BMC Biotechnology

RESEARCH Open Access

Genome-wide identification and diversity of *FAD2*, *FAD3* and *FAE1* genes in terms of biotechnological importance in *Camelina* species

Rostyslav Y. Blume[1*], Vitaliy Y. Hotsuliak[1], Tara J. Nazarenus[2], Edgar B. Cahoon[2] and Yaroslav B. Blume[1]

Abstract

Background False flax, or gold-of-pleasure (*Camelina sativa*) is an oilseed that has received renewed research interest as a promising vegetable oil feedstock for liquid biofuel production and other non-food uses. This species has also emerged as a model for oilseed biotechnology research that aims to enhance seed oil content and fatty acid quality. To date, a number of genetic engineering and gene editing studies on *C. sativa* have been reported. Among the most common targets for this research are genes, encoding fatty acid desaturases, elongases, and diacylglycerol acyltransferases. However, the majority of these genes in *C. sativa* are present in multiple copies due to the allohexaploid nature of the species. Therefore, genetic manipulations require a comprehensive understanding of the diversity of such gene targets.

Results Here we report the detailed analysis of *FAD2*, *FAD3* and *FAE1* gene diversity in five *Camelina* species, including hexaploid *C. sativa* and four diploids, namely *C. neglecta*, *C. laxa*, *C. hispida* var. *hispida* and var. *grandiflora*. It was established that *FAD2*, *FAD3* and *FAE1* homeologs in *C. sativa* retain very high conservancy, despite their allohexaploid inheritance. High sequence conservancy of the identified genes along with their different expression patterns in *C. sativa* suggest that subfunctionalization of these homeologs is mainly grounded on the transcriptional balancing between subgenomes. Finally, fatty acid composition of seed lipids in different *Camelina* species was characterized, suggesting potential variability in the activity of fatty acid elongation/desaturation pathways may vary among these taxa.

Conclusion It was shown that the *FAD2*, *FAD3* and *FAE1* genes retain high conservation, even in allohexaploid *C. sativa* after polyploidzation, in which the subfunctionalization of the described homeologs is mainly grounded on the expressional differences. The major differences in FA accumulation patterns within the seeds of different species were identified as well. These results provide a foundation for future precise gene editing, which would be based on targeting of particular *FAD2*, *FAD3* and *FAE1* gene copies in *C. sativa* that allow regulating the dosage of the mentioned genes, thus shaping the desired FA composition in cultivated false flax.

*Correspondence:
Rostyslav Y. Blume
blume.rostislav@gmail.com

Full list of author information is available at the end of the article

Keywords *Camelina*, Crop wild relatives, Fatty acid, Desaturase, Fatty acid elongase, Oilseed

Introduction

False flax, or gold-of-pleasure (*Camelina sativa*), is an emerging oilseed crop, which has gained a renewed interest as a platform for genetic engineering and gene editing, aimed at altering seed fatty acid (FA) composition [1, 2]. This crop is currently viewed as one of the most promising candidates for production of oil-based liquid biofuels and, particularly, sustainable aviation fuel (SAF) [3, 4]. Lower content of very long-chain fatty acids (VLCFA), compared to other Brassicaceae, relatively high abiotic and biotic stress resilience and short vegetation cycle have contributed to the high research interest for this crop [5, 6]. Moreover, close genetic relation of *C. sativa* to another widely-used plant model, *Arabidopsis thaliana*, and high transformation amenability of false flax [3, 7, 8] make this crop an ideal candidate for oilseed biotechnology research.

Despite the potential of *C. sativa* as an oilseed crop, self-pollination nature of this crop and abandonment of its cultivation in the middle of the 20th century led to the decrease of genetic heterogeneity of this crop and loss of varietal diversity [9–12]. In addition, its allohexaploid nature has contributed to genetic paucity, as *C. sativa* had faced at least three major genetic diversity bottlenecks during the evolution [13]. Therefore, the wild relatives of this crop are considered potential germplasm donors for enhancing the genetic diversity of *C. sativa* [13–15]. Currently, the main progenitors of the cultivated false flax are well known. Among them *Camelina neglecta*, which is believed to be an ancestor for at least two subgenomes of *C. sativa* (N^6 –subgenome 1 and N^7 – subgenome 2) that were inherited from intermediate tetraploid species, *Camelina intermedia* nom. provis. (N^6N^7 genomes) [16]. *Camelina hispida* is considered to be the second diploid ancestor, which after the hybridization with *C. intermedia*, contributed the third subgenome (H^7) to *C. sativa-C. microcarpa* ancestral lineage [13, 15, 17].

