Engineering Arabidopsis long-chain acyl-CoA synthetase 9 variants with enhanced enzyme activity

Yang Xu[1], Kristian Mark P. Caldo[1], Roman Holic[2], Elzbieta Mietkiewska[1], Jocelyn Ozga[1], Syed Masood Rizvi[3], Guanqun Chen[1], and Randall J. Weselake[1*]

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

[2]Centre of Biosciences, Institute of Animal Biochemistry and Genetics, Slovak Academy of Sciences, Dúbravská cesta 9, 840 05 Bratislava, Ivanka pri Dunaji, 900 28, Slovakia

[3]Corteva Agriscience, Agriculture Division of DowDuPont, Site 600, RR #6 PO Box 12 Saskatoon, Saskatchewan, Canada S7K 3J9

* To whom correspondence should be addressed: Randall J. Weselake, Phone: (+1) 306 261-0560; E-mail: randall.weselake@ualberta.ca

Running title: Performance-improved AtLACS9 variants

Abstract:

Long-chain acyl-CoA synthetase (LACS, EC 6.2.1.3) catalyzes the ATP-dependent activation of free fatty acid to form acyl-CoA, which in turn serves as the major acyl donor for various lipid metabolic pathways. Increasing the size of acyl-CoA pool by enhancing LACS activity appears to be a useful approach to improve the production and modify the composition of fatty acid-derived compounds, such as triacylglycerol. In this study, we aimed to improve the enzyme activity of *Arabidopsis thaliana* LACS9 (AtLACS9) by introducing random mutations into its cDNA using error-prone PCR. Two AtLACS9 variants containing multiple amino acid residue substitutions were identified with enhanced enzyme activity. To explore the effect of each amino acid residue substitution, single site mutants were generated and the amino acid substitutions C207F and D238E were found to be primarily responsible for the increased activity of the two variants. Furthermore, evolutionary analysis revealed that the beneficial amino acid site C207 is conserved among LACS9 from plant eudicots, whereas the other beneficial amino acid site D238 might be under positive selection. Together, our results provide valuable information for production of LACS variants for applications in the metabolic engineering of lipid biosynthesis in oleaginous organisms.

Key words: LACS; error-prone PCR; protein engineering; *Arabidopsis thaliana*; *Saccharomyces cerevisiae*

Abbreviations list:

The abbreviations used are: ACS, acyl-CoA synthetase; *AtLACS, Arabidopsis thaliana LACS*; CoA, coenzyme A; BEB, Bayes empirical Bayes; ER, endoplasmic reticulum; GTR, general time reversible; LACS, long-chain acyl-CoA synthetase; SD, standard deviation; TAG, triacylglycerol; WT, wild type.

Introduction

Fatty acids are carboxylic acids with highly reduced acyl chains, which act as the major energy reservoir in eukaryotic cells and the building blocks for all cellular lipids, including phospholipids, triacylglycerols (TAGs), isoprenoids, sterols, cutins, suberins and jasmonates. Free fatty acids are activated to their coenzyme A (CoA) derivatives prior to further metabolism in an ATP-dependent process catalyzed by long-chain acyl-CoA synthetase (LACS, EC 6.2.1.3). This reaction involves a two-step ping-pong reaction mechanism, in which firstly an adenylyl from ATP is transferred to a fatty acid forming an enzyme-bound acyl-adenylate intermediate and pyrophosphate and secondly, the acyl-adenylate intermediate is attacked by a CoA yielding an acyl-CoA and an AMP [1].

LACS activity was first reported in guinea pig (*Cavia porcellus*) liver by Kornberg and Pricer (1953) [2] and was later found in numerous other organisms [1]. In plants and many other organisms, multiple *LACS* genes are present, which encode enzymes that appear to function in different aspects of lipid metabolism. For instance, the model plant *Arabidopsis thaliana* (hereafter Arabidopsis) contains nine *LACS* (*AtLACS*) genes with distinct expression patterns, subcellular localizations and functions [3]. AtLACS6 and AtLACS7 are peroxisomal-localized enzymes required for the activation of fatty acids for *β*-oxidation during seedling development [4,5]. In contrast, AtLACS1 and AtLACS2, which are localized in the endoplasmic reticulum (ER), are involved in surface lipid biosynthesis [6–9]. In addition, AtLACS4, another ER-bound LACS, functionally overlaps with AtLACS1 in mediating the synthesis of lipids for pollen coat formation [10]. LACS may also play a crucial role in TAG biosynthesis in plants. Indeed, *de novo* fatty acids synthesized in the plastid are required to be activated to acyl-CoAs by LACS for use in TAG assembly in the ER [11]. The presence of LACS activity in the plastidial envelope was demonstrated [12], and further evidence suggested that LACS was specifically associated with the outer envelope [13,14]. In Arabidopsis, AtLACS9 is the only LACS associated with plastid outer envelope and thus was regarded as the most likely candidate for activating and exporting plastidially-derived fatty acids for TAG assembly [15]. The function of AtLACS9, however, is still debatable as a recent study has shown that AtLACS9 might contribute to lipid trafficking from the ER back to the plastid [16].

Regardless of the multiple roles of LACS enzymes in plants, the applications of *LACS* genes in engineering oleaginous microorganisms have been widely explored. Over-expressing

LACS has been shown to increase the production of fatty acid esters, fatty alcohols, waxes, and TAGs in *Escherichia coli* and yeast (*Saccharomyces cerevisiae*). For example, metabolic engineering of *E. coli* to produce fatty acid esters, fatty alcohols and waxes was achieved by over-expression of *LACS* in combination with other genes [17]. In addition, heterologous over-expression of *LACS* from diatoms (*Phaeodactylum tricornutum* and *Thalassiosira pseudonana*) or higher plants (Arabidopsis and *Brassica napus*) was applied to facilitate fatty acid uptake and stimulate oil deposition in yeast [18–21]. Since LACS provides the substrates for acyl-CoA-dependent acyltransferases to produce various lipids, over-expression of *LACS* in these microorganisms appears to contribute to directing the carbon flux to the lipid biosynthesis pathways by enhancing the size of the acyl-CoA pool. In addition, improved LACS production may directly promote substrate channeling for lipid biosynthesis considering the possible direct association or cooperation of LACS with the down-stream lipid biosynthetic enzymes [22, 23]. In these regards, further improvement of the enzyme activity of LACS via protein engineering might represent novel perspectives for engineering lipid production in oleaginous organisms. Indeed, over-expression of cDNAs encoding improved enzyme variants has been demonstrated to be a more efficient strategy to increase lipid production than using wild-type enzymes in various plant and microorganism species [24-29]. In addition, it has been shown that the beneficial mutations identified from one enzyme could be used to improve the performance of an enzyme from another species [24, 30], and thus could potentially benefit the *in planta* improvement of enzyme action using non-transgenic approaches such as CRISPR [31].

A few crystal structures of acyl-CoA synthetases (ACSs) from bacteria and mammals have been solved, including a medium-chain specific ACS from human (*Homo spaines*) [32], a long-chain specific ACS from *Thermus thermophilus* [33], and a very-long-chain specific ACS from *Mycobacterium tuberculosis* [34]. The lack of a detailed three-dimensional structure for AtLACS9, however, does not facilitate a rational design approach to modify the enzyme. Indeed, protein modification by rational design approaches are often constrained in practice since the protein structure and function relationship is intricate and hard to predict [35]. As an alternative, error-prone PCR has been used in generating enzyme variants with improved catalytic properties including increased activity, altered substrate specificity and increased temperature tolerance in the absence of detailed structure information [36–39]. This approach is so powerful and robust

that it could even identify beneficial mutations that were overlooked in rational design experiments [40,41].

In the current study, our strategy is to engineer enhanced performance in LACS using error-prone PCR and site-directed mutagenesis. Two AtLACS9 variants with multiply amino acid substitutions were generated with increased enzyme activity. The possible function of key amino acid residues affecting enzyme activity was further evaluated through *in vitro* enzyme assays and evolutionary analysis. The identified amino acid residue substitutions provide valuable information for the modification of LACS from different species.

Experimental

Cloning, random mutagenesis and site-directed mutagenesis of *AtLACS9*

The coding sequence of *AtLACS9* was amplified using a cDNA preparation from Arabidopsis developing seeds and was cloned into the pYES2.1 vector (pYES2.1-V5/HIS vector, Invitrogen, Burlington, ON, Canada) under the control of *GAL1* promoter and *CYC1* terminator to yield pYES-AtLACS9. The stop codon of *AtLACS9* was removed for in-frame fusion with a C-terminal V5 epitope. Random mutagenesis of *AtLACS9* was carried out by error-prone PCR using the GeneMorph II Random Mutagenesis kit (Agilent Technologies, Santa Clara, CA) and pYES-AtLACS9 as a template. The DNA fragment containing *AtLACS9* coding region and the regions on the pYES2.1 backbone (250 bp before and 155 bp after the AtLACS9 coding region) was amplified using Mutazyme II DNA polymerase and 2000 ng and 200 ng of pYES-AtLACS9 plasmid. PCR reaction was performed for 30 cycles of 95^{o}C for 30 s (denaturation), 55^{o}C for 30 s (annealing), 72^{o}C for 2 min 10 s (extension). Site-directed mutagenesis within *AtLACS9* was conducted using the QuikChange™ Site-Directed Mutagenesis Kit (Stratagene Cloning Systems, La Jolla, CA) and pYES-AtLACS9 as a template. The primers used for the random mutagenesis and construction of various single-site mutants are listed in Supplementary Table S1.

Selection of positive clones and heterologous expression of *AtLACS9* variants in yeast

The product of error-prone PCR was purified and co-transformed with the linearized pYES2.1 vector backbone into *S. cerevisiae* strain *BYfaa1,4Δ* (*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0, faa1 Δ::HIS3, faa4 Δ::LYS2*) [23] using the *S.c.* EasyComp Transformation Kit (Invitrogen) for recombination. *S. cerevisiae* yeast transformants were selected on minimal medium [0.67% (w/v) yeast nitrogen base and 0.2 % (w/v) SC-URA] containing 2% (w/v)

galactose and 1% (w/v) raffinose, 100 mM oleic acid and 45 mM cerulenin. Tyloxapol 1% (v/v) was also added into plates to disperse the fatty acids. After incubating at 30ºC for 2~3 days, the individual colonies grown on the selection plates were then used to inoculate minimal medium containing 2% (w/v) galactose and 1% (w/v) raffinose (refer to as induction medium) in 96-well plates. The yeast cultures were grown at 30ºC for 24 h and 48 h before subjected to Nile red assay (using the protocol described below).

