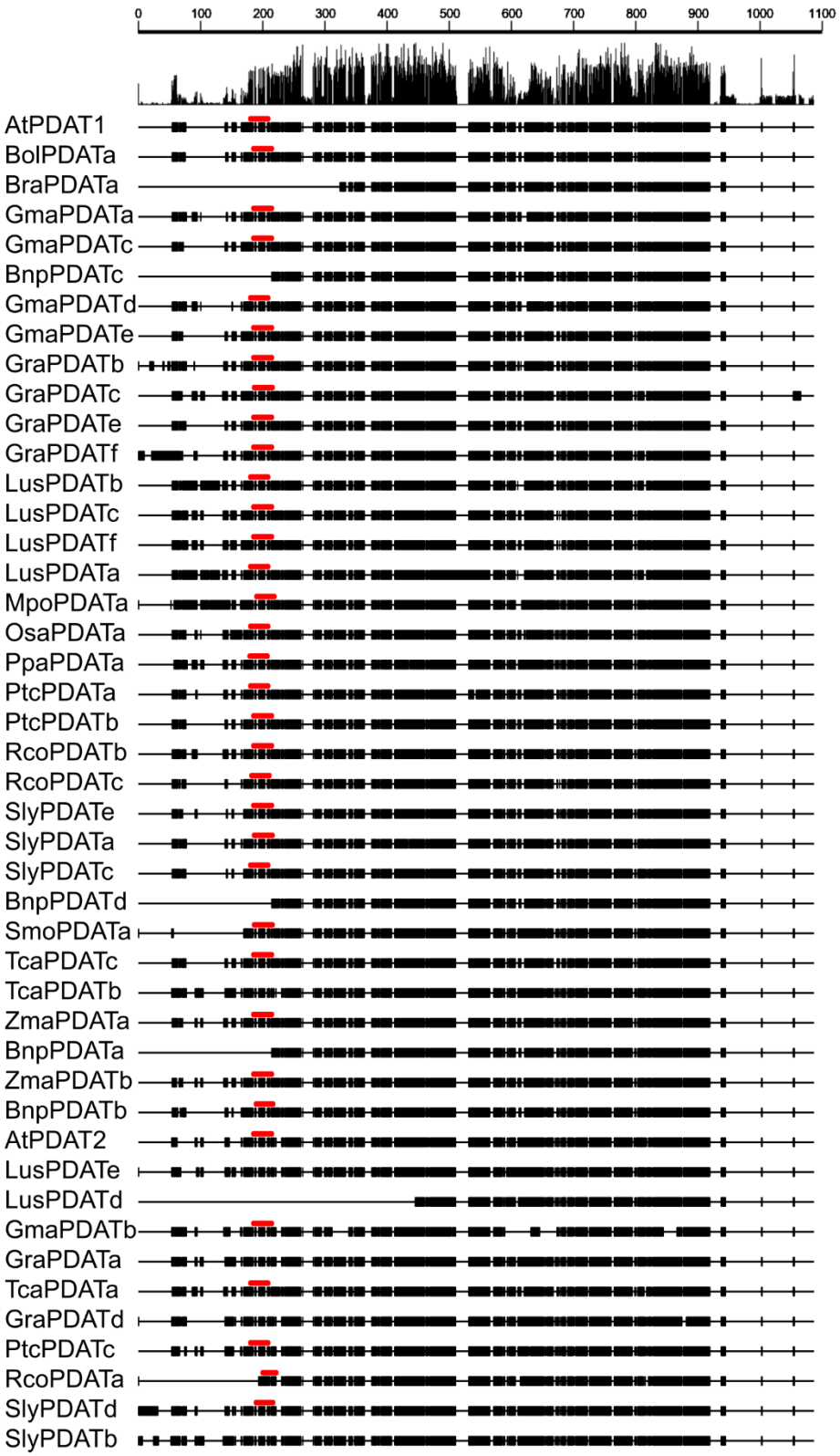


Figure B.2: B) (Continued)

Figure B.3. The sequence logos of the conserved motifs identified in the MEME v4.12.0 program. The E-value of each motif and the number of occurrences is also shown in each motif legend. Each letter in the sequence logos represents an amino acid and follows the usual abbreviation. The height of the letter represents its relative frequency in the alignment.

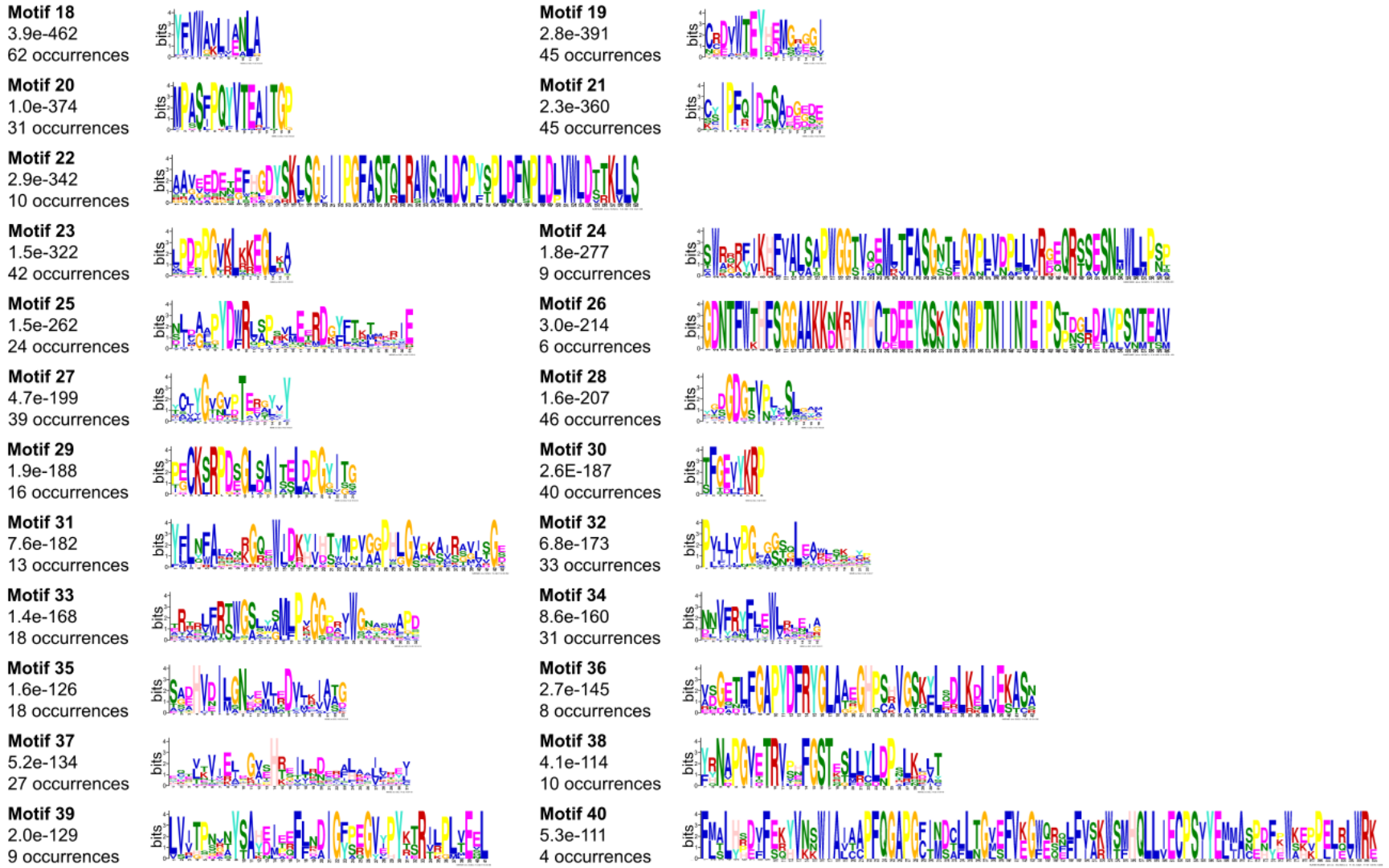


Figure B.3 (Continued)

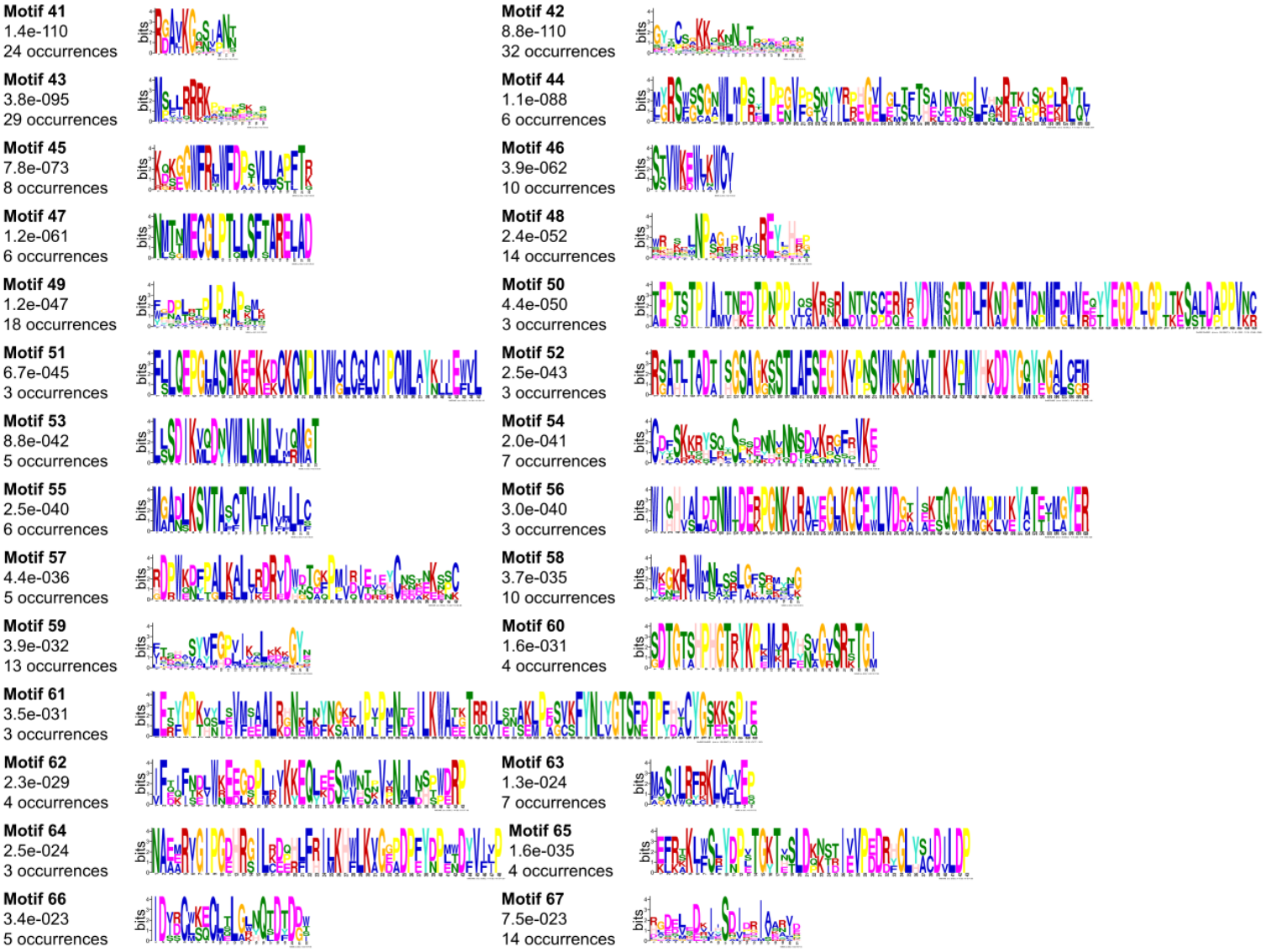


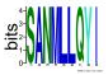
Figure B.3 (Continued)