The direct wild hexaploid progenitor of the cultivated false flax, *C. microcarpa*, is often viewed as a good candidate for interspecific hybridization; however it also suffers from the same genetic limitations [18–20]. Moreover, its use is limited by the presence of two distinct cytotypes with different chromosome counts and genome organization (Type 1, $2n=40$, $N^6N^7H^7$; and Type 2, $2n=38$, $N^6N^7N^6$) [15, 17, 18]. Other, more distant diploid relatives, like *Camelina laxa* and mentioned *C. hispida* and *C. neglecta*, are almost not amenable for crossing with *C. sativa*, as well as tetraploid *C. rumelica*, which is partially cross-compatible [13]. Therefore, transgenic methods and gene editing of *C. sativa* might be considered the most promising approaches for the improvement of this crop.

The fatty acid composition of *C. sativa* seeds is distinguished by high proportions of polyunsaturated FAs, including α-linolenic acid (ALA, 18:3). The high content of omega-3 ALA in false flax seeds made this oil of high interest for nutritional and industrial (e.g., drying oil) applications [5]. In addition, the amenability of *C. sativa* to transgenesis has also the metabolic engineering of seed oils with a wide range of fatty acid compositions. For example, significant progress was achieved towards the production of omega-3 long-chain polyunsaturated FAs, particularly docosahexaenoic (C22:6ω3) acid in *C. sativa* seeds [21–23], which has demonstrated the value for aquaculture feed [24, 25]. Conversely, *C. sativa* was genetically modified to produce seed oils with high content of medium-chain length fatty acids (MCFA), including C10-C14 fatty acids by introduction of specific lysophosphatic acid acyltransferase and FatB thioesterase genes from *Cuphea* species [26–29]. Increased content of MCFA in the oil appears to be beneficial for biojet fuel production, as such lipids could be more easily converted into SAF [28]. In addition, *C. sativa* was used as platform for accumulation of the industrially important terpenes or other compounds [30].

C. sativa has also been used as a model species for gene editing, aimed on regulating gene dosage effect [6]. For example, *C. sativa* plants were edited to decrease content of glucosinolates in seeds [31] or to alter seed storage protein accumulation [32]. However, manipulating seed lipid accumulation and their FA composition are the most popular aims of *C. sativa* gene editing. It has been shown that knockout of multiple *PDAT1* and *DGAT1* copies in cultivated false flax lead to a significant decrease in seed lipid accumulation, consistently with the number of mutated homeologs [33]. Similarly, knockout of multiple *FAE1* [34] or *FAD2* [35–37] homeologs results in reductions of specific fatty acids in a gene dosage-dependent manner.

However, the majority of *C. sativa* genes are present in multiple copies due to the allohexaploid nature of the species [38], which complicates precise gene editing [6] and requires a comprehensive understanding of the diversity of gene targets. Fatty acid desaturase (*FAD2*, *FAD3*) and elongase (*FAE1*) genes which have been primary targets for gene editing, were partially characterized for *C. sativa* [39]. These analyses, however, were conducted before the whole genome sequence was reported. Since the genome of cultivated false flax was sequenced [38], the understanding of this species evolution has been greatly expanded [13, 16] as well as the role

of different subgenomes in transcriptional balance in this polyploid [17]. Moreover, availability of whole-genome sequences of the diploid wild relatives and progenitor species [40–42] now allows tracking evolution of *FAD2*, *FAD3*, *FAE1* and origin of their diversity in the allohexaploid *C. sativa*.