For heterologous expression of *AtLACS9* variants, the coding sequences of the selected *AtLACS9* variants were amplified and re-cloned into the pYES2.1 vector and the resulting plasmids were sequenced and used to transform *S. cerevisiae* mutant *BYfaa1,4Δ*. The recombinant yeast cells were cultured in 2% (w/v) raffinose minimal medium. After overnight growth, the yeast cultures were inoculated into induction medium to an optical density of 0.4 at 600 nm (OD_{600}). Yeast cultures were grown at 30ºC with shaking at 220 rpm.

Lipid analysis

Neutral lipid analysis in yeast cells was performed using the Nile red fluorescence assay with a Synergy H4 Hybird reader (Biotek, Winooskit, VT, USA) as described previously [30]. In brief, 100 μL aliquots of yeast culture were placed in 96-well dark plates and the first fluorescence was measured with excitation at 485 nm and emission at 538 nm. Five microliters of Nile red solution (0.1 mg/mL in methanol) were then added into the yeast culture before the measurement of the second fluorescence under the same conditions. The change in fluorescence from the two measurements (ΔF TAG) is correlated with the amount of neutral lipids in the yeast culture. The Nile red values were calculated based on amount of neutral lipids (ΔF TAG) as a function of OD_{600} (ΔF TAG/OD_{600}).

Protein extraction and Western blotting

Microsomal fractions were recovered from the recombinant yeast cells as described previously [30]. In brief, the recombinant yeast cells were collected at the similar OD_{600} values (~7) during the log growth phase if not stated otherwise and then resuspended in lysis buffer containing 20 mM Tris-HCl pH 7.9, 10 mM $MgCl_2$, 1 mM EDTA, 5% (v/v) glycerol, 300 mM ammonium sulfate and 2 mM dithiothreitol before homogenization using a bead beater (Biospec, Bartlesville, OK, USA). The crude homogenate was centrifuged for 30 min at 10 000 g, and the supernatant was further centrifuged at 105 000 g for 70 min to separate the microsomal fractions. The microsomal fractions were resuspended in 3 mM imidazole buffer (pH 7.4) containing 125

mM sucrose. All procedures were carried out at 4ºC. The protein content was determined by Bradford assay using BSA as a standard [42].

For Western blotting, 5 μg of microsomal proteins were separated on 10% SDS-PAGE gel and electrotransferred (overnight at 30 mA and 4ºC) onto polyvinylidene difluoride membrane (Amersham, GE Healthcare, Mississauga, ON, Canada). The membrane was first blocked with 2% ECL prime blocking reagent (Amersham) and then was incubated with V5-HRP-conjugated antibody (Invitrogen). HRP conjugated antibody was detected using ECL Advance Western Blotting Detection Kit (Amersham) with a FluorChem SP imager (Alpha Innotech Corp., San Leandro, CA, USA). The band densities of AtLACS9 variants were quantified using ImageJ software [43].

***In vitro* LACS enzyme assays**

The LACS assay was performed as described previously [23] with slight modifications. In brief, the enzyme assay was carried out in a 60-μL reaction mixture containing 100 mM Bis-Tris-propane (pH 7.6), 10 mM $MgCl_2$, 5 mM ATP, 2.5 mM dithiothreitol, 1 mM CoA, 20 μM [1-^{14}C] oleic acid (56.3 mCi/mmol, PerkinElmer, Waltham, MA, USA) and 2 to 10 μg of microsomal protein. The reaction was initiated by adding microsomal protein and quenched with 10 μL of 10% (w/v) SDS after incubation at 30ºC for 5 min with shaking. The entire reaction mixture was washed 4 times using 900 μL of 50% (v/v) isopropanol saturated hexane for each wash. An aliquot of the aqueous phase was analyzed for radioactivity by a LS 6500 multi-purpose scintillation counter (Beckman-Coulter, Mississauga, ON, Canada). For substrate specificity assay, 20 μM [1-^{14}C] fatty acids, including palmitic acid (60 mCi/mmol, PerkinElmer), stearic acid (58.9 mCi/mmol, American Radiolabeled Chemicals, St. Louis, MO, USA), oleic acid, and linoleic acid (58.2 mCi/mmol, PerkinElmer) were used in the assay.

Sequence alignment, positive selection and protein three-dimensional structure prediction

Forty-five LACS sequences were collected from different species (Supplementary Table S2). Multiple sequence alignment of LACS proteins was performed using ClustalW in MEGA 7 under the default setting [44]. A neighbour-joining with 1000 bootstrap repetitions tree was built using the same software. A web server PAL2NAL (http://www.bork.embl. de/pal2nal/) was then used to construct a multiple codon alignment based on the corresponding aligned amino acid sequences. The output alignment was imported into the jModelTest 2 program [45] to determine the best-fitting evolutionary model. The general time reversible (GTR) model plus Gamma

distribution plus invariant site model of molecular evolution (GTR + G + I) was determined as the best-fit substitution model based on the lowest value of the Akaike Information Criterion. A maximum likelihood phylogenetic tree was then constructed with the PhyML webserver (http://www.atgc-montpellier.fr/phyml/; accessed on 06 Dec 2018) [46,47] according to the best-fit predictive model. The posterior probabilities of sites under positive selection were calculated using CodeML program in the PAML version 4 software [48] based on site-specific Bayes empirical Bayes probabilities [49]. Three sets of models were carried out using the F3X4 codon frequency model, including M0 (one ratio) vs. M3 (discrete); M1 (nearly neutral) vs. M2 (positive selection); and M7 (β) vs. M8 (β + ω). The statistical significance of each pair of nested models was evaluated by the likelihood ratio test (LRT).

The AtLACS9 structure was obtained through homology modeling using PHYRE2 protein fold recognition server [50]. Several homologous structures were identified as possible templates including carboxylic acid reductases (24-25% identity) and acetyl-CoA synthetases (less than 20% identities). A *Nocardia iowensis* carboxylic acid reductase [51] exhibiting 24% identity with AtLACS9 was used as a template to generate a model with 83% sequence coverage and a high confidence level. The following AtLACS9 residues were included in the model: 59-86; 92-320; 328-371; 395-459; 468-525; 530-607; 623-691. A model based on an acetyl-CoA synthetase structure was also obtained to assess the structure. To further verify the quality of the model, I-TASSER was used to predict the 3D structure and the best models from the 2 softwares were assessed and overlaid [52].

Statistical analysis

Data are means ± standard deviation (SD) for the number of independent experiments as indicated. All statistical analyses were performed using the SPSS statistical package (SPSS 16.0, Chicago, IL, USA). Significant differences between two groups were determined using a two-tailed Student's t-test. The equality of variances was determined by the Levene's test. When the variances were equal, the unpaired Student's t-test assuming equal variances was performed. When the variances were unequal, the unpaired Student's t-test with Welch corrections assuming unequal variances was used.

Results

Selection and characterization of active AtLACS9 variants

To select active AtLACS9 variants, randomly mutated *AtLACS9* cDNAs libraries were transformed into *S. cerevisiae* strain *BYfaa1,4Δ* (a double mutant with both *FAA1* and *FAA4* genes knocked out, and it thus contains less than 10% of yeast endogenous LACS activity [18]). This yeast mutant cannot grow on the selection media containing fatty acids and cerulenin (an endogenous fatty acid synthesis inhibitor) due to acyl-CoA deficiency. The growth of yeast cells can be rescued, however, by the introduction of a cDNA encoding an active AtLACS9, which would import the exogenous fatty acids into cells and activate them to acyl-CoAs [23]. The positive colonies grown on the selection plates were cultivated in induction media for 24 h and 48 h and subjected to Nile red assay. The positive colonies were screened based on their abilities to produce neutral lipids as reflected by the Nile red assay. Two colonies were found to produce neutral lipid at levels higher than the yeast expressing wild type (WT) *AtLACS9* (data not shown). The coding sequences of these variant *AtLACS9*s were sequenced, re-cloned into the pYES2.1 vector and transformed into yeast strain *BYfaa1,4Δ* for detailed characterization. Yeast transformed with *AtLACS9* variant cDNAs showed similar growth rates to the yeast harbouring WT *AtLACS9* or *LacZ* control (Supplementary Figure S1). Consistent with the screening results, expression of *AtLACS9* variants in yeast resulted in higher or similar levels of neutral lipid accumulation (ΔF TAG/OD_{600}) relative to yeast expressing WT *AtLACS9* at the early stationary phase (Figure 1). Moreover, yeast cells producing *AtLACS9* and its variants resulted in higher neutral lipid content than the *LacZ* control (Figure 1).

To analyze the production profiles of AtLACS9 variants in yeast mutant *BYfaa1,4Δ*, yeast cells producing AtLACS9 variants were collected periodically from the log to the stationary growth phase, and the corresponding microsomal fractions were prepared for the analyses of *in vitro* LACS activity and protein accumulation by Western blotting. The activity of the recombinant AtLACS9 enzyme and the variants remained at high levels during the log phase, and then decreased after reaching the stationary phase (Figure 2A). AtLACS9 variants displayed the highest activity at the early log phase, whereas the highest activity of WT AtLACS9 occurred at the late log or early stationary phase. Increased LACS activity was observed for variants L12F/C207F/L656F and D238E/P659S. The recombinant AtLACS9 polypeptide accumulation

in the microsomal fraction displaying the highest activity from each variant (Figure 2B) was then analyzed by Western blotting. The AtLACS9 variants displayed different polypeptide accumulation levels in yeast. Variant D238E/P659S had higher polypeptide accumulation while variant L12F/C207F/L656F had lower polypeptide accumulation compared to that of WT AtLACS9 (Figure 2C). After normalizing the enzyme activity to the corresponding protein accumulation [30], both variants displayed 3-fold higher normalized activity relative to the WT enzyme (Figure 2D).