Motif 68
1.1e-021
6 occurrences

Motif 69
1.2e-021
34 occurrences

Motif 70
5.4e-023
7 occurrences

Motif 71
7.0e-020
6 occurrences



Motif 72
3.2e-019
3 occurrences

Motif 73
4.9e-018
4 occurrences

Motif 74
8.1e-017
10 occurrences

Motif 75
1.0e-015
26 occurrences

Motif 76
2.6e-015
9 occurrences

Motif 77
1.4e-013
7 occurrences

Motif 78
2.2e-011
9 occurrences

Motif 79
3.6e-010
8 occurrences

Motif 80
5.7e-014
5 occurrences

Motif 81
2.0e-009
3 occurrences

Motif 82
6.0e-011
3 occurrences

Motif 83
6.8e-013
4 occurrences

Motif 84
1.7e-012
3 occurrences

Motif 85
3.3e-011
4 occurrences

Motif 86
1.5e-008
5 occurrences

Motif 87
1.5e-008
4 occurrences

Motif 88
1.4e-007
4 occurrences

Motif 89
1.1e-006
3 occurrences

Motif 90
1.5e-005
6 occurrences

Motif 91
1.1e-005
5 occurrences

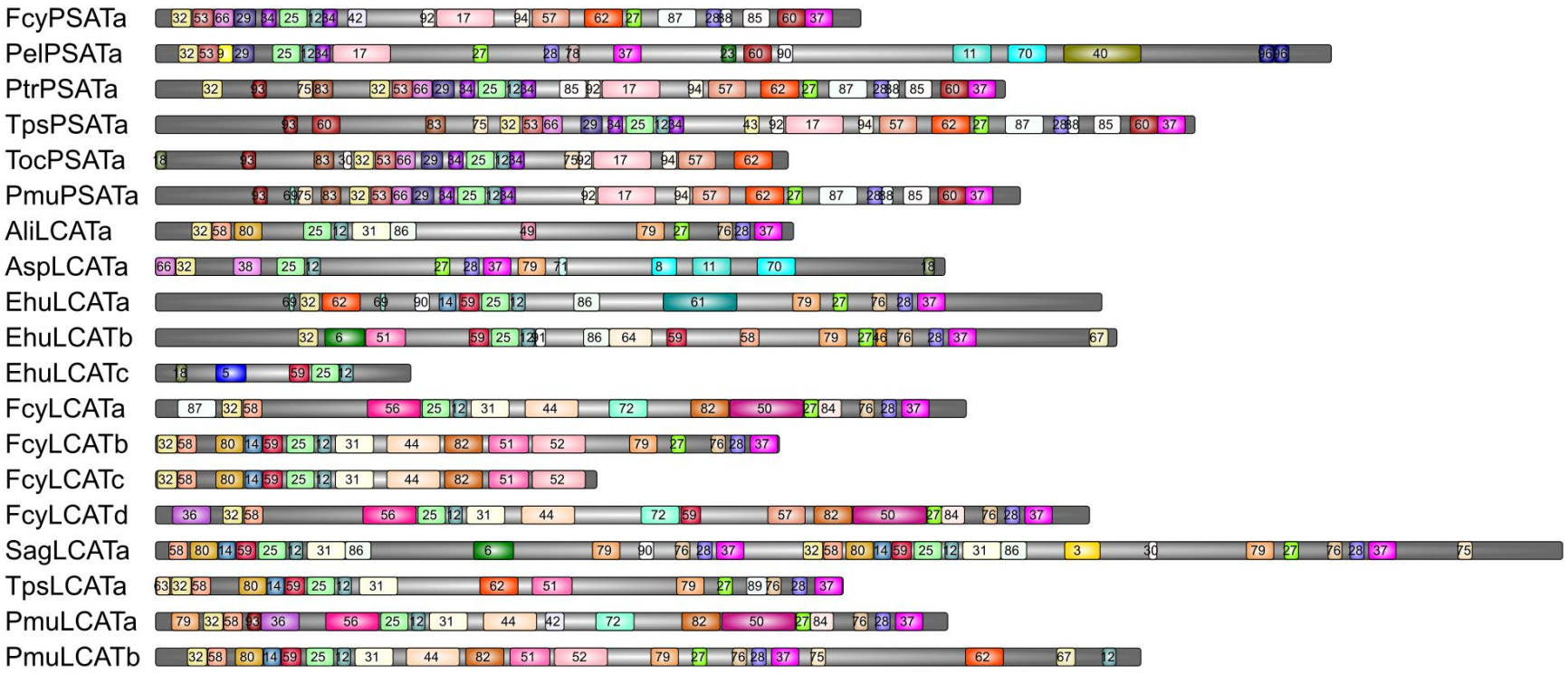
Motif 92
3.3e-005
5 occurrences

Motif 93
4.3e-005
4 occurrences

Motif 94
7.3e-005
5 occurrences

Figure B.3. (Continued)

Figure B.4. Conserved motifs identified in the MEME analysis of all sequences used in this study. Grey bars represent the sequences. The motifs are shown as bars with different colors and they are identified by a number inside this bar.



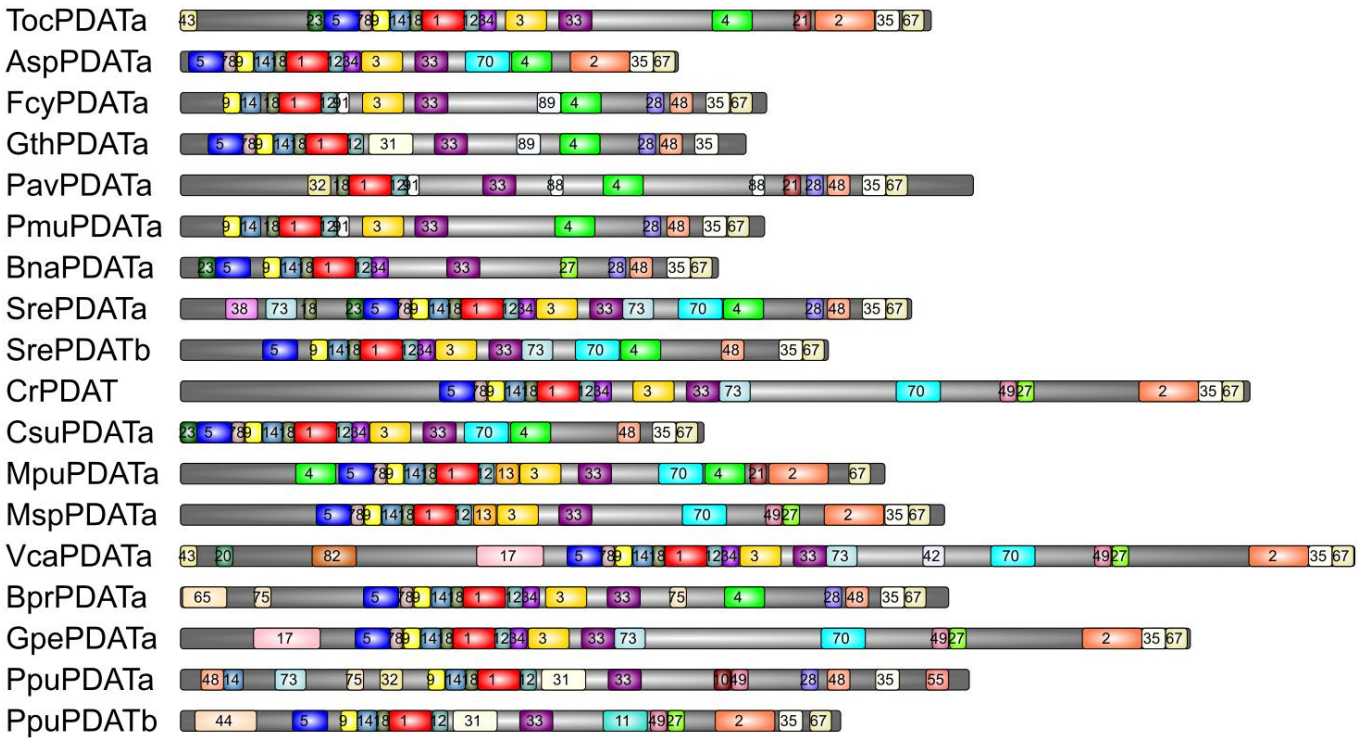


Figure B.4 (Continued)

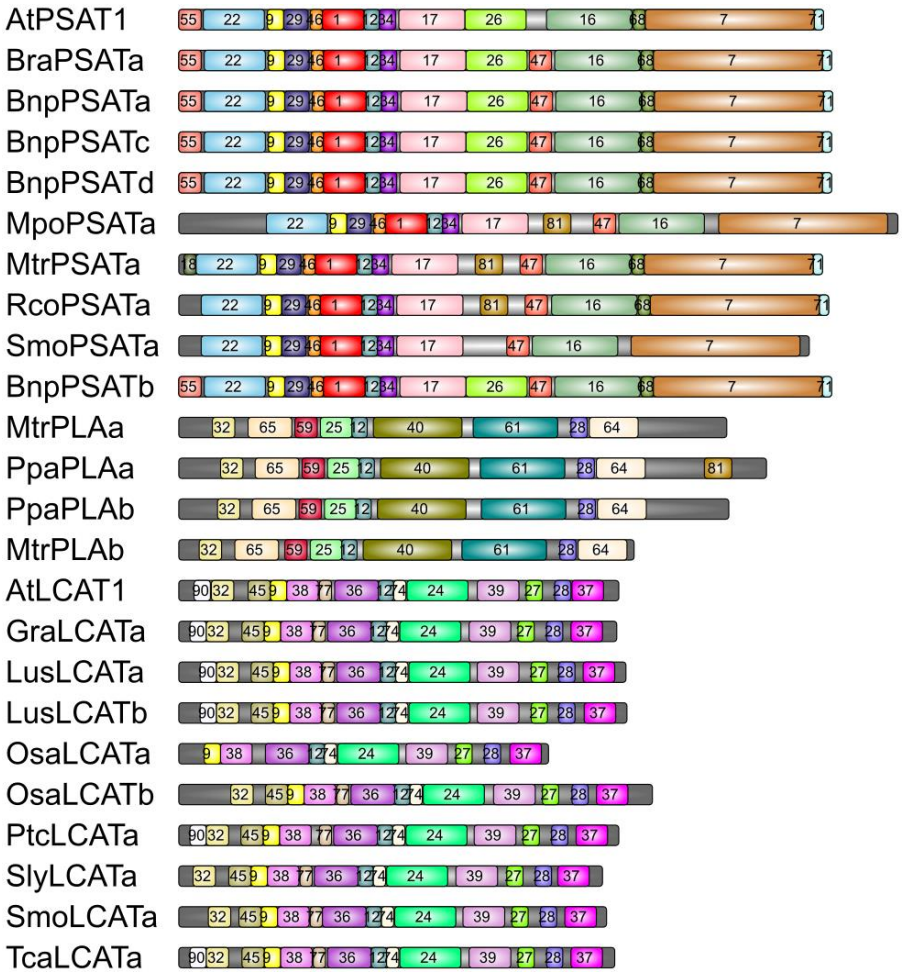


Figure B.4 (Continued)