In the present study, we aimed to identify and characterize *FAD2*, *FAD3*, *FAE1* in five *Camelina* species (*C. sativa*, *C. neglecta*, *C. laxa*, *C. hispida* var. *hispida* and var. *grandiflora*), for which complete genome assemblies are available to the date. A comprehensive characterization of these genes was aimed to reveal the conservancy of these genes, expressional differences in *C. sativa* and possible influence of such differences on observed fatty acid composition of seed lipids in different *Camelina* species, in order to simplify use of these genes as targets for genetic manipulation and would shed light on the role of different homeologs in fatty acid biosynthesis in *C. sativa*.

Materials and methods

Gene identification and analysis of their genomic organization

The initial identification for *FAD2*, *FAD3* and *FAE1* sequences in the genomes of *Camelina* species was conducted via series of BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) searches, using coding sequences of the *A. thaliana* genes as templates (*AtFAD2* – AT3G12120; *AtFAD3* – AT2G29980; *AtFAE1* – AT4G34520). We have analyzed the search results and discarded short and non-meaningful hits. The genome-wide search involved the annotated reference assembly of *C. sativa* (cv. DH55) genome (GCA_000633955.1), deposited in NCBI database [38]. Information on gene location, full genomic, coding and protein sequences was acquired from the NCBI database, as well.

Additionally, four unannotated genome assemblies of diploid *Camelina* species were included in the study, in particular: *C. neglecta* (GCA_023864065.1), *C. laxa* (GCA_024034495.1), *C. hispida* var. *hispida* (GCA_023657505.1) and *C. hispida* var. *grandiflora* (GCA_023864115.1) [40]. In this case, after the BLAST search the genomic region, confirmed to contain coding sequence of *FAD2*, *FAD3* or *FAE1* gene was extracted and further annotated, using WebScipio (https://www.webscipio.org/search) software [43]. This allowed identification of genes exon-intron structure, extraction of the coding sequences and putative peptide sequences.

A multiple sequence alignment of the *FAD2*, *FAD3* and *FAE1* genes CDS was performed with MUSCLE algorithm [44]. Exon-intron structure of *FAD2*, *FAD3* and *FAE1* genes was visualized using the Gene Structure Display Server (http://gsds.cbi.pku.edu.cn/) [45].

The genomic organization of upstream promoter regions of the identified genes was inferred as well. To do that 2 kbp upstream regions of the respective genes were downloaded from NCBI and analyzed against PlantCARE database (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/) [46], which allowed detection of *cis*-acting regulatory elements in the target sequences. Data on abundance of *cis*-elements of in the upstream regions of the identified genes was visualized using TBtools v2.045 [47]. Prediction of upstream transcription factors (TFs) in *C. sativa FAD2*, *FAD3* and *FAE1* homeologs was performed against JASPAR-2024 database (https://jaspar.elixir.no/) [48], using non-redundant Viridiplantae PFMs database. PWM-searches for TFs binding sites (TFBSs) in gene upstream promoter regions were conducted using MOODS package (https://github.com/jhkorhonen/MOODS) [49]. Data on the TFBSs were visualized using ggplot2 package in R.

Draft information on the orthology of the identified genes was retrieved from KEGG database (https://www.genome.jp), while locus ID (in CsaXXgXXXXX format) was determined via EnsemblPlants database (http://plants.ensembl.org) search using genomic coordinates of particular gene as a query. Further, these loci IDs were used to verify presence/absence of tubulin genes on homologous chromosomes. The values of *Ka/Ks* ratio were also calculated in TBtools v2.045 [47].

Gene expression analyses

Transcriptomic data for *C. sativa* (cv. DH55) were obtained from a publicly available database (https://bar.utoronto.ca/eplant_camelina/) [50]. The expression levels of the identified genes in twelve different tissues (at different developmental stages) were taken for the analysis. The expression was analyzed in the following tissues: germinating seed (GS), cotyledon (C), young leaf (YL), senescing leaf (SL), stem (S), root (R), flower bud (B), flower (F) and seeds/fruits at various developmental stages – early (ESD), early-mid (EMSD), late-mid (LMSD) and late (LSD). Expression heatmaps were constructed using Heatmap tool in TBtools v2.045 software [47].

Protein sequence and structure conservancy analysis

Domain organization of the identified peptides was analyzed using InterPro (https://www.ebi.ac.uk/interpro) tool [51]. Peptide sequences were searched against the databases of functional domains (Pfam) [52] and CDD [53]. Identified peptide motifs and domains were visualized using TBtools v2.045 software [47].