Effect of single site mutations on enzyme activity

Since the two AtLACS9 variants (L12F/C207F/L656F and D238E/P659S) with increased LACS activity contained more than one amino acid residue substitution, the effect of each amino acid residue substitution on enzyme activity was separately determined. Five single site mutants (L12F, C207F, L656F, D238E and P659S) were generated and expressed in yeast mutant *BYfaa1,4Δ*. WT *AtLACS9* and *LacZ* were used as positive and negative controls, respectively. The microsomal fractions containing the recombinant enzymes were used for enzyme assays and Western blotting (Figure 3). Compared to the WT enzyme, increased microsomal enzyme activity and polypeptide accumulation were observed for variants L12F/C207F/L656F and D238E/P659S along with the following variants with single amino acid substitution: C207F, L656F, D238E and P659S (Figure 3A and B). Variant L12F, however, displayed comparable microsomal activity but decreased protein accumulation to those of WT enzyme. The enzyme activity for each variant was then normalized to the corresponding protein accumulation level (Figure 3C) and single site mutants C207F and D238E were found to possess the highest normalized activity, which could have mainly contributed to the increased enzyme activity of the original variants with multiple amino acid residue substitutions.

The substrate specificity of AtLACS9 and its single site variants C207F and D238E was assessed using different radiolabeled fatty acids as substrates (Figure 4). AtLACS9 or its variants was able to utilize all fatty acids tested, with linoleic acid ($18{:}2\Delta^{9cis,\,12cis}$) being the most effective substrate in each case. No significant differences in substrate preference, however, were observed for AtLACS9 and its variants.

Multiple sequence alignment, positive selection prediction for LACS proteins and structure prediction

Given that several amino acid residue substitutions in AtLACS9 were shown to affect AtLACS9 activity (Figures 2 and 3), it is useful to further explore the relationship between the identified beneficial amino acid residue and the putative amino acid residue sites with functional importance in LACS proteins. In this regard, sequence-based approaches, including multiple sequence alignment and positive selection prediction were performed for various LACS proteins. Multiple sequence alignment is an effective approach to identify conserved functional motifs and subfamily specific positions. In addition to the conserved sites, some unconserved sites, such as positively selected sites, may also affect protein function. Positive selection or Darwinian selection is considered to drive the sweep and fixation of the advantageous mutations throughout a population [53], and thus may have crucial roles in the evolution of protein function [54]. To test the presence of positive selection in the LACS sequence, the site-specific non-synonymous (dN) to synonymous (dS) substitutions ratio (dN/dS or ω) test was conducted by using three sets of models (M0 versus M3, M1 versus M2, and M7 versus M8) from the PAML version 4 software [48]. The likelihood ratio test (LRT) of the comparison between the model pair of M1 (null and neutral) versus M2 (selection) did not give a significant result to reject the null hypothesis of neutral selection (Table 1). However, the comparison between the model pairs, M0 (null and neutral) versus M3 (selection), and M7 (null and neutral) versus M8 (selection) yielded the LRT statistics of 7909.3 and 783.0, respectively, suggesting that certain sites were indeed under selective pressures in LACS proteins (Table 1). No positively selected sites were detected from the model M3 ($\omega2=0.89758$), whereas in total 63 amino acid residues were identified as sites of positive selection from the model M8 ($\omega=1.22349$) using Bayes empirical Bayes (BEB) analysis [49].

By mapping the detected positively selected sites and putative functional motifs [33,55,56] along the aligned sequences of LACS (Figure 5A), it became apparent that the positively selected sites were mainly located at the N- and C- termini of the enzymes, whereas no sites on the putative functional motifs were observed under positive selection. The beneficial amino acid residue substitutions in the identified AtLACS9 variants were further compared with the positively selected sites and putative functional motifs. Most of the beneficial amino acid residue substitutions were found to reside in the less conserved regions. L12 and D238 were predicted as positive selection sites despite that the posterior probabilities of both sites are only higher than 50%. Furthermore, the phylogenetic analysis revealed that AtLACS9-C207 is highly

conserved among the LACS9 from plant eudicots, whereas AtLACS9-D238 is more divergent and the substitution of D238E exists naturally in other LACS sequences (Figure 5B).

A three-dimensional structure of AtLACS9 was then obtained to further map the two identified beneficial amino acid residue substitution sites (Figure 5C). The best template was identified as a carboxylic acid reductase [51], which is a soluble enzyme. AtLACS9, on the other hand, has been experimentally demonstrated to reside in the envelope of the chloroplast [6, 16]. The first 20 amino acid residues of AtLACS9 are predicted to constitute a membrane-spanning segment [23] despite that the prediction results varied among different programs (Supplementary Figure S2A and S2B). Indeed, the predicted membrane-associated nature agrees with its microsomal localization in yeast (Supplementary Figure S2C) and chloroplastidial localization in Arabidopsis [6, 16]. Since both the experimental results [6, 16] and the TargetP1.1 prediction (http://www.cbs.dtu.dk/services/TargetP/; accessed on 06 Dec 2018) suggest the chloroplastic subcellular localization of AtLACS9, the protein sequence of AtLACS9 was subjected to the signal (chloroplast transit) peptides prediction using ChloroP 1.1 (http://www.cbs.dtu.dk/services/ChloroP/; accessed on 06 Dec 2018) and SOSUIsignal (http://harrier.nagahama-i-bio.ac.jp/sosui/sosuisignal/sosuisignal_submit.html; accessed on 06 Dec 2018). Although the N-terminal region (19 amino acid residues) of AtLACS9 is predicted as a signal peptide by SOSUIsignal, no chloroplast transit peptides are predicted from ChloroP 1.1. Therefore, the N-terminal fragment with the putative transmembrane domain was not included in the model as this section did not exhibit any homology to the carboxylic acid reductase and was thus separately added to the structure to show localization in the membrane. The identified beneficial mutation sites are shown in red, whereas the putative ATP and fatty acid binding motifs in the AtLACS9 model structure are shown in blue and green, respectively. These substrate binding sites are close to one another, suggesting that these sites may be able to facilitate the transfer of an AMP moiety from ATP to the carboxylate group of fatty acids. The identified beneficial sites, however, are present at distal sites relative to the putative substrate binding sites, indicating that these mutations do not directly affect substrate binding. Although the AtLACS9 sequence exhibits only 20% sequence identity with an acetyl-CoA synthetase with a reported three-dimensional structure, the homology structure of AtLACS9 using this as a template also gave similar orientations of the putative substrate binding sites and the beneficial sites for mutagenesis (Supplementary Figure S3). Furthermore, results of I-TASSER modeling

confirmed that the two models predicted from different software programs have a similar overall fold as shown by the overlaid structures (Supplementary Figure S4).

Discussion

The current study reports on the generation of performance-enhanced variants of AtLACS9 using protein engineering. To improve the enzyme performance of AtLACS9, we introduced random mutations into the coding sequence by error-prone PCR, transformed the mutagenized *AtLACS9* into the yeast mutant *BYfaa1,4Δ*, and screened for the active AtLACS9 variants by identifying positive clones on selection media followed by fluorescence detection of neutral lipid accumulation in yeast cells. After screening, two AtLACS9 variants were identified which slightly increased neutral lipid accumulation in yeast cells (Figure 1) and thus were selected for further characterization by analyzing *in vitro* LACS activity and polypeptide accumulation using yeast microsomal proteins. These enzyme variants were found to display increased yeast microsomal activity and altered polypeptide accumulation (Figures 2 and 3). To eliminate the influence of differences in protein abundance, the microsomal activity of each variant was then normalized to the corresponding polypeptide accumulation level. The two LACS variants had higher normalized activity compared to that of the WT enzyme (Figure 2D). Since the two activity-improved variants (L12F/C207F/L656F and D238E/P659S) contained more than one amino acid residue substitution, recombinant enzymes with single amino acid residue substitution were generated using site-directed mutagenesis to further elucidate the contribution of each substitution. The single site mutants C207F and D238E were considered mainly responsible for the increased enzyme activity, although amino acid residue substitutions of L12F and P659S also led to increases in enzyme activity to some extent (Figure 3).

To interpret the effects of amino acid residue substitutions in the two AtLACS9 variants, the substituted amino acid residue sites from each variant and the predicted positively selected sites were mapped onto the multiple-aligned sequences of LACS9 proteins (Figure 5A), since both moderately conserved sites (e.g. subfamily specific positions) [57] and unconserved sites (particularly positively selected sites) [54] appear to have crucial roles in affecting protein function. The beneficial amino acid residue substitution of C207F is at a moderately conserved site, whereas two additional activity-improved variants (L12F and D238E) were found with amino acid residue substitutions at the predicted positively selected sites (Figure 5A). Indeed, the

substitution of D238E is naturally present in LACS from other plant species (Figure 5B). As for the substitution of C207F, the change from the polar C residue to the non-polar (aromatic) F residue appears to be dramatic, especially considering the capability of formation of a disulfide bond between two cysteine residues, the replacement of which might lead to changes in the tertiary structure. It should be noted, however, that although C207 is conserved among LACS9 from eudicots, a non-polar and aromatic Y residue is also found at that position of LACS from many other species. Furthermore, PSIPRED analysis predicts C207 and D238E to be part of β-sheet and loop region, respectively, in agreement with the PHYRE modeling results. C207F mutation may lead to secondary structure changes whereas D238E may lead to changes in loop region (Supplementary Figure S5).