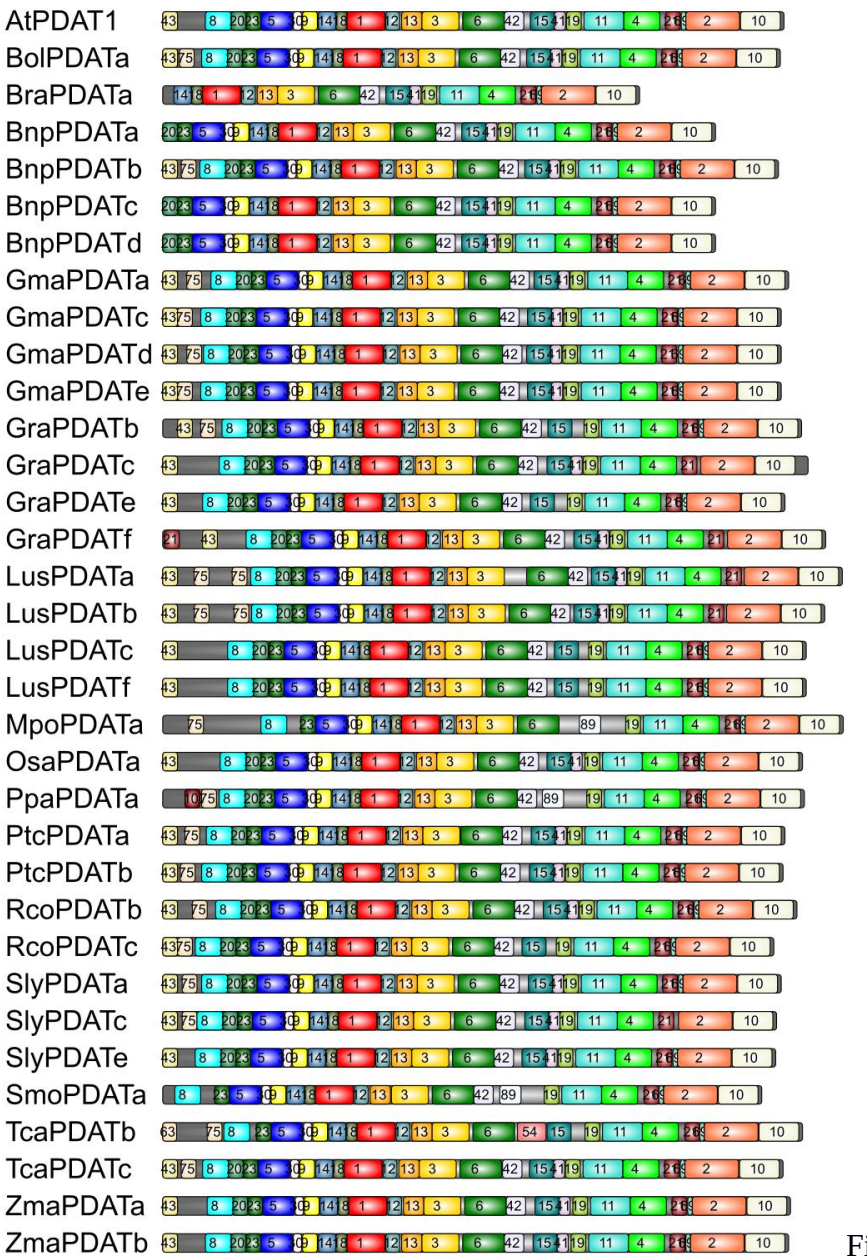


Figure B.4 (Continued)

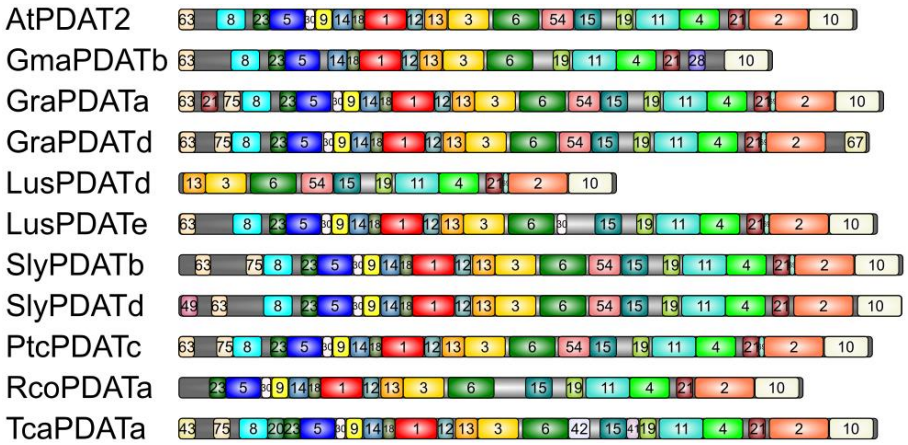
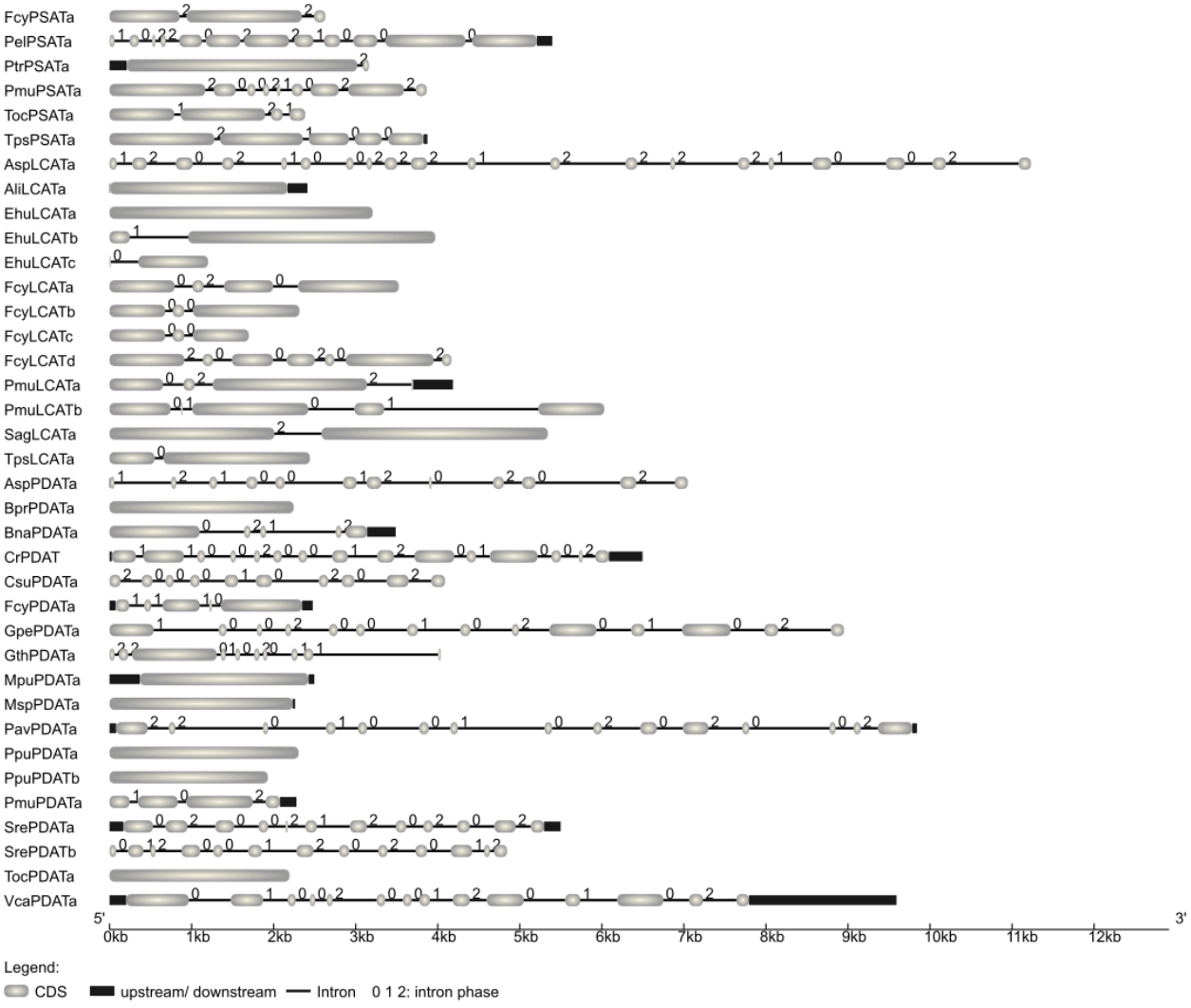


Figure B.4 (Continued)

Figure B.5. Schematic diagram showing the exons and introns of sequences used in the analyses. Grey bars represent the exons and black lines represent the introns. The intron phase is shown as a number above the introns.