Localization of transmembrane domains of FAD2, FAD3 and FAE1 was conducted using DeepTMHMM algorithm (https://dtu.biolib.com/DeepTMHMM) [54], a more recent and precise version of TMHMM 2.0 [55],

since more common InterPro tool [51] failed to identify transmembrane domains correctly, as it relies on the older version of TMHMM. Allocation of transmembrane domains within the analyzed peptides was visualized using TBtools v2.045 software [47]. Data on amino acid sites conservancy was retrieved from ClustalX 2.1 [56].

3D structures of the identified proteins were inferred using ColabFold tool [57]. Pairwise calculation of RMSD values of the constructed models and model alignment was performed using BioPython package [58] and visualized using TBtools v2.045 software [47]. 3D model were visualized using RCSB Protein Data Bank Mol*3D Viewer (https://www.rcsb.org/3d-view) [59].

Phylogenetic and synteny analyses

For the phylogenetic analysis, the peptide sequences of the identified *FAD2*, *FAD3* and *FAE1* genes were used, as well as the previously reported sequences of *FAD2* and *FAE1* of different *Camelina* species [39], including *C. rumelica* and *C. microcarpa*, for which no whole-genome sequencing were reported to date. Such sequences are given in Table 1.

The amino acid sequences of FAD2, FAD3 and FAE1 were aligned using MUSCLE algorithm [44]. Optimal substitution model was identified using ModelFinder [60] for Maximum Likelihood tree reconstruction. For the set of FAD2 proteins JTTDCMut+I was determined as the optimal model, for FAD3 – cpREV+I, and for FAE1 – JTTDCMut+G4. Phylogenetic analysis (ML) was performed using web version of IQ-TREE tool (http://iqtree.cibiv.univie.ac.at/) [61, 62] with the boostrap support of 1000 iterations, involving the usage of UFBoot for ultrafast bootstrapping [63]. The resulting trees were visualized using the web-version of iTOL v6 tool (https://itol.embl.de) [64].

In order to prepare unannotated *Camelina* genomes for the further comparative genomics analyses, AUGUSTUS 3.3.2 [65] was run on the *C. neglecta*, *C. laxa*, *C. hispida* var. *hispida* and *C. hispida* var. *grandiflora* genome assemblies available in NCBI database. Noteworthy, the genome sequence of *C. neglecta* was obtained using the same specimen, which was used for the species description [40, 66]. This allowed prediction of genes in silico, using *A. thaliana* as a template for annotation. Syntenic relations between the identified *FAD2*, *FAD3* and *FAE1* in the five *Camelina* genomes were analyzed based on the coding sequencing data in TBtools v2.045 software [47], using MCScanX algorithm [67]. The results were further visualized as a circos plot in the same software.

Table 1 Previously reported sequences of *FAD2* and *FAE1*, used in the phylogeny reconstruction

Protein type	Gene name	Genebank nucleotide ID	Genebank protein ID
FAD2	*CsFAD2-A*	GU929417.1	ADN10824.1
	CsFAD2-B	GU929418.1	ADN10825.1
	CsFAD2-C	GU929419.1	ADN10826.1
	ChFAD2	GU929426.1	ADN10829.1
	ClFAD2	GU929429.1	ADN10830.1
	CmFAD2-A	GU929432.1	ADN10831.1
	CmFAD2-B	GU929433.1	ADN10832.1
	CmFAD2-C	GU929434.1	ADN10833.1
	CrFAD2-1	GU929438.1	ADN10834.1
	CrFAD2-2	GU929439.1	ADN10835.1
FAE1	*CsFAE1-A*	GU929420.1	ADN10812.1
	CsFAE1-B	GU929421.1	ADN10813.1
	CsFAE1-C	GU929422.1	ADN10814.1
	ChFAE1-1	GU929427.1	ADN10816.1
	ChFAE1-2	GU929428.1	ADN10817.1
	ClFAE1-1	GU929430.1	ADN10818.1
	ClFAE1-2	GU929431.1	ADN10819.1
	CmFAE1-A	GU929435.1	ADN10820.1
	CmFAE1-B	GU929436.1	ADN10821.1
	CmFAE1-C	GU929437.1	ADN10822.1
	CrFAE1-1	GU929440.1	ADN10823.1
	CrFAE1-2	GU929441.1	ADN10836.1