It is possible that the increased enzyme activity was caused by a more favorable conformation in support of catalysis, such as improved affinity to the substrates. However, mapping of the beneficial amino acid residue substitution sites onto the predicted three-dimensional structure of AtLACS9 further revealed that these beneficial sites are apparently remote from the putative substrate binding sites (Figure 5C). Indeed, beneficial substitutions at residues far away from the putative functional sites are not unusual in variants generated via random mutagenesis, although these sites are easily overlooked in rational design [40,41]. For example, directed evolution of diacylglycerol acyltransferase 1 from soybean (*Glycine max*) [24] and canola-type *Brassica napus* [30,25] also identified several amino acid residue substitutions affecting enzyme activity which are far away from the enzyme's functional sites. Mutations in these distal sites may influence enzyme conformation or regulatory sites resulting in more active conformations. Regulatory sites can be found distal from active site motifs. For example, it was recently shown that several amino acid residues substitutions within the hydrophilic N-terminal domain of *B.napus* diacylglycerol acyltransferase 1 could increase enzyme activity [25]. This N-terminal domain represents a regulatory domain that is located away from the putative active sites near the C-terminal region [58]. Though far from the putative catalytic sites, amino acid residues substitution within the acyltransferase's regulatory domain resulted in significant changes in enzyme activity [25]. It should also be noted that the beneficial substitutions may affect enzyme stability and thus lead to increased enzyme activity. Nonetheless, it would be worthwhile to further explore the effects of the beneficial substitutions on the enzyme's conformation and stability.

418 The observed fatty acid specificity of microsomal AtLACS9 (Figure 4) is generally
419 consistent with previously reported substrate specificity of AtLACS9 [3]. The only difference is
420 that AtLACS9 was observed to prefer linoleic acid slightly more than oleic acid in this study,
421 whereas Shockey et al. observed AtLACS9 showed a slightly higher preference towards oleic
422 acid over linoleic acid [3]. It should be noted that different sources of recombinant AtLACS9
423 were used in the assays in these two studies. Shockey et al. used the *E. coli* mutant strain K27 to
424 produce recombinant AtLACS9, which provided a clean background for LACS assay due to the
425 complete loss of endogenous LACS activity in *E. coli* K27. In our study, *AtLACS9* and its
426 variants were expressed in yeast stain *BYfaa1,4*Δ, which contains less than 10% LACS enzymes
427 in yeast background [18], and as a result, the control microsomes (LacZ) displayed a low level of
428 LACS activity (see Figures 2A and 3A). The low level of LACS activity from the yeast
429 background would therefore only have a limited influence on substrate specificity data depicted
430 in Figure 4. Recently, AtLACS9 was shown to be involved in establishing a specific linoleoyl-
431 CoA pool, which is connected to lipid trafficking from the PC to the plastid [16]. Indeed, in the
432 current study, recombinant AtLACS9 produced in yeast exhibited an enhanced specificity for
433 linoleic acid relative to palmitic, stearic and oleic acid (Figure 4).
434 As the only LACS associated with the outer envelope of the plastid, AtLACS9 appears to
435 function in the activation of free fatty acids and transport of fatty acyl moieties between the
436 plastid and extra-plastidial compartment [15,16]. Although the direction of AtLACS9 mediated
437 transport of fatty acyl moieties is still debatable, increasing AtLACS9 activity through metabolic
438 engineering could potentially shed light on the specific contribution of AtLACS9. In addition,
439 AtLACS9 variants with enhanced enzyme activity may also provide potential candidates for
440 engineering of oleginious organisms to produce fatty acid-derived compounds, including TAG.
441 Indeed, introduction of various LACS from diatoms or higher plants into yeast has been shown
442 to increase oil deposition [18–21]. The current study also suggested that the introduction of
443 activity-improved AtLACS9 variants led to a slight increase in the yeast neutral lipid
444 accumulation (Figure 1). Combined heterologous expression of an *AtLACS9* variant with cDNA
445 encoding one or more of the acyl-CoA-dependent acyltransferases of the Kennedy pathway
446 involved in seed oil biosynthesis (see Chapman and Ohlrogge, 2012) [11] may even result in
447 greater accumulation of TAG in yeast. A similar metabolic engineering strategy might also be
448 explored as means of increasing the TAG content of seeds or vegetative tissue. Furthermore, the

beneficial amino acid substitutions of AtLACS9 provide valuable information for systematically engineering LACS from other species since these amino acid substitutions are at sites conserved among different species (Figure 5B).

In conclusion, two activity-enhanced variants of AtLACS9 were developed via random mutagenesis combined with site-directed mutagenesis. The recombinant AtLACS9 variants produced in yeast were characterized by analysis of *in vitro* enzyme activity and polypeptide accumulation in microsomes. The beneficial amino acid substitution sites were further analyzed by sequence and structural analyses. Amino acid residue substitution at the moderately conserved site C207 and the predicted positively selected site D238 were responsible for the increases in enzyme activity of corresponding enzyme variants. These findings provide valuable information for improving LACS enzymes from plants which may be useful in the *in vivo* alteration of acyl-CoA pools.

Conflict of interest

The authors declare that they have no conflicts of interest with the content of this article.

Funding information

The research was supported by Genome Canada, Genome Prairie, Dow AgroScience (Corteva Agriscience, Agriculture division of DowDupont), Alberta Innovates Bio Solutions (R.J.W.), the Natural Science and Engineering Research Council of Canada (NSERC) Discovery Grants (Discovery grant number 163306 to J.O.; RGPIN-2016-05926 to G.C.; and RGPIN-2014-04585 to R.J.W.) and the Canada Research Chairs Program (R.J.W. and G.C.).

Author contributions

RJW, YX and GC designed the research; GC and RJW supervised the experiments; YX performed most of the experiments, analyzed the data and drafted the manuscript. KMPC performed the analysis of the predicted 3D structure of AtLACS9. KMPC, RH, EM, JO, GC, and SMR contributed valuable discussion during this study. All co-authors contributed in further revising the manuscript.

Table 1. Parameter estimates and likelihood scores of LACS9 for site models. Positive selection by site models was performed using CODEML program in PAML. The number of positively selected sites is also shown, with the Bayes empirical Bayes (BEB) posterior probability in blankets. df, degrees of freedom; LRT, likelihood ratio test; lnL, log likelihood scores; 2ΔlnL, twice the log-likelihood difference of the models compared.

Model	Estimates of parameters	lnL	LRT pairs	df	2ΔlnL	p-value	Positively selected sites
M0: one ratio	ω= 0.19727	-63498.3	M0/M3	4	7909.3	<0.0001	none
M3: discrete	p0=0.82713, p1=0.15115, p2=0.02172, ω0=0.01143, ω1=0.18961, ω2=0.89758	-59543.7					
M1: nearly neutral	p0= 0.96980, p1=0.03020, ω0=0.06927, ω1=1.00000	-60172.6	M1/M2	2	0	1.00000	none
M2: positive selection	p0=0.96980, p1=0.02844, p2=0.00176, ω0=0.06927, ω1=1.00000, ω2=1.00000	-60172.6					
M7: β	p=0.08048, q=0.83695	-59838.9	M7/M8	2	783.0	<0.0001	30 sites (>50%)
M8: β + ω	p0=0.99029, p=0.14171, q=2.31642, p1=0.00971, ω=1.22349	-59447.4					15 sites (>95%) 18 sites (>99%)

Reference:

1 Watkins, P. A. (1997) Fatty acid activation. Prog. Lipid Res. **36**, 55–83.

2 Kornberg, A. and Pricer, W. E. (1953) Enzymatic synthesis of the coenzyme A derivatives of long chain fatty acids. J. Biol. Chem. **204**, 329–343.

3 Shockey, J. M., Fulda, M. S. and Browse, J. A. (2002) Arabidopsis contains nine long-chain acyl-coenzyme A synthetase genes that participate in fatty acid and glycerolipid metabolism. Plant Physiol. **129**, 1710–1722.

4 Fulda, M., Schnurr, J., Abbadi, A., Heinz, E. and Browse, J. (2004) Peroxisomal Acyl-CoA synthetase activity is essential for seedling development in *Arabidopsis thaliana*. Plant Cell **16**, 394–405.

5 Fulda, M., Shockey, J., Werber, M., Wolter, F. P. and Heinz, E. (2002) Two long-chain acyl-CoA synthetases from *Arabidopsis thaliana* involved in peroxisomal fatty acid beta-oxidation. Plant J. **32**, 93–103.

6 Schnurr, J., Shockey, J. and Browse, J. (2004) The Acyl-CoA synthetase encoded by LACS2 is essential for normal cuticle development in Arabidopsis. Plant Cell **16**, 629–642.

7 Tang, D., Simonich, M. T. and Innes, R. W. (2007) Mutations in LACS2, a long-chain acyl-coenzyme A synthetase, enhance susceptibility to avirulent *Pseudomonas syringae* but confer resistance to *Botrytis cinerea* in Arabidopsis. Plant Physiol. **144**, 1093–1103.

8 Lü, S., Song, T., Kosma, D. K., Parsons, E. P., Rowland, O. and Jenks, M. A. (2009) Arabidopsis *CER8* encodes LONG-CHAIN ACYL-COA SYNTHETASE 1 (LACS1) that has overlapping functions with LACS2 in plant wax and cutin synthesis. Plant J. **59**, 553–564.

9 Weng, H., Molina, I., Shockey, J. and Browse, J. (2010) Organ fusion and defective cuticle function in a lacs1 lacs2 double mutant of Arabidopsis. Planta **231**, 1089–1100.

10 Jessen, D., Olbrich, A., Knüfer, J., Krüger, A., Hoppert, M., Polle, A. and Fulda, M. (2011) Combined activity of LACS1 and LACS4 is required for proper pollen coat formation in Arabidopsis. Plant J. **68**, 715–726.

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

12 Roughan, P. G. and Slack, C. R. (1977) Long-chain acyl-coenzyme A synthetase activity of spinach chloroplasts is concentrated in the envelope. Biochem. J. **162**, 457–459.

13 Andrews, J. and Keegstra, K. (1983) Acyl-CoA synthetase is located in the outer membrane and acyl-CoA thioesterase in the inner membrane of pea chloroplast envelopes. Plant Physiol. **72**, 735–740.

14 Block, M. A., Joyard, J. and Deuce, R. (1983) The acyl-CoA synthetase and acyl-CoA thioesterase are located on the outer and inner membrane of the chloroplast envelope, respectively. FEBS Lett. **153**, 377–381.

15 Zhao, L., Katavic, V., Li, F., Haughn, G. W. and Kunst, L. (2010) Insertional mutant analysis reveals that long-chain acyl-CoA synthetase 1 (LACS1), but not LACS8, functionally overlaps with LACS9 in Arabidopsis seed oil biosynthesis. Plant J. **64**, 1048–1058.