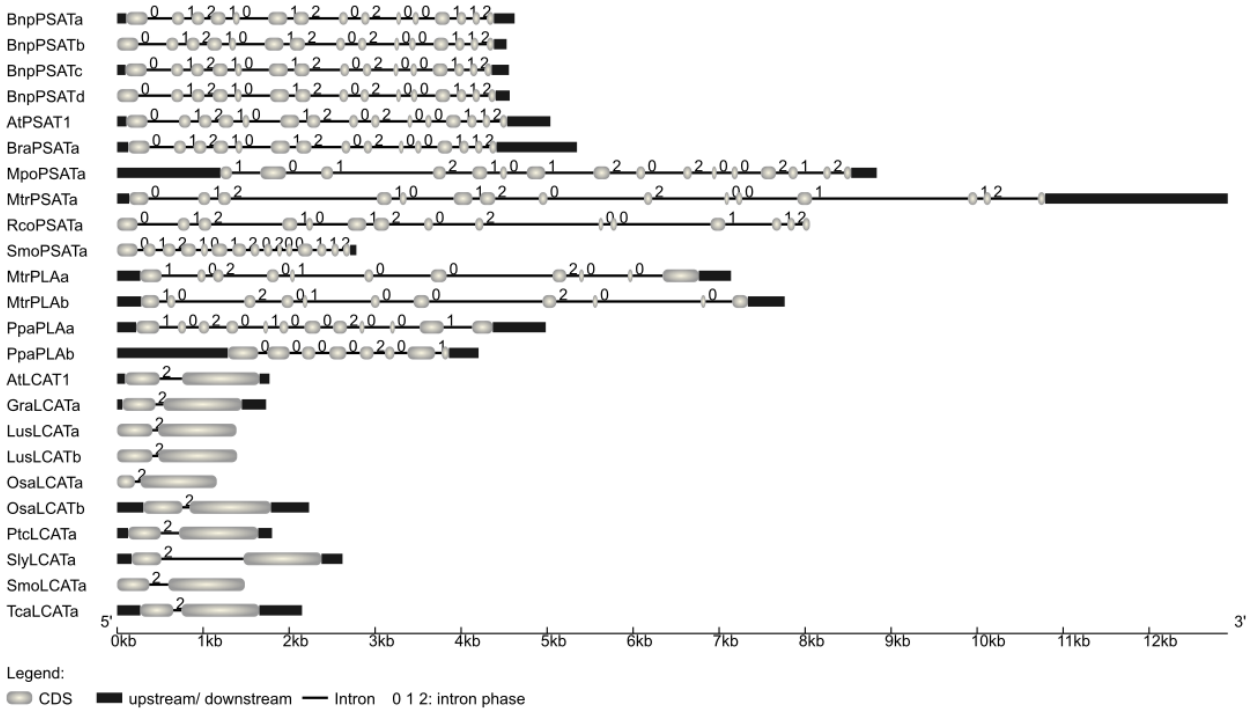


Figure B.5 (continued)

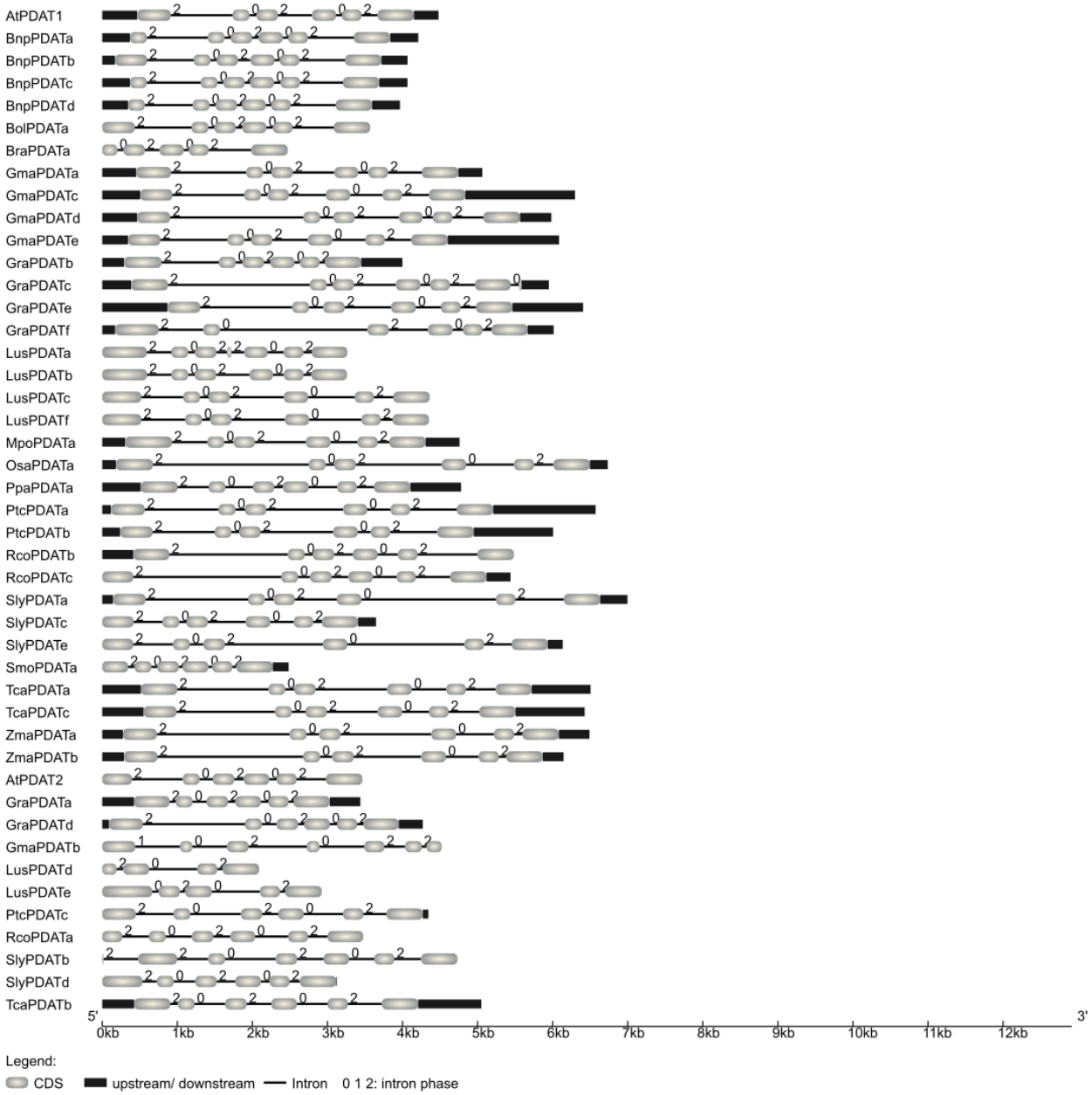


Figure B.5 (continued)

Figure B.6. Phylogenetic analysis of algal LCAT-family enzymes. *Porphyridium purpureum* PDATs (PpuPDATa and PpuPDATb, shown in red) forms a basal group to the green algae and chromists. The Bayesian inference phylogenetic tree was reconstructed with MrBayes v3.2.6. Branch-supporting values (shown as a percentage) represent Bayesian posterior probabilities and are shown on the branches. Sequences were classified into PDAT, PSAT, PLA or LCAT-related based on PANTHER implemented in InterProScan v5.22.61. LCAT, lecithin-cholesterol acyltransferase; PDAT, phospholipid:diacylglycerol acyltransferase; PSAT, phospholipid:sterol O-acyltransferase.

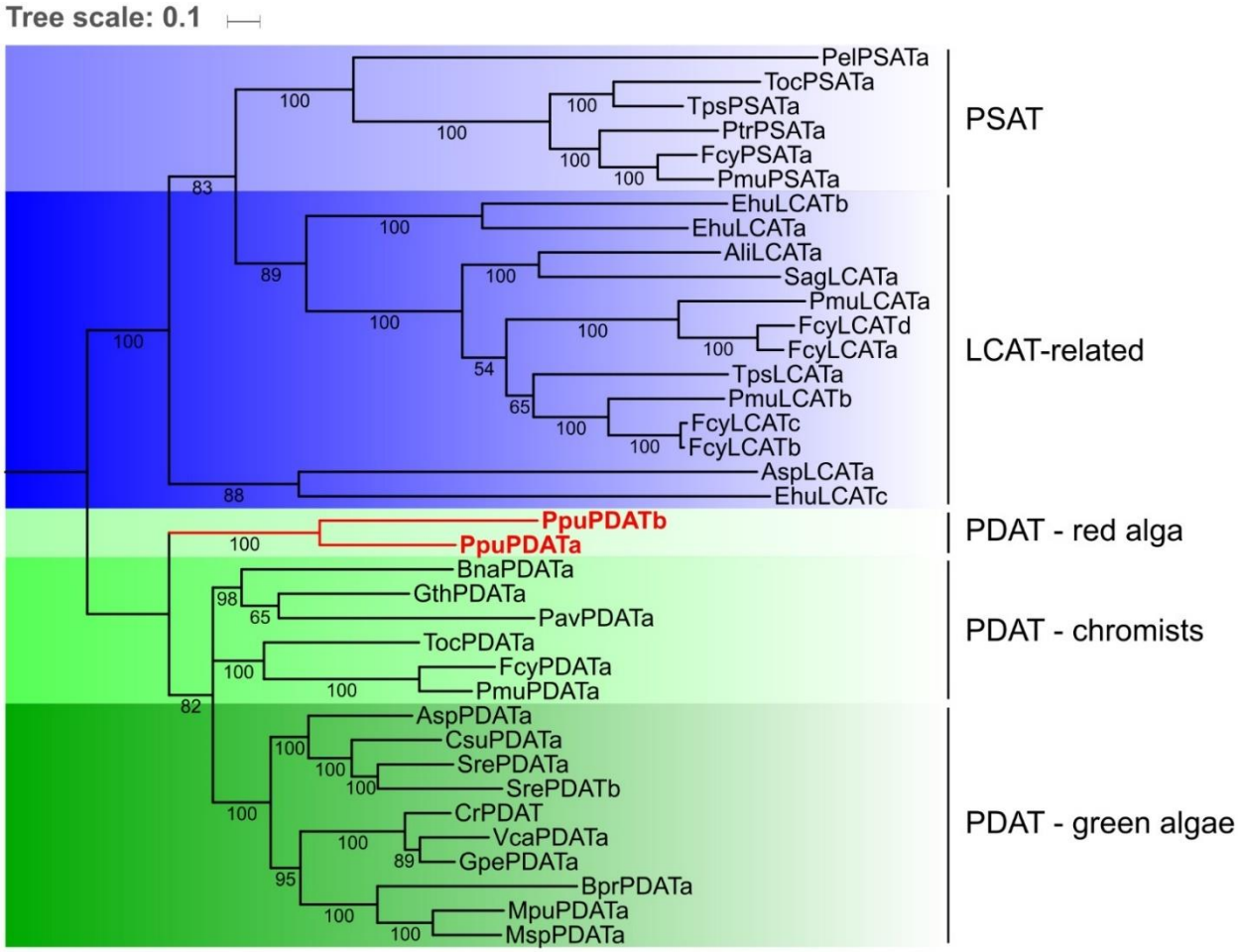


Figure B.7. Phylogenetic analysis algal and plant LCAT-family enzymes. The Bayesian inference phylogenetic tree was reconstructed with MrBayes v3.2.6. Branch-supporting values (shown as a percentage) are shown on the branches. Sequences were classified into PDAT, PSAT, PLA or LCAT-related based on PANTHER implemented in InterProScan v5.22.61. LCAT, lecithin-cholesterol acyltransferase; PDAT, phospholipid:diacylglycerol acyltransferase; PLA, phospholipase; PSAT, phospholipid:sterol O-acyltransferase.