Determination of seed fatty acid composition

The plant material of *Camelina* sp. genotypes, used in the present study was obtained from USDA National Plant Germplasm System (USDA-NPGS), while *C. sativa* accessions were received from the collection M.M. Gryshko National Botanical Garden of Natl. Academy of Sciences of Ukraine (Kyiv, Ukraine). The list of the accessions is provided in Table 2.

Table 2 List of the *Camelina* sp. accessions, used in the study

Species	Accession No.	Name	The country of origin
C. sativa	-	cv. Mirazh	Ukraine
	-	cv. Klondaik	Ukraine
C. alyssum	PI650132	CA-CAM21	Germany
C. microcarpa	Ames 31219	GE.2011-02	Georgia
	PI633187	Index Seminum 2468	Poland
	PI633191	NU 60689	USA, Montana
	PI633186	No. 61	Hungary
	PI650136	CM-CAM6	Germany
C. rumelica	PI650134*	160-0933-66	Spain
	PI650138	161-3724-75	Iran
C. neglecta	PI650135**	Index Seminum 238	France
C. laxa	Ames 32852	AM-2014-12	Armenia
C. hispida var. *grandiflora*	PI650133***	158-6281-83	Turkey

*The accession is misidentified in US NPGS as *C. microcarpa*; **This accession was used for the species description [66] and for the genome sequencing [40]; ***This accession was used for the genome sequencing [40]

Fatty acid methyl esters (FAMEs) were prepared via transesterification, using trimethylsulphonium hydroxide (TMSH). Single seeds were directly crushed in 50 µL of TMSH in glass GC vials. Heptane (400 µL) was added to each vial. After the incubation at room temperature with agitation for 30 min, FAMEs were analyzed by gas chromatography as described previously [68].

Specific coefficients, describing fatty acid biosynthesis were used, in order to characterize overall differences among the fatty acid profiles of the investigated *Camelina* sp. accessions. ER (elongation ratio) and DR (desaturation ratio) are relative values, showing the relative share of oleic acid (18:1) elongation or desaturation pathways, respectively [69]. The calculation is based on the content of particular fatty acids (given in mol%), which appear to be a result of oleic acid conversion (e.g. linoleic (18:2) and linolenic (18:3) acids in the case of desaturation pathway), divided by the total content of oleic acid and its desaturation/elongation products, observed within the analyzed fatty acid profile of seed lipids.

ODR (oleic desaturation ratio) and LDR (linoleic desaturation ratio) coefficients are aimed on evaluation of the activity of individual desaturation enzymes [70], in this case the activity of FAD2 and FAD3. Similarly, these ratios evaluate the efficiency of oleic acid desaturation (ODR) or the desaturation of linoleic acid to linolenic (LDR). In order to evaluate the efficiency oleic (18:1) and gondoic (20:1) acids elongation, we proposed two additional equations. GER (gondoic acid elongation ratio) describes the efficiency of oleic acid conversion into its primary elongated product – gondoic acid and its further conversion to erucic acid. EER (erucic elongation ratio) indicates the efficiency of erucic acid biosynthesis out of gondoic acid. Despite both stages are catalyzed by FAE1, the relative share of this elongation stages might significantly differ. The equations for all used coefficients are provided below:

$$ER = \frac{\%C20:1 + \%C22:1}{C18:1 + \%C18:2 + \%C18:3 + \%C20:1 + \%C22:1}$$

$$DR = \frac{\%C18:2 + \%C18:3\%}{C18:1 + \%C18:2 + \%C18:3\% + C20:1 + \%C22:1}$$

$$ODR = \frac{\%C18:2 + \%C18:3\%}{C18:1 + \%C18:2 + \%C18:3\%}$$

$$LDR = \frac{\%C18:3\%}{\%C18:2 + \%C18:3\%}$$

$$GER = \frac{\%C20:1 + \%C22:1}{C18:1 + \%C20:1 + \%C22:1}$$

$$EER = \frac{\%C22:1}{\%C20:1 + \%C22:1}$$

Statistical processing of data

All statistical processing of the obtained data was conducted using OriginPro 2019b software. Deviations of all means were calculated as a standard deviation (SD). To identify the significance of differences in different parameters between the studied genotypes, one-way ANOVA was used, which included the calculation of Fisher's least significant differences (LSDs). PCA- and dot-plots were also constructed using OriginPro 2019b software. The LSDs were used to identify homogeneous groups for values of particular parameter at different level of significance $p<0.05$, $p<0.01$, $p<0.001$.