16 Jessen, D., Roth, C., Wiermer, M. and Fulda, M. (2015) Two activities of long-chain acyl-CoA synthetase are involved in lipid trafficking between the endoplasmic reticulum and the plastid in Arabidopsis. Plant Physiol. **167**, 351–366.

17 Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., del Cardayre, S. B. and Keasling, J. D. (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature **463**, 559–562.

18 Guo, X., Jiang, M., Wan, X., Hu, C. and Gong, Y. (2014) Identification and biochemical characterization of five long-chain acyl-coenzyme A synthetases from the diatom *Phaeodactylum tricornutum*. Plant Physiol. Biochem. **74**, 33–41.

19 Tan, X., Zheng, X., Zhang, Z., Wang, Z., Xia, H., Lu, C. and Gu, S. (2014) Long chain acyl-coenzyme A synthetase 4 (BnLACS4) gene from Brassica napus enhances the yeast lipid contents. J. Integr. Agric. **13**, 54–62.

20 Tonon, T., Qing, R., Harvey, D., Li, Y., Larson, T. R. and Graham, I. A. (2005) Identification of a long-chain polyunsaturated fatty acid acyl-coenzyme A synthetase from the diatom *Thalassiosira pseudonana*. Plant Physiol. **138**, 402–408.

21 Pulsifer, I. P., Kluge, S. and Rowland, O. (2012) Arabidopsis long-chain acyl-CoA synthetase 1 (LACS1), LACS2, and LACS3 facilitate fatty acid uptake in yeast. Plant Physiol. Biochem. **51**, 31–39.

22 Xu, N., Zhang, S. O., Cole, R. A., McKinney, S. A., Guo, F., Haas, J. T., Bobba, S., Farese, R. V and Mak, H. Y. (2012) The FATP1-DGAT2 complex facilitates lipid droplet expansion at the ER-lipid droplet interface. J. Cell Biol. **198**, 895–911.

549 23 Xu, Y., Holic, R., Li, D., Pan, X., Mietkiewska, E., Chen, G., Ozga, J. and Weselake, R. J.
550 (2018) Substrate preferences of long-chain acyl-CoA synthetase and diacylglycerol
551 acyltransferase contribute to enrichment of flax seed oil with α-linolenic acid. Biochem. J.
552 **475**, 1473–1489.
553 24 Roesler, K., Shen, B., Bermudez, E., Li, C., Hunt, J., Damude, H. G., Ripp, K. G., Everard,
554 J. D., Booth, J. R., Castaneda, L., et al. (2016) An improved variant of soybean type 1
555 diacylglycerol acyltransferase increases the oil content and decreases the soluble
556 carbohydrate content of soybeans. Plant Physiol. **171**, 878–893.
557 25 Chen, G., Xu, Y., Siloto, R. M. P., Caldo, K. M. P., Vanhercke, T., Tahchy, A. El, Niesner,
558 N., Chen, Y., Mietkiewska, E. and Weselake, R. J. (2017) High performance variants of
559 plant diacylglycerol acyltransferase 1 generated by directed evolution provide insights into
560 structure-function. Plant J. **92**, 167-177.
561 26 Kamisaka, Y., Kimura, K., Uemura, H. and Shibakami, M. (2010) Activation of
562 diacylglycerol acyltransferase expressed in *Saccharomyces cerevisiae*: overexpression of
563 *Dga1p* lacking the N-terminal region in the Deltasnf2 disruptant produces a significant
564 increase in its enzyme activity. Appl. Microbiol. Biotechnol. **88**, 105–15.
565 27 Xu, J., Francis, T., Mietkiewska, E., Giblin, E. M., Barton, D. L., Zhang, Y., Zhang, M.
566 and Taylor, D. C. (2008) Cloning and characterization of an acyl-CoA-dependent
567 diacylglycerol acyltransferase 1 (DGAT1) gene from *Tropaeolum majus*, and a study of
568 the functional motifs of the DGAT protein using site-directed mutagenesis to modify
569 enzyme activity and oil content. Plant Biotechnol. J. **6**, 799–818.
570 28 Greer, M. S., Truksa, M., Deng, W., Lung, S. C., Chen, G. and Weselake, R. J. (2015)
571 Engineering increased triacylglycerol accumulation in *Saccharomyces cerevisiae* using a
572 modified type 1 plant diacylglycerol acyltransferase. Appl. Microbiol. Biotechnol. **99**,
573 2243–2253.
574 29 O'Quin, J. B., Mullen, R. T. and Dyer, J. M. (2009) Addition of an N-terminal epitope tag
575 significantly increases the activity of plant fatty acid desaturases expressed in yeast cells.
576 Appl. Microbiol. Biotechnol. **83**, 117–125.
577 30 Xu, Y., Chen, G., Greer, M. S., Caldo, K. M. P., Ramakrishnan, G., Shah, S., Wu, L.,
578 Lemieux, M. J., Ozga, J. and Weselake, R. J. (2017) Multiple mechanisms contribute to
579 increased neutral lipid accumulation in yeast producing recombinant variants of plant

580 diacylglycerol acyltransferase 1. J. Biol. Chem. **292**, 17819–17831.
581 31 Belhaj, K., Chaparro-Garcia, A., Kamoun, S. and Nekrasov, V. (2013) Plant genome
582 editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas
583 system. Plant Methods **9**, 39.
584 32 Kochan, G., Pilka, E. S., von Delft, F., Oppermann, U. and Yue, W. W. (2009) Structural
585 snapshots for the conformation-dependent catalysis by human medium-chain acyl-
586 coenzyme A synthetase ACSM2A. J. Mol. Biol. **388**, 997–1008.
587 33 Hisanaga, Y., Ago, H., Nakagawa, N., Hamada, K., Ida, K., Yamamoto, M., Hori, T., Arii,
588 Y., Sugahara, M., Kuramitsu, S., et al. (2004) Structural basis of the substrate-specific
589 two-step catalysis of long chain fatty acyl-CoA synthetase dimer. J. Biol. Chem. **279**,
590 31717–31726.
591 34 Andersson, C. S., Lundgren, C. A. K., Magnúsdóttir, A., Ge, C., Wieslander, Å., Molina,
592 D. M. and Högbom, M. (2012) The *Mycobacterium tuberculosis* very-long-chain fatty
593 acyl-CoA synthetase: Structural basis for housing lipid substrates longer than the enzyme.
594 Structure **20**, 1062–1070.
595 35 Siloto, R. M. P., Truksa, M., Brownfield, D., Good, A. G. and Weselake, R. J. (2009)
596 Directed evolution of acyl-CoA:diacylglycerol acyltransferase: development and
597 characterization of *Brassica napus DGAT1* mutagenized libraries. Plant Physiol. Biochem.
598 **47**, 456–461.
599 36 Baek, S. C., Ho, T.-H., Lee, H. W., Jung, W. K., Gang, H.-S., Kang, L.-W., Kim, H., Lu,
600 C., Napier, J. A., Clemente, T. E., et al. (2017) Improvement of enzyme activity of β-1,3-
601 1,4-glucanase from *Paenibacillus* sp. X4 by error-prone PCR and structural insights of
602 mutated residues. Appl. Microbiol. Biotechnol. 4073–4083.
603 37 Porter, J. L., Boon, P. L. S., Murray, T. P., Huber, T., Collyer, C. A. and Ollis, D. L. (2014)
604 Directed evolution of new and improved enzyme functions using an evolutionary
605 intermediate and multi-directional search. ACS Chem. Biol. **10**, 611–621.
606 38 Ford, T. J. and Way, J. C. (2015) Enhancement of *E. coli* acyl-CoA synthetase FadD
607 activity on medium chain fatty acids. PeerJ **3**, e1040.
608 39 McCullum E.O., Williams B.A.R., Zhang J., Chaput J.C. (2010) Random Mutagenesis by
609 Error-Prone PCR. In: Braman J. (eds) In Vitro Mutagenesis Protocols. Methods in
610 Molecular Biology (Methods and Protocols), vol 634. Humana Press, Totowa, NJ.40

Meyer, A., Schmid, A., Held, M., Westphal, A. H., Rothlisberger, M., Kohler, H.-P. E., van Berkel, W. J. H. and Witholt, B. (2002) Changing the substrate reactivity of 2-hydroxybiphenyl 3-monooxygenase from *Pseudomonas azelaica* HBP1 by directed evolution. J. Biol. Chem. **277**, 5575–5582.

41 Fortin, P. D., MacPherson, I., Neau, D. B., Bolin, J. T. and Eltis, L. D. (2005) Directed evolution of a ring-cleaving dioxygenase for polychlorinated biphenyl degradation. J. Biol. Chem. **280**, 42307–42314.

42 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.

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

44 Kumar, S., Stecher, G. and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. **33**, msw054.

45 Santorum, J. M., Darriba, D., Taboada, G. L. and Posada, D. (2014) Jmodeltest.org: Selection of nucleotide substitution models on the cloud. Bioinformatics **30**, 1310–1311.

46 Guindon, S., Lethiec, F., Duroux, P. and Gascuel, O. (2005) PHYML Online - A web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res. **33**, 557–559.

47 Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Asessing the performance of PhyML 2.0. Syst. Biol. **59**, 307–321.

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

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

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

51 Gahloth, D., Dunstan, M. S., Quaglia, D., Klumbys, E., Lockhart-Cairns, M. P., Hill, A. M., Derrington, S. R., Scrutton, N. S., Turner, N. J. and Leys, D. (2017) Structures of carboxylic acid reductase reveal domain dynamics underlying catalysis. Nat. Chem. Biol.

13, 975–981.

52 Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. (2015) The I-TASSER Suite: protein structure and function prediction. Nature methods. **12**:7.

53 Biswas, S. and Akey, J. M. (2006) Genomic insights into positive selection. Trends Genet. **22**, 437–446.

54 Yuan, H., Wu, J., Wang, X., Chen, J., Zhong, Y., Huang, Q. and Nan, P. (2017) Computational identification of amino-acid mutations that further improve the activity of a chalcone-flavonone osomerase from *Glycine max*. Front. Plant Sci. **8**, 1–8.