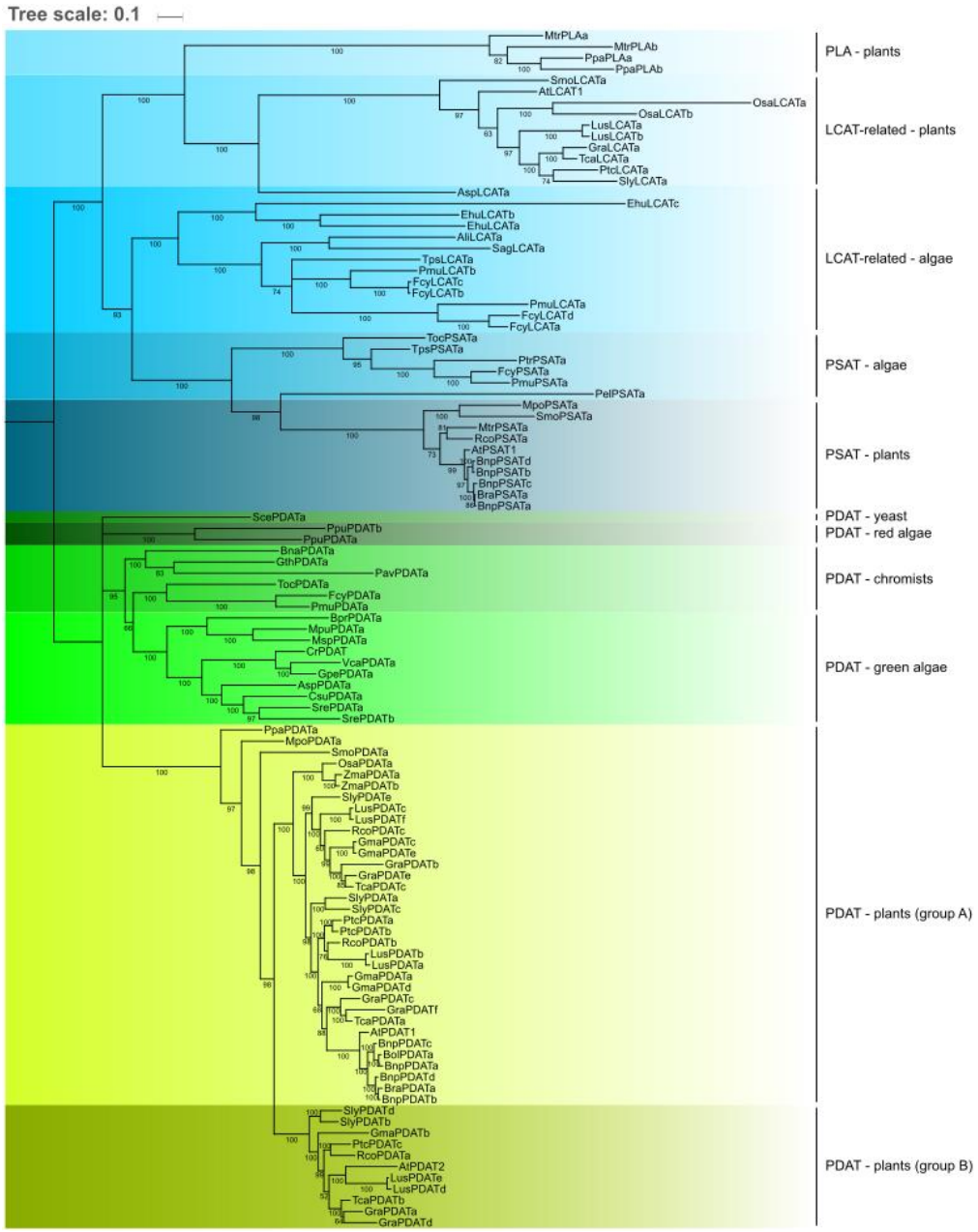


Figure B.8. Phylogenetic trees used in the Clade model C analyses. A) Green algal PDATs as the foreground group and plants (group A) PDATs as the background group. B) Green algal PDATs as the foreground group and plants (group B) PDATs as the background group. In both trees, the green algal PDATs are shown with a red color branch.

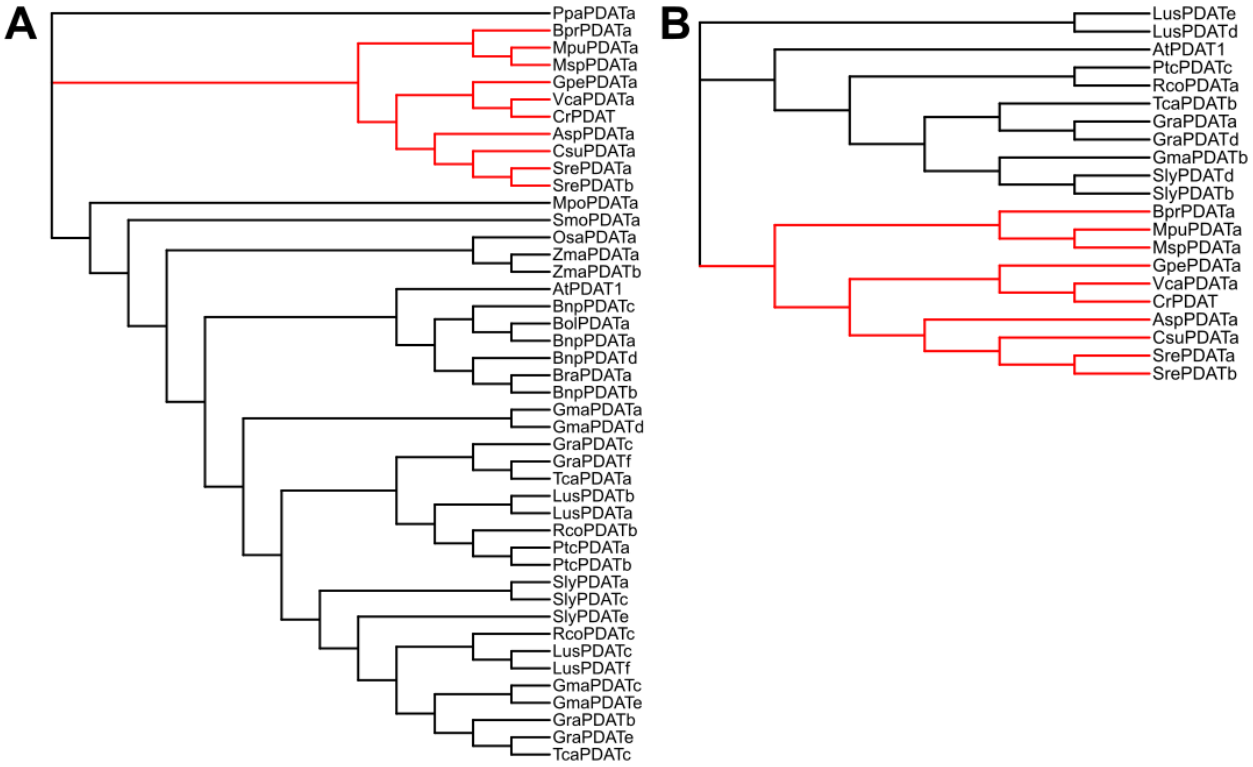


Figure B.9. Phylogenetic trees used in the branch-site model analyses. Foreground partition is shown with red branches. A) Comparison of green algal PDATs as the foreground group with plant (group A) sequences. B) Comparison of green algal PDATs with plant (group A) PDATs as the foreground partition. C) Comparison of green algal PDATs as the foreground group with plant (group B) sequences. D) Comparison of green algal PDATs with the plant (group B) PDATs as the foreground partition

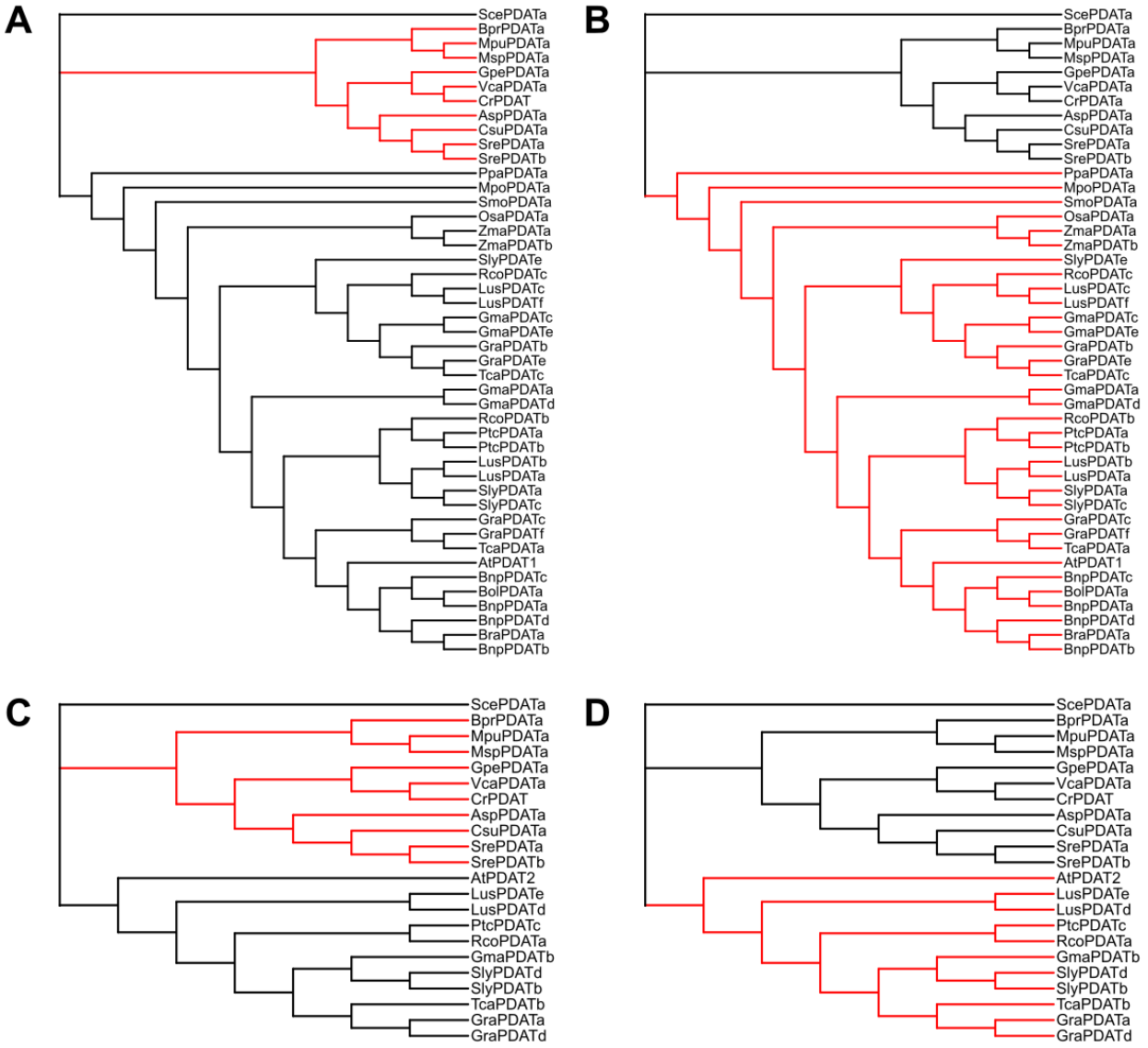
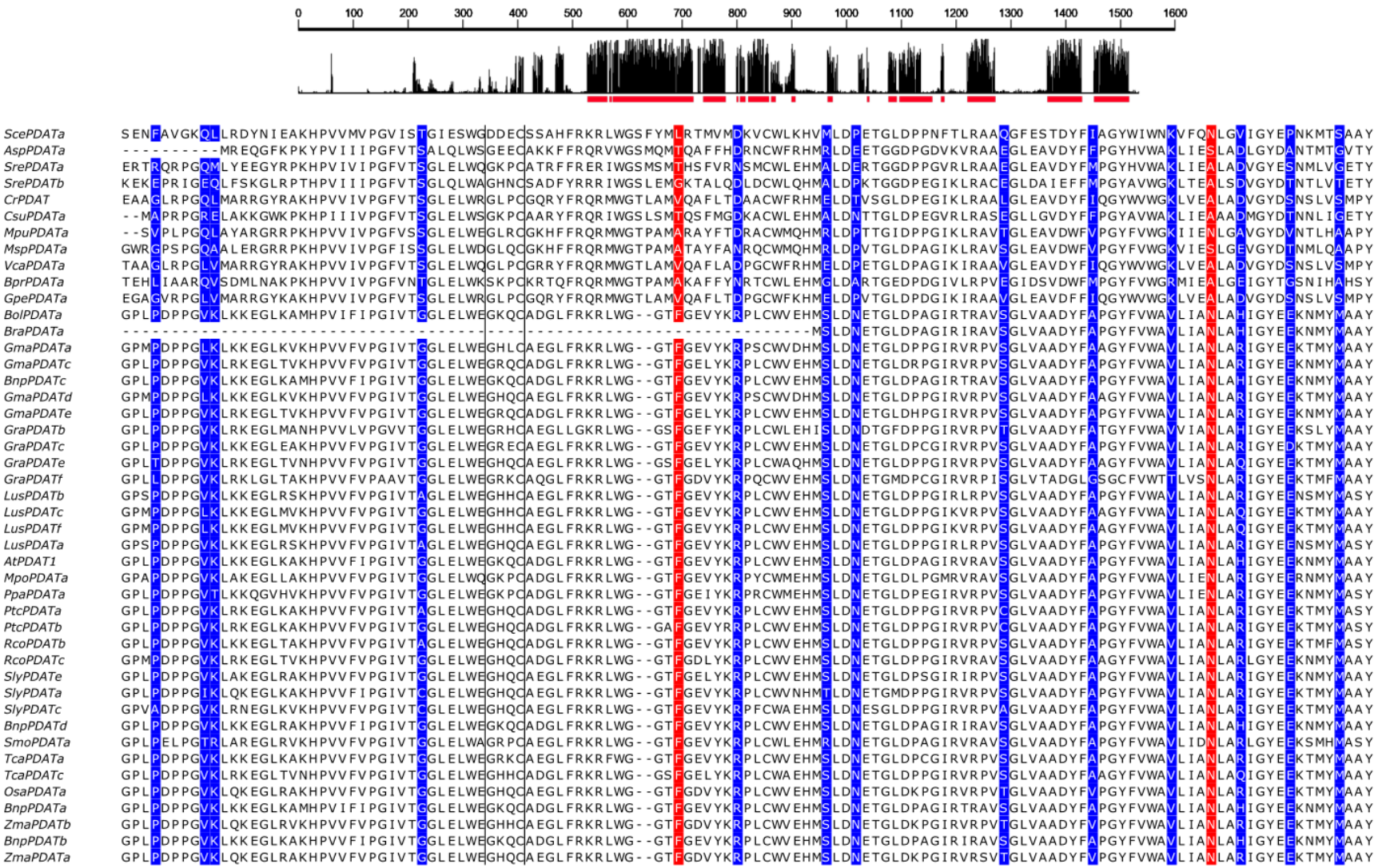