Results

Diversity of *FAD2* genes in *Camelina* species

We initially identified *FAD2* genes within the genomes of five *Camelina* species, resulting in the identification of seven *FAD2* genes (Table 3). Allohexaploid *C. sativa* contained three genes (one per each subgenome), while the diploid species had only one *FAD2* each. The genes *CsFAD2-A*, *CsFAD2-B*, *CsFAD2-C* are located in homologous chromosomes Cs19, Cs01 and Cs15, respectively. All three genes allocated in the same ancestral genomic block F, if previously published *Camelina* genome evolution models were taken into account [17, 38]. This suggests that the triplet of *CsFAD2* genes arose in result of *C. sativa* allopolyploidy, indicating that these three genes are likely homeologs.

Among a broader panel of *Camelina* species, which includes *C. microcarpa* and *C. rumelica* (both of which have polyploid genomes), *FAD2* genes show very high conservancy rate. Notably ≥95.8% of encoded amino acid residues tend to be invariable among the species of the genus. Such high conservancy rate allows tracking of the origin of *CsFAD2* genes from different subgenomes. A reconstruction of FAD2 phylogeny shows clear differentiation of the proteins, corresponding to the subgenome, to which a particular *FAD2* gene belongs (Fig. 1a). For instance, *FAD2* contained in *C. neglecta*-type subgenomes (N^6 and N^7, A and B subgenomes respectively) formed a separate group that includes sister clades of *CsFAD2-A—CmFAD2-A* and *CsFAD2-B—CmFAD2-B*. This major group of N^{6-7} subgenomes included also *CnFAD2* in a basal branch of N^7 clade, since this gene

Table 3 Identified *FAD2* genes in five *Camelina* species

Name	Gene ID	Location	Strand
CsFAD2-A	104764975	19:5522581-5526159	-
CsFAD2-B	104776214	1:4948902-4952339	-
CsFAD2-C	104745425	15:5229286-5232157	-
CnFAD2	-	3:4930838-4931992	-
ClFAD2	-	1:4527318-4528472	-
ChvhFAD2	-	3:6167354-6168508	-
ChvgFAD2	-	3:6167354-6168508	-

might be evolutionary close to the ancestral form of FAD2 for N^{6-7} group.

Synteny analysis of the identified *FAD2* genes confirmed their potential orthologous nature in *Camelina* species (Fig. 1b). The analysis of *C. sativa* genome compared to four diploid *Camelina* genomes showed that each of *CsFAD2* genes form four syntelogous pairs with *FAD2* genes in the other species. Synteny analysis itself was unable to clarify the origin of *FAD2* homeologs from the diploid species, unlike the phylogeny reconstruction. This suggests that the *FAD2* genes have not undergone duplications (except allopolyploidy-mediated WGD) during the evolution of the *Camelina* genus, which is highly consistent with their conserved nature. Moreover, exon-intron structure of these genes tend to be also invariable (Fig. 1c). Almost all identified *FAD2* genes consisted contained a single exon. The exception is *CsFAD2-A*, in which the coding region was split into two exons by an 890 b.p. long intron in the GCA_000633955.1 assembly. However, the search in a more recent genome assembly GCA_030686135.1 suggest that the coding region of *CsFAD2-A* is not split into two exons. Therefore, we strongly believe that two exon structure of *CsFAD2-A* in the earlier GCA_000633955.1 assembly results from a genome assembly artefact. Figure 1c shows *CsFAD2-A* as a single-exon gene. Calculated *Ka/Ks* ratios for *CsFAD2-A/B/C* indicate that these genes have not faced any significant selective pressure after polyploidization, since *Ka/Ks* values are very low – about 0.019–0.025 (Table S2).