55 Black, P. N., Zhang, Q., Weimar, J. D. and DiRusso, C. C. (1997) Mutational analysis of a fatty acyl-coenzyme A synthetase signature motif identifies seven amino acid residues that modulate fatty acid substrate specificity. J. Biol. Chem. **272**, 4896–4903.

56 Weimar, J. D., DiRusso, C. C., Delio, R. and Black, P. N. (2002) Functional role of fatty acyl-coenzyme A synthetase in the transmembrane movement and activation of exogenous long-chain fatty acids. J. Biol. Chem. **277**, 29369–29376.

57 Suplatov, D. A., Besenmatter, W., Švedas, V. K. and Svendsen, A. (2012) Bioinformatic analysis of alpha /beta-hydrolase fold enzymes reveals subfamily-specific positions responsible for discrimination of amidase and lipase. Protein Eng. Des. Sel. **25**, 689–697.

58 Caldo, K. M. P., Acedo, J. Z., Panigrahi, R., Vederas, J. C., Weselake, R. J. and Lemieux, M. J. (2017) Diacylglycerol acyltransferase 1 is regulated by its N-terminal domain in response to allosteric effectors. Plant Physiol. **175**, 667–680.

59 Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. and Barton, G. J. (2009) Jalview Version 2-A multiple sequence alignment editor and analysis workbench. Bioinformatics **25**, 1189–1191.

667 **Figure legends**

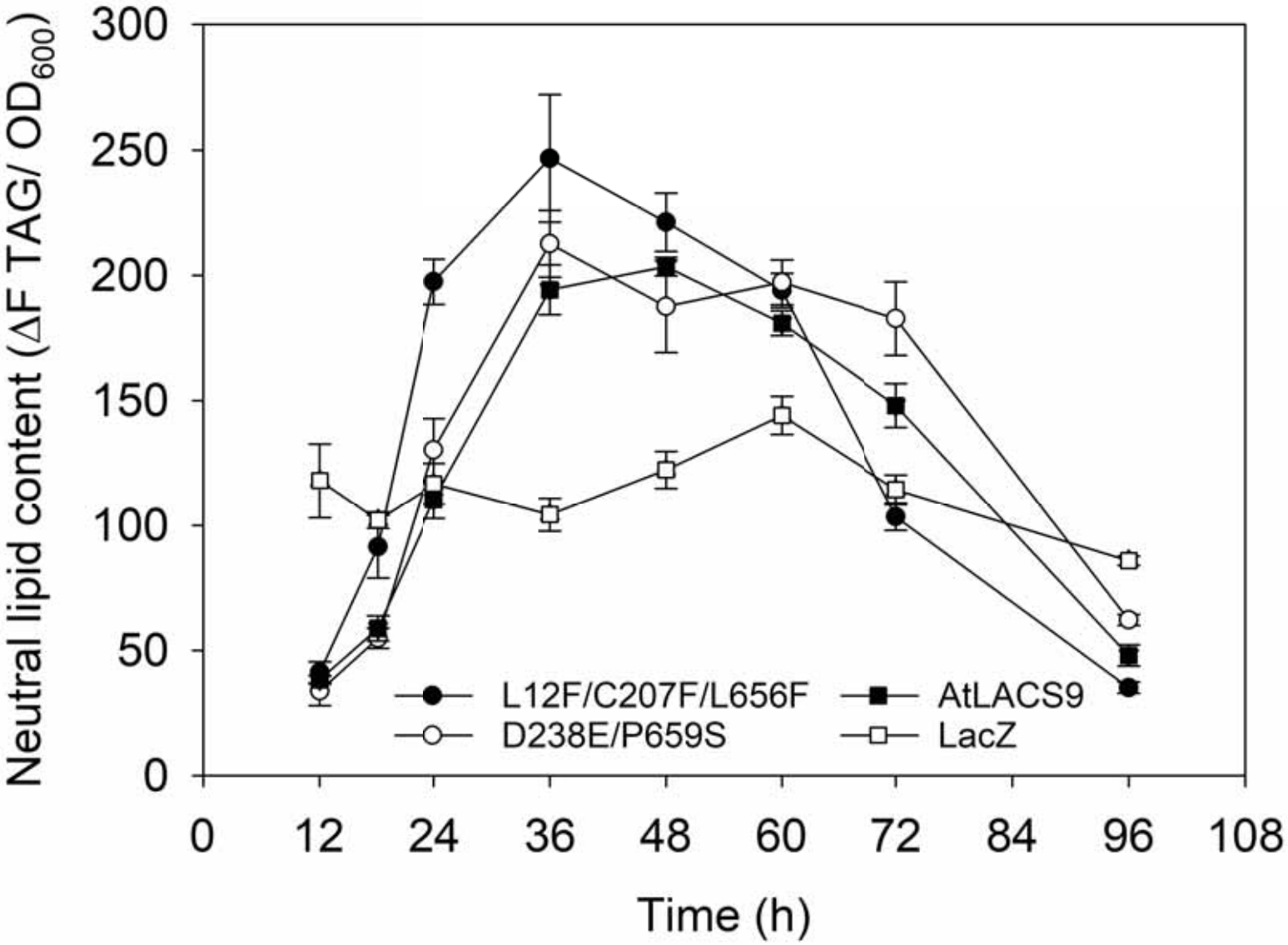
668 Figure 1. Neutral lipid content of yeast producing AtLACS9 variants. Neutral lipid content was
669 analyzed using the Nile red assay and the values are calculated based on the Nile red
670 fluorescence (ΔF TAG) as a function of the optical density (OD_{600}) at 600 nm (ΔF TAG/OD_{600}).
671 Data are means ± SD, n = 3.
672
673
674 Figure 2. Characterization of AtLACS9 variants. A, *In vitro* LACS activities of different
675 AtLACS9 variants. Microsomal fractions from yeasts producing recombinant AtLACS9 variants
676 were harvested at different time points after induction and used for the enzyme assay. The
677 growth curve was monitored by measuring OD_{600}. B, Relative enzyme activities of AtLACS9
678 variants. The highest activity of each variant is shown, with the wild type (WT) AtLACS9
679 activity set as 1.0. C, Relative protein abundance. Five micrograms of microsomal protein from
680 the same batch of microsomes used to assess enzyme activity were used for Western blotting
681 analysis. The relative protein accumulation of recombinant WT AtLACS9 was set as 1.0. D, The
682 normalized relative activity of each enzyme variant was obtained by dividing the enzyme activity
683 value by relative protein abundance, with recombinant WT AtLACS9 activity set as 1.0. For A,
684 B, C and D, data are means ± SD; n = 2 for A, n=4 for B, n = 3 for C and D. The asterisks
685 indicate significant differences in activity (B) and protein abundance (C) of the microsomes
686 containing recombinant AtLACS9 variants versus recombinant WT AtLACS9 (t-test, ** $P<0.01$,
687 * $P<0.5$). ND, not determined.
688
689 Figure 3. Enzyme activity and corresponding protein abundance of AtLACS9 single site mutants.
690 A, Relative enzyme activities of single site mutants. The wild type (WT) AtLACS9 activity was
691 set as 1.0. B, Relative protein abundance. Five micrograms of microsomal protein from the same
692 batch of microsomes used to assess enzyme activity were used for Western blotting analysis. The
693 relative abundance of recombinant WT AtLACS9 was set as 1.0. C, The normalized relative
694 activity of each mutant was obtained by dividing the enzyme activity value by relative protein
695 accumulation, with recombinant WT AtLACS9 activity set as 1.0. For A, B and C, data are
696 means ± SD, n = 3. The asterisks indicate significant differences in activity (A), and protein

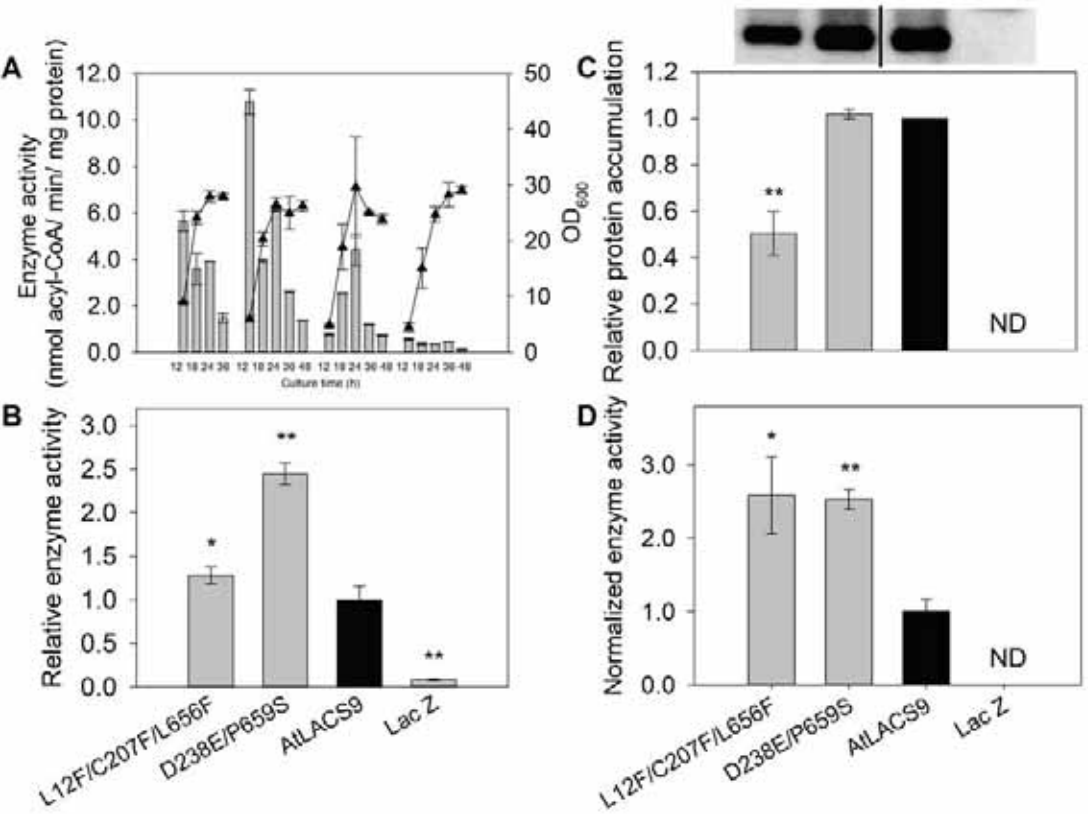
abundance (B) of the microsomes containing recombinant AtDGAT9 variants versus recombinant WT AtLACS9 (t-test, ** P<0.01, * P < 0.05). ND, not determined.