Figure B.10. Sites under positive selection identified from green algal and plant (group A) PDATs by the Bayes empirical Bayes method. The alignment consensus is shown at the top and the red bars below it represents the regions visualized in the figure. These regions are separated by black vertical lines in the alignment. Sites highlighted in red are under positive selection in green algal PDATs. Sites highlighted in blue are under positive selection in plant PDATs from group A. The amino acid position further studied in AtPDAT1 is marked with a star (★)


0 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600
ScePDATa
AspPDATa
SrePDATa
SrePDATb
CrPDAT
CsuPDATa
MpuPDATa
MspPDATa
VcaPDATa
BprPDATa
GpePDATa
BolPDATa
BraPDATa
GmaPDATa
GmaPDATc
BnpPDATc
GmaPDATd
GmaPDATe
GraPDATb
GraPDATc
GraPDATe
GraPDATf
LusPDATb
LusPDATc
LusPDATf
LusPDATa
AtPDAT1
MpoPDATa
PpaPDATa
PtcPDATa
PtcPDATb
RcoPDATb
RcoPDATc
SlyPDATe
SlyPDATa
SlyPDATc
BnpPDATd
SmoPDATa
TcaPDATa
TcaPDATc
OsaPDATa
BnpPDATa
ZmaPDATb
BnpPDATb
ZmaPDATa

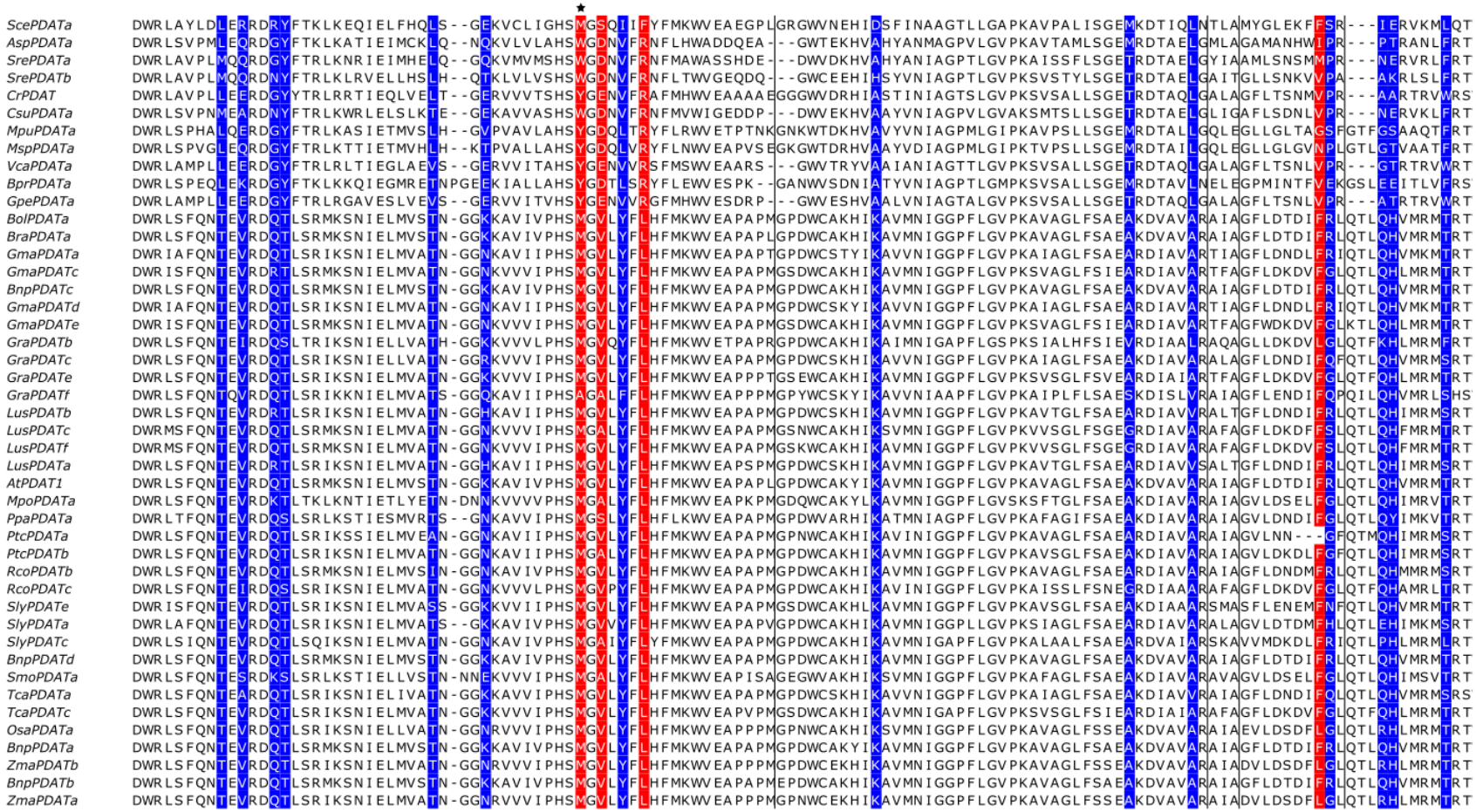


Figure B.10 (continued)

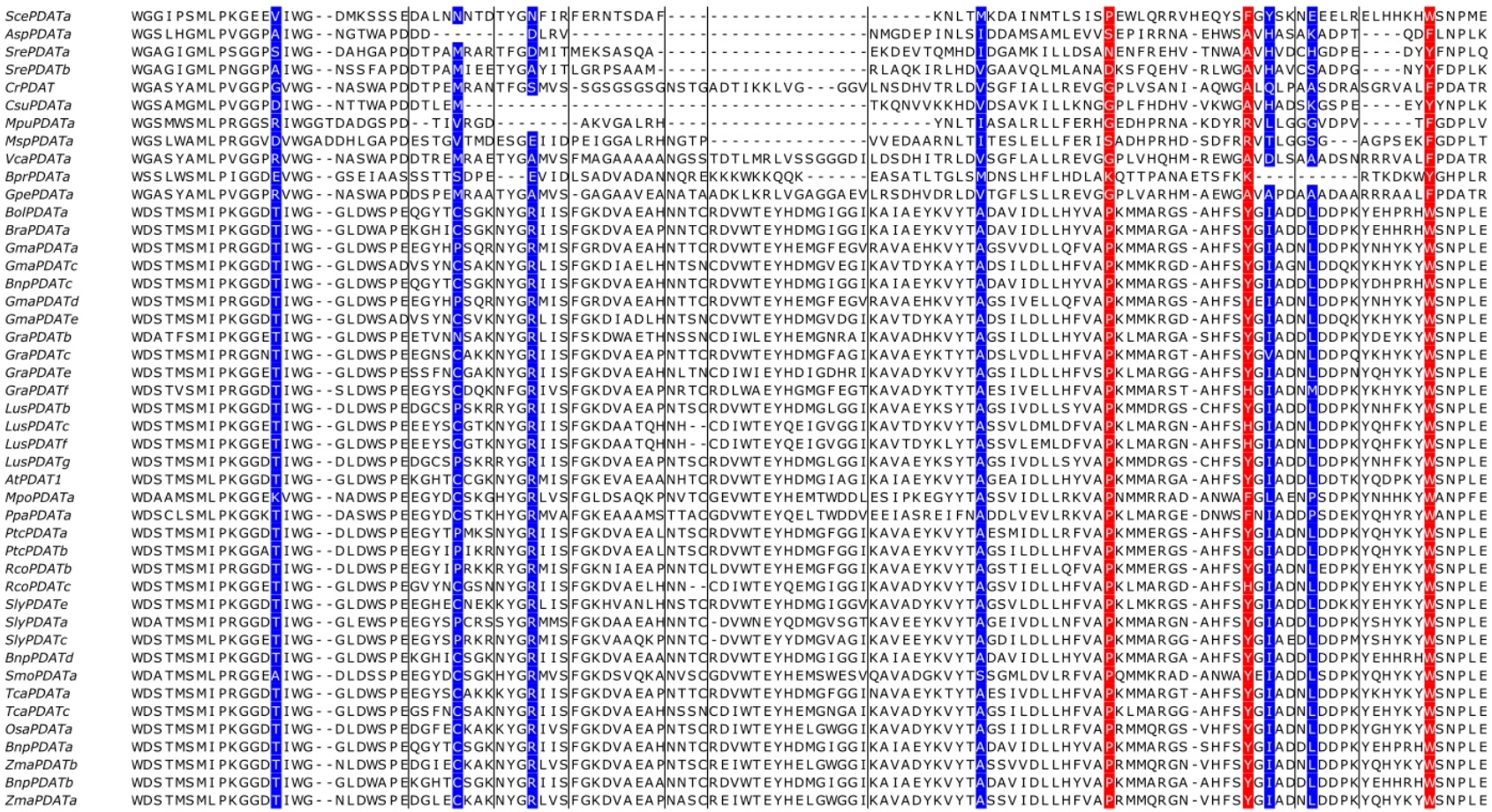


Figure B.10 (continued)