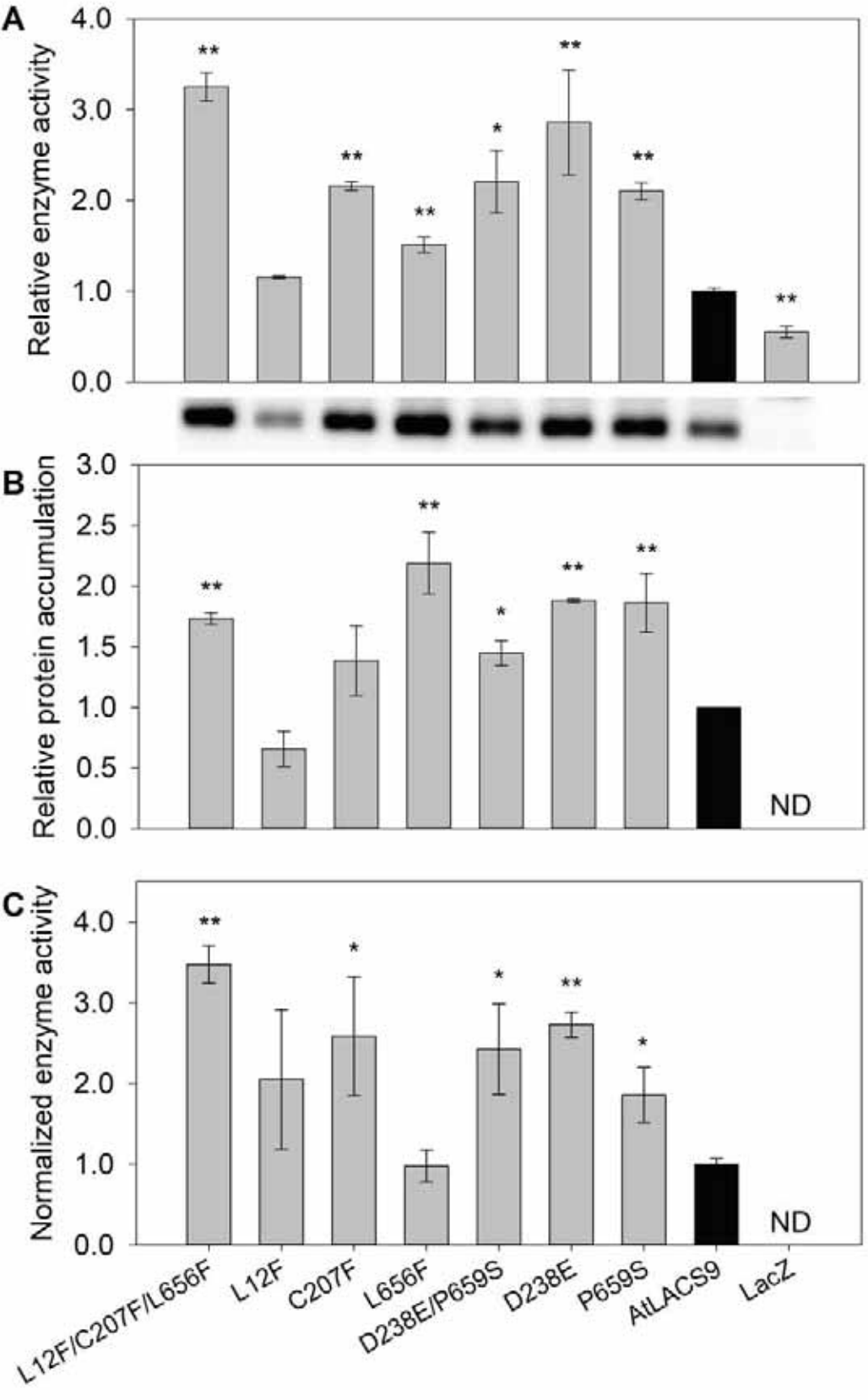
Figure 4. Substrate specificity of AtLACS9 variants. Enzyme activity data were normalized to activity observed using oleic acid (18:1Δ^{9cis}) as the substrate (i.e., oleic acid supported activity was set at 100%). The microsomal activities of AtLACS9, C207F, and D238E were 4.92±0.09, 6.06±0.25, 8.17±0.14 nmol [^{14}C] oleoyl-CoA/ min/ mg protein, respectively. Microsomal preparations from the yeast mutant *BYfaa1,4*Δ producing AtLACS9 variants were used for analysis of enzyme assay. Data represent means ± SD, n = 3. 16:0, palmitic acid; 18:0, stearic acid; 18:1, oleic acid; 18:2, linoleic acid (18:2$\Delta^{9cis, 12cis}$).

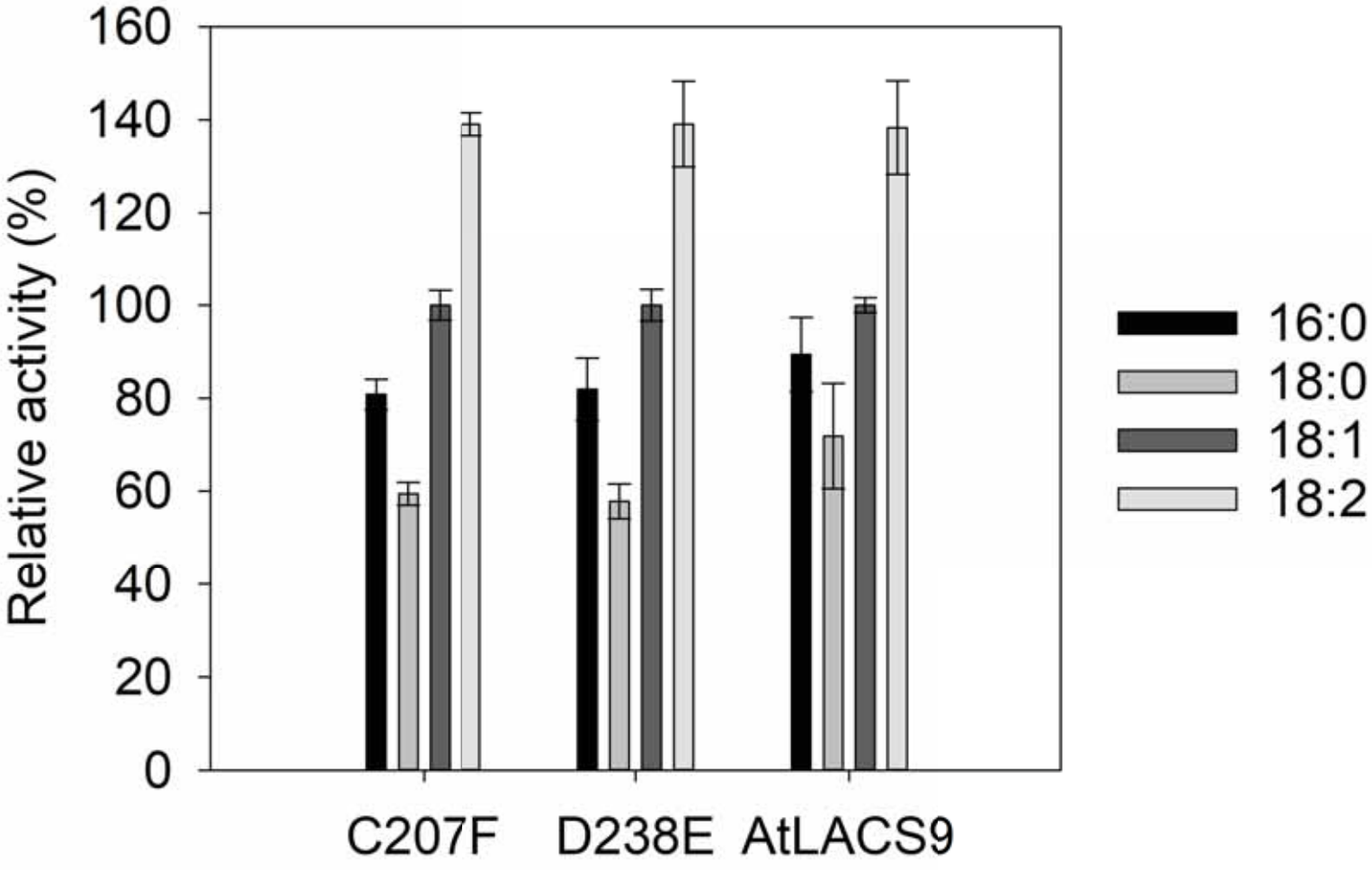
Figure 5. Sequence analysis and predicted structure of AtLACS9. A, Sequence alignment of LACS9 protein from seven typical plant species. Conserved sites are shaded. Positively selected sites with a Bayes Empirical Bayes posterior probability higher than (≥) 50%, higher than (≥) 95%, and higher than (≥) 99% are indicated by the amino acid sites in green, blue, and red background, respectively. The single mutation sites are indicated by red-filled triangle. The bar above the sequence corresponds to the ATP/AMP signature motifs (I & II) and the fatty acyl-CoA synthetase signature motif (III). The putative active sites are indicated by black-filled star. The alignment was visualized and displayed using Jalview [59]. B, Amino acid sequence analysis of LACS proteins from different species. *Ah, Arachis hypogaea; At, Arabidopsis thaliana; Bd, Brachypodium distachyon; Bn, Brassica napus; Cas, Camelina sativa; Cs, Cucumis sativus; Eg, Elaeis guineensis; Fv, Fragaria vesca subsp. vesca; Gh, Gossypium hirsutum; Gm, Glycine max; Ha, Helianthus annuus; Lu, Linum usitatissimum; Mt, Medicago truncatula; Os, Oryza sativa; Rc, Ricinus communis; Si, Sesamum indicum; Sl, Solanum lycopersicum; Vv, Vitis vinifera; Zm, Zea mays*. Phytozome/Genbank accession number for each sequence is shown in brackets. Phylogenetic relationship among protein sequences of LACS was constructed using the neighbour-joining method. Bootstrap values are shown at the tree nodes. The amino acid substitution sites of AtLACS9 variants are marked with red-filled triangle. C, Homology model of AtLACS9 using PHYRE2 software and a carboxylic acid reductase as a template. The putative binding sites for ATP and fatty acid are shown in blue and green, respectively. The identified beneficial mutation sites C207 and D238 are shown in red. About 83% of the sequence was

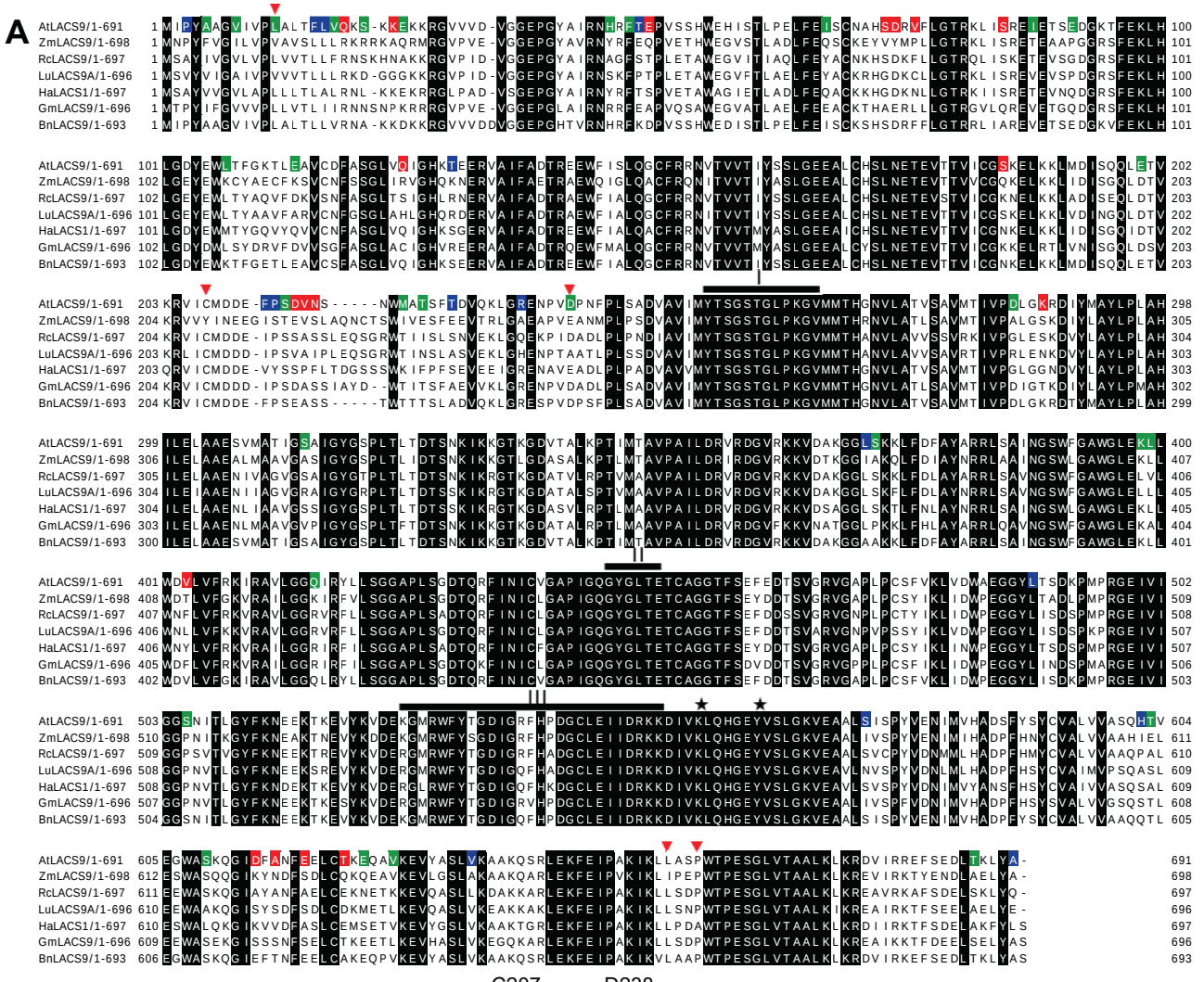
modelled with high confidence. The first 58 N-terminal residues were not included in the model as the template is a soluble enzyme. Based on TMHMM and SOSUI analyses, this N-terminal segment is predicted to constitute one N-terminal transmembrane domain, which was added to the structure.

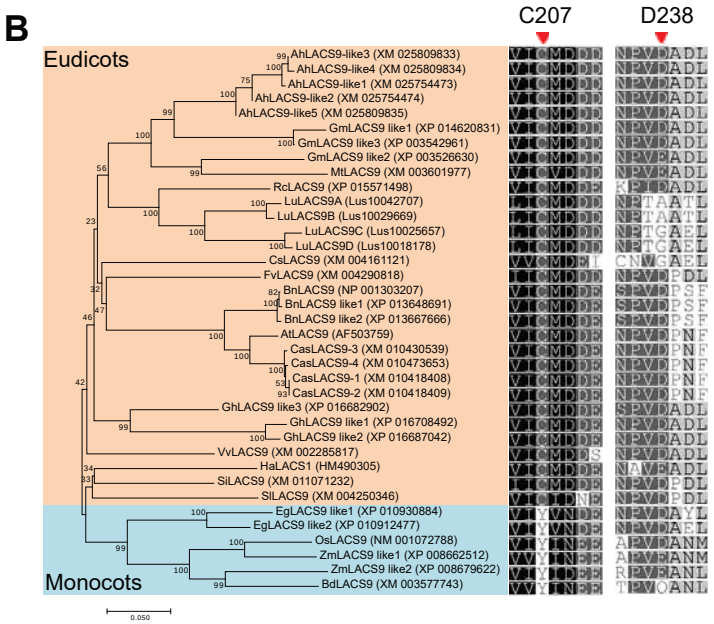

Neutral lipid content (ΔF TAG/ OD600)
300
250
200
150
100
50
0
0
12
24
36
48
60
72
84
96
108
Time (h)
L12F/C207F/L656F
AtLACS9
D238E/P659S
LacZ

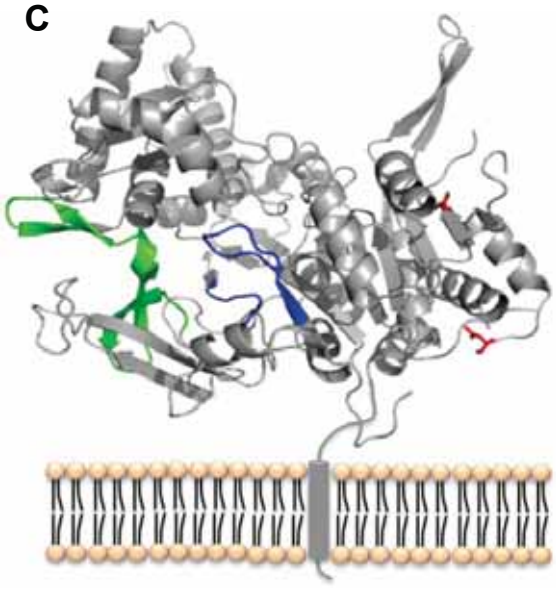

A
Enzyme activity
(nmol acyl-CoA/ min/ mg protein)
OD600
Culture time (h)
B
Relative enzyme activity
C
Relative protein accumulation
ND
D
Normalized enzyme activity
ND
L12F/C207F/L656F
D238E/P659S
AtLACS9
Lac Z


A
Relative enzyme activity
4.0
3.0
2.0
1.0
0.0
B
Relative protein accumulation
3.0
2.5
2.0
1.5
1.0
0.5
0.0
ND
C
Normalized enzyme activity
4.0
3.0
2.0
1.0
0.0
ND
L12F/C207F/L656F
L12F
C207F
L656F
D238E/P659S
D238E
P659S
AtLACS9
LacZ


160
140
120
100
80
60
40
20
0
Relative activity (%)
C207F
D238E
AtLACS9
16:0
18:0
18:1
18:2


A
AtLACS9/1-691
ZmLACS9/1-698
RcLACS9/1-697
LuLACS9A/1-696
HaLACS1/1-697
GmLACS9/1-696
BnLACS9/1-693


B
C207
D238
Eudicots
Monocots
AhLACS9-like3 (XM 025809833)
AhLACS9-like4 (XM 025809834)
AhLACS9-like1 (XM 025754473)
AhLACS9-like2 (XM 025754474)
AhLACS9-like5 (XM 025809835)
GmLACS9 like1 (XP 014620831)
GmLACS9 like3 (XP 003542961)
GmLACS9 like2 (XP 003526630)
MtLACS9 (XM 003601977)
RcLACS9 (XP 015571498)
LuLACS9A (Lus10042707)
LuLACS9B (Lus10029669)
LuLACS9C (Lus10025657)
LuLACS9D (Lus10018178)
CsLACS9 (XM 004161121)
FvLACS9 (XM 004290818)
BnLACS9 (NP 001303207)
BnLACS9 like1 (XP 013648691)
BnLACS9 like2 (XP 013667666)
AtLACS9 (AF503759)
CasLACS9-3 (XM 010430539)
CasLACS9-4 (XM 010473653)
CasLACS9-1 (XM 010418408)
CasLACS9-2 (XM 010418409)
GhLACS9 like3 (XP 016682902)
GhLACS9 like1 (XP 016708492)
GhLACS9 like2 (XP 016687042)
VvLACS9 (XM 002285817)
HaLACS1 (HM490305)
SiLACS9 (XM 011071232)
SiLACS9 (XM 004250346)
EgLACS9 like1 (XP 010930884)
EgLACS9 like2 (XP 010912477)
OsLACS9 (NM 001072788)
ZmLACS9 like1 (XP 008662512)
ZmLACS9 like2 (XP 008679622)
BdLACS9 (XM 003577743)
0.050


C