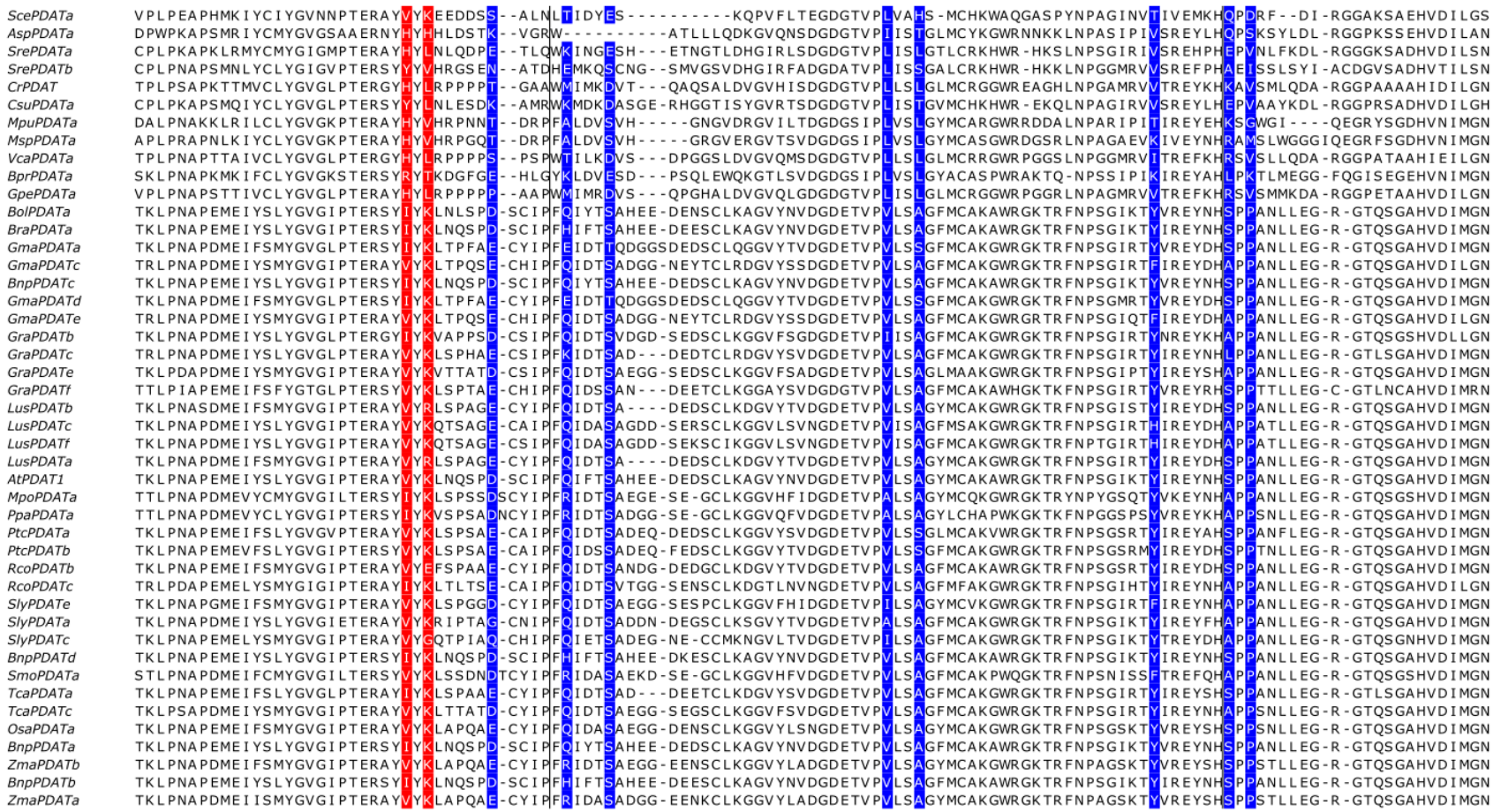


Figure B.10 (continued)

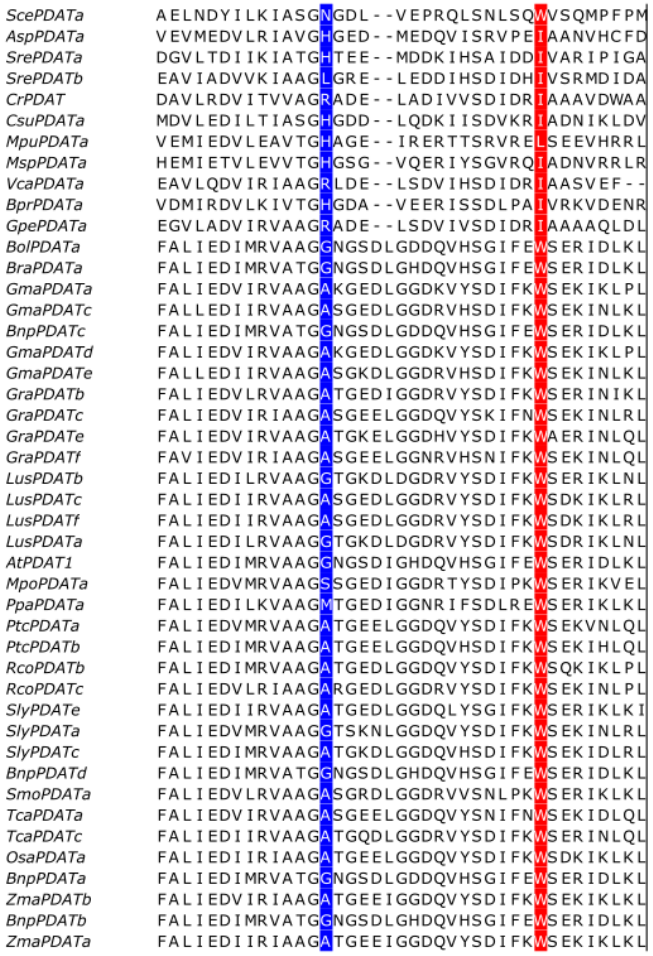
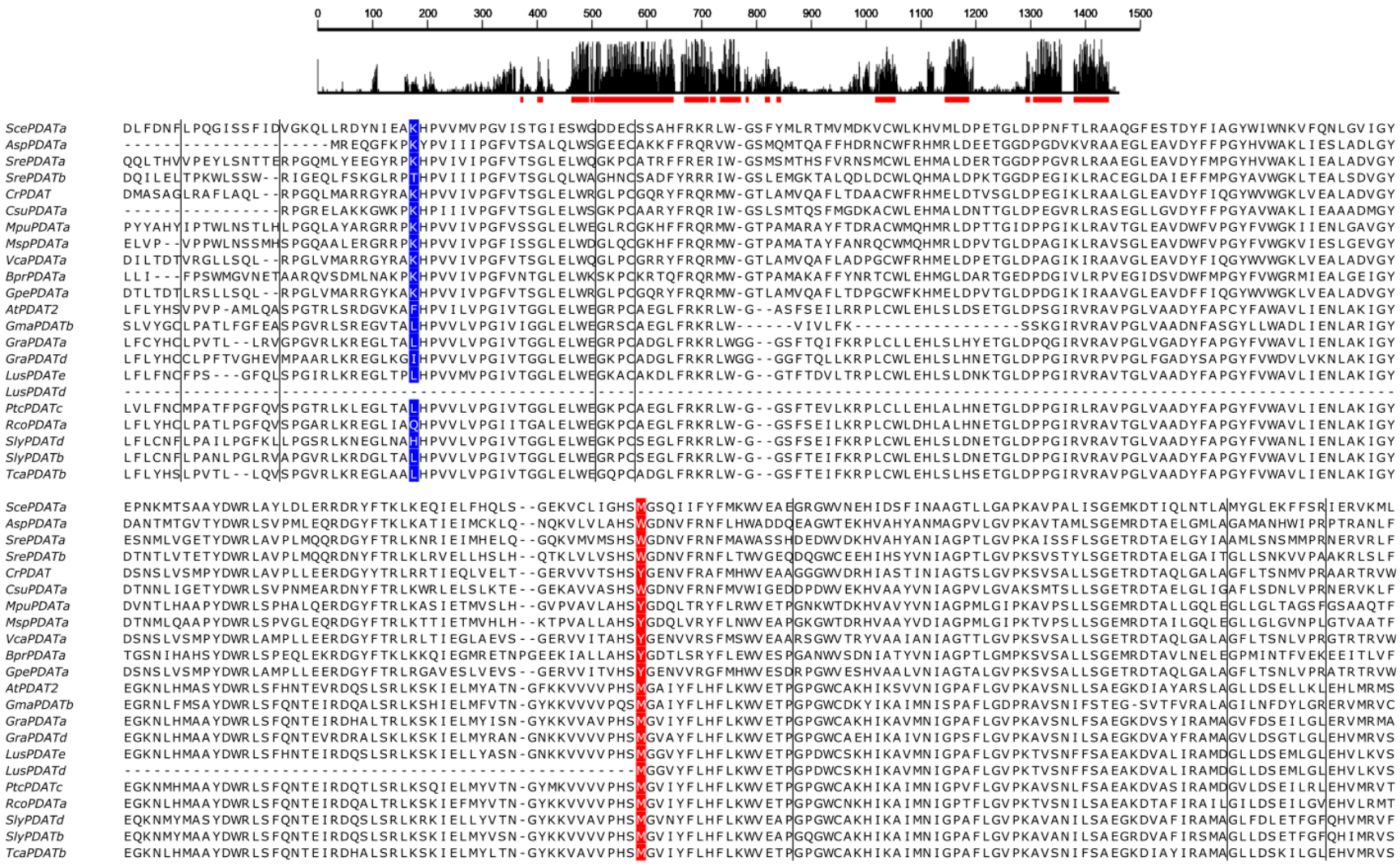


Figure B.10 (continued)

Figure B.11. Sites under positive selection identified from green algal and plant (group B) PDATs by the Bayes empirical Bayes method. The alignment consensus is shown at the top and the red bars below it represents the regions visualized in the figure. These regions are separated by black vertical lines in the alignment. The site highlighted in red is under positive selection in green algal PDATs. Sites highlighted in blue are under positive selection in plant PDATs from group A


0 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500
ScePDATa
AspPDATa
SrePDATa
SrePDATb
CrPDAT
CsuPDATa
MpuPDATa
MspPDATa
VcaPDATa
BprPDATa
GpePDATa
AtPDAT2
GmaPDATb
GraPDATa
GraPDATd
LusPDATe
LusPDATd
PtcPDATc
RcoPDATa
SlyPDATd
SlyPDATb
TcaPDATb

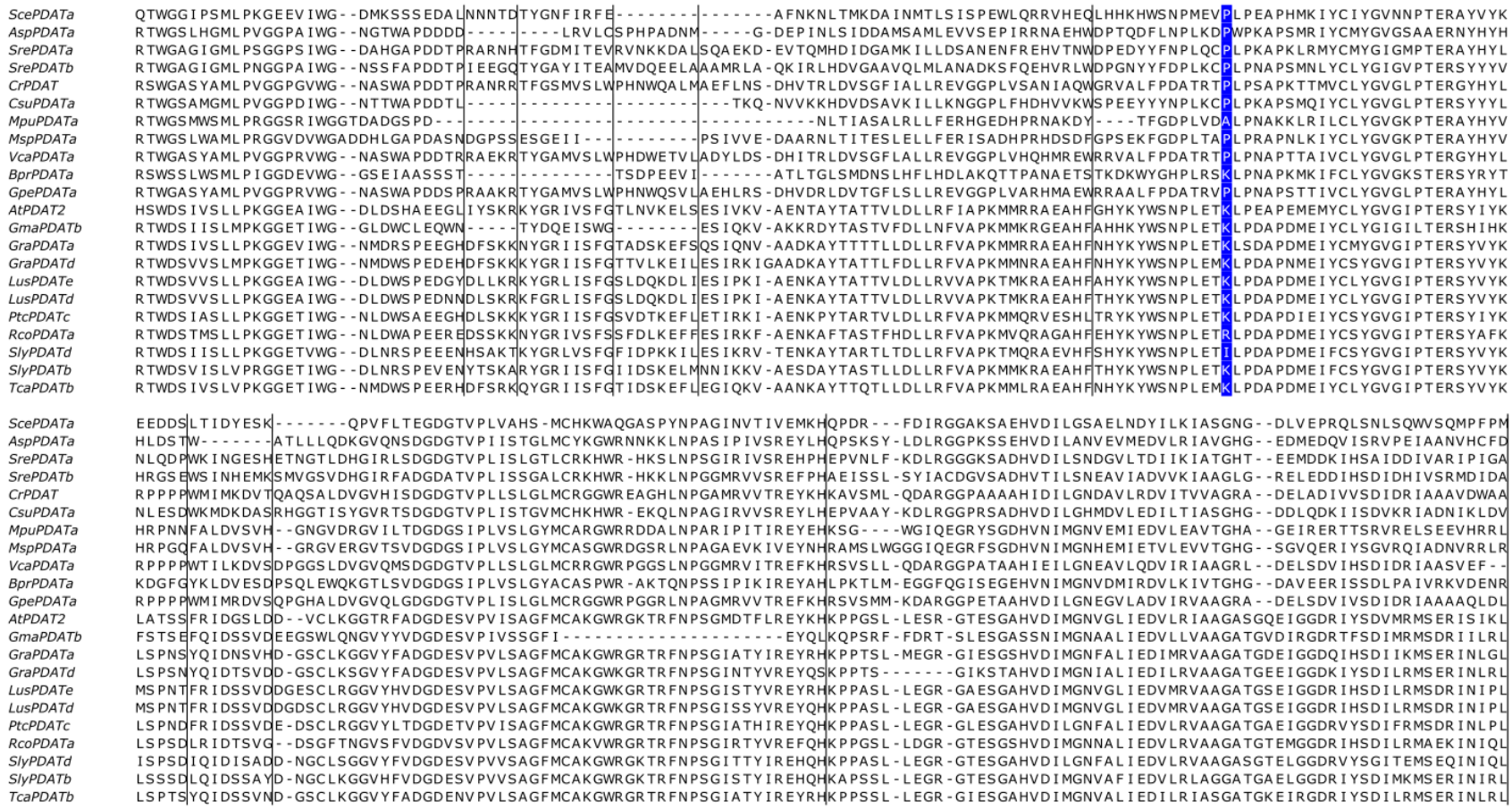


Figure B.11 (continued)

Figure B.12. Predicted structures of A) *Arabidopsis thaliana* PDAT1 and B) *Chlamydomonas reinhardtii* PDAT with the sites under positive selection shown in red. The atoms of the catalytic residues are shown in green (carbon), blue (nitrogen), and red (oxygen). This figure was generated by Dr. Kristian Caldo.

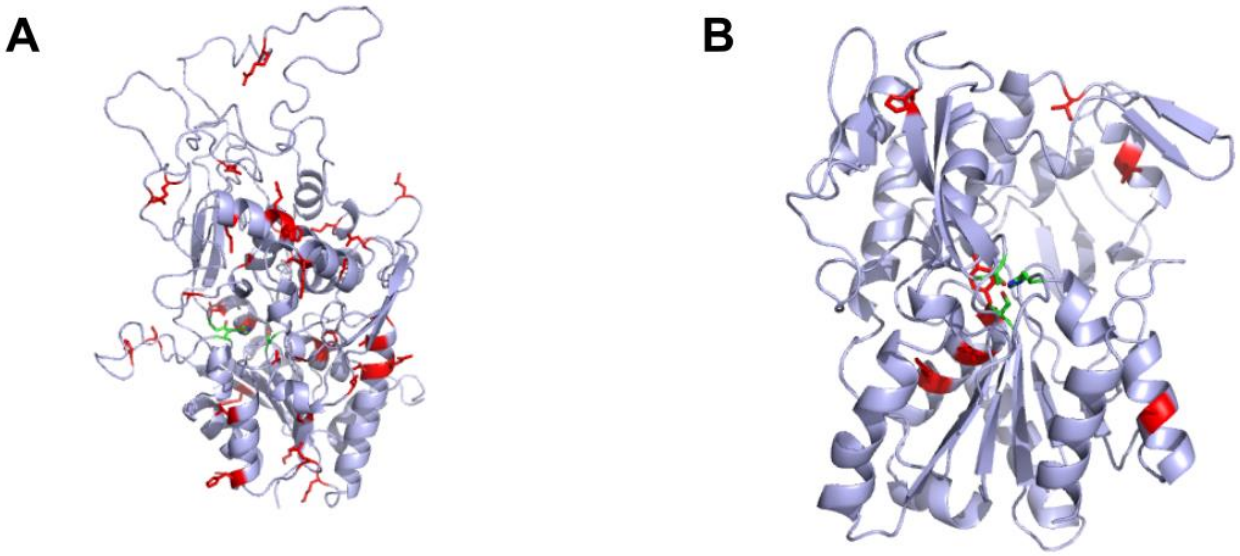


Figure B.13. Predicted interaction of sites under positive selection with other amino acids in *Arabidopsis thaliana PDAT1* (AtPDAT1). A) AtPDAT1 structure showing the location of positive selection sites in orange. The catalytic triad is also shown in green, red and blue. B) Some positive selection sites may influence or stabilize secondary structures through polar contacts with the main chain residues. C) Other positive selection sites may influence or stabilize tertiary structure through polar contacts. The distance between the atoms participating in polar contacts (yellow dash) is indicated in angstrom. This figure was generated by Dr. Kristian Caldo.

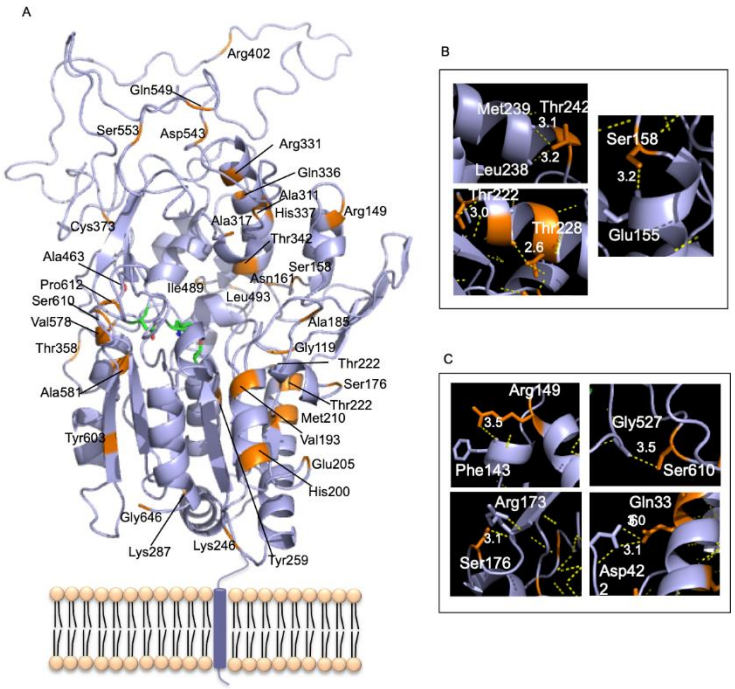


Figure B.14. Structure analysis of human LCAT. A) Homology model of the human LCAT obtained with SWISS-MODEL software. B-D) Close-up view of the active site pocket wherein the side chains for the catalytic triad (S205, D369, H401) are shown in different colors with green, red and blue corresponding to carbon, oxygen and nitrogen, respectively. The positive selection site (L206) is also shown in yellow. A simulation of the mutation of this site with methionine (L206M) and tyrosine (L206Y) is also shown in yellow. This figure was generated by Dr. Kristian Caldo.

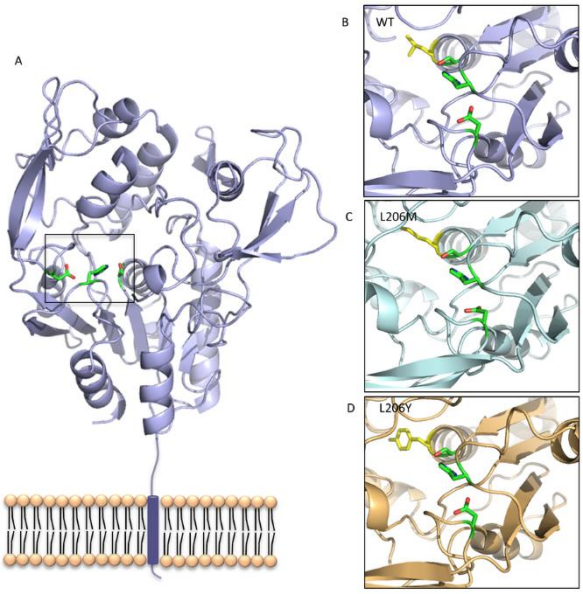


Figure B.15. Structure analysis of *Chlamydomonas reinhardtii* PDAT (CrPDAT) A) Homology model of CrPDAT obtained with PHYRE2 software. B-D) Close-up view of the active site pocket wherein the side chains for the catalytic triad (S401, D940, H994) are in different colors with green, red and blue corresponding to carbon, oxygen and nitrogen, respectively. The positive selection site (Y402) and its simulated mutation (Y402M) is also shown in yellow. This figure was generated by Dr. Kristian Caldo.