High sequence conservancy of all analyzed *FAD2* genes, and *CsFAD2* homeologs in particular, suggest that transcriptional balancing between homeologs may play a significant role in subfunctionalization. Understanding the expression differences between homeologs is key for elucidating the most important gene copy, which has greater impact on trait formation, so such a loci may become a subject of further breeding or a target for gene editing. Therefore, the observed expression rates of different *CsFAD2* homeologs (Fig. 1d) suggest that the expression of all three genes reaches its peak during seed development, especially at early-mid stage. However, the genes are not expressed equally. Thus, the highest expression level was noted for *CsFAD2-C*, which was at least 10-fold higher than the expression of *CsFAD2-A* and 1.45-fold higher than *CsFAD2-B*. Moreover, the domination of *CsFAD2-C* transcripts was observed on other seed development stages, as well as during other stages, like floral development, cotyledon stage and young leaf growth, etc. *CsFAD2-B* was the second most expressed *FAD2* homeolog, however, its expression was usually 1.5-3.4-fold lower, than the expression of *CsFAD2-C*. In the majority of tissues the expression of *CsFAD2-A* was 6-10.5-fold lower, than such of *CsFAD2-C* (except the senescing leaf tissues and during the late seed development – only 2.8 to 3.4-fold lower).

Analysis of upstream promoter region of the identified genes has revealed that *C. sativa* homeologs demonstrated different composition of *cis*-acting elements in such regions (Fig. S1a). TATA-box and CAAT-box elements appeared to be most abundant motifs, detected within the upstream regions of *FAD2* genes. Despite the number of TATA-box elements is usually considered to be associated with general levels of expression, the highest expressed homeolog *CsFAD2-C* had only 29 TATA-box elements in upstream region, while the other *Camelina FAD2* genes typically contained more than 30 of such motifs (except for the *CsFAD2-A*). CAAT-box elements were contatined in notably higher quantities in *CsFAD2-B* and *CsFAD2-C*, compared to *CsFAD2-A*. Interestingly, that *CsFAD2-A*, is inherited from *C. neglecta* genome, in which *CnFAD2* contains significantly lower number of CAAT-box in upstream region (26 vs. 37–39 in *C. laxa* and *C. hispida* species). This fact may suggest that organization of promoter region of these genes may be highly dependent on the promoter strength in parental species, rather than solely on post-polyploidy gene dosage balancing in *C. sativa*.

Prediction of possible TFBSs in *C. sativa FAD2* genes has shown that the majority of the detected sites are associated with DOF family TFs (commonly, DOF3.4, DOF3.6, DOF5.8, etc.) (Fig. S1b, c, d). Similarly, one of the most abundant sites was associated with DOF-cycling factor, namely CDF5 [71]. Interestingly, that higher-expressed homeologs, *CsFAD2-B* and *CsFAD2-C* numerous potential binding sites for BPC1 and BPC5 TFs (*CsFAD2-C* promoter region contained 1.8-fold more of such sites), which are generally associated with control of *Seedstick* (*STK*) gene [72], controlling ovule identity and flowering at different stages (Fig. S1c, d). Second the most abundant were the sites for C2H2 zinc finger factors (*RAMOSA1*), controlling meristemal activity, inflorescence architecture, flower and seed size [73, 74]. Noteworthy, the *CsFAD2-C*, which is highest-expressed *FAD2* homeolog during EMSD, contained the most of predicted sites for *RAMOSA1*-like TFs (up to 36) (Fig. S1d). It may be assumed that such upstream promoter region organization may condition the highest expressional activity of this gene during seed development (Fig. 1d).

Investigation of the protein domain distribution did not reveal any significant differences. Detection of the conversed functional domains suggested that the peptides of the identified *FAD2* genes tend to possess typical domains: Fatty acid desaturase domain (PF00487 – Pfam ID) or larger Delta 12-FADS-like domain (cd03507 – CDD ID) (Fig. S2). The domains were retained at conserved positions, whereas their location was shifted by 1 aa in only CsFAD2-C, due to a single aa insertion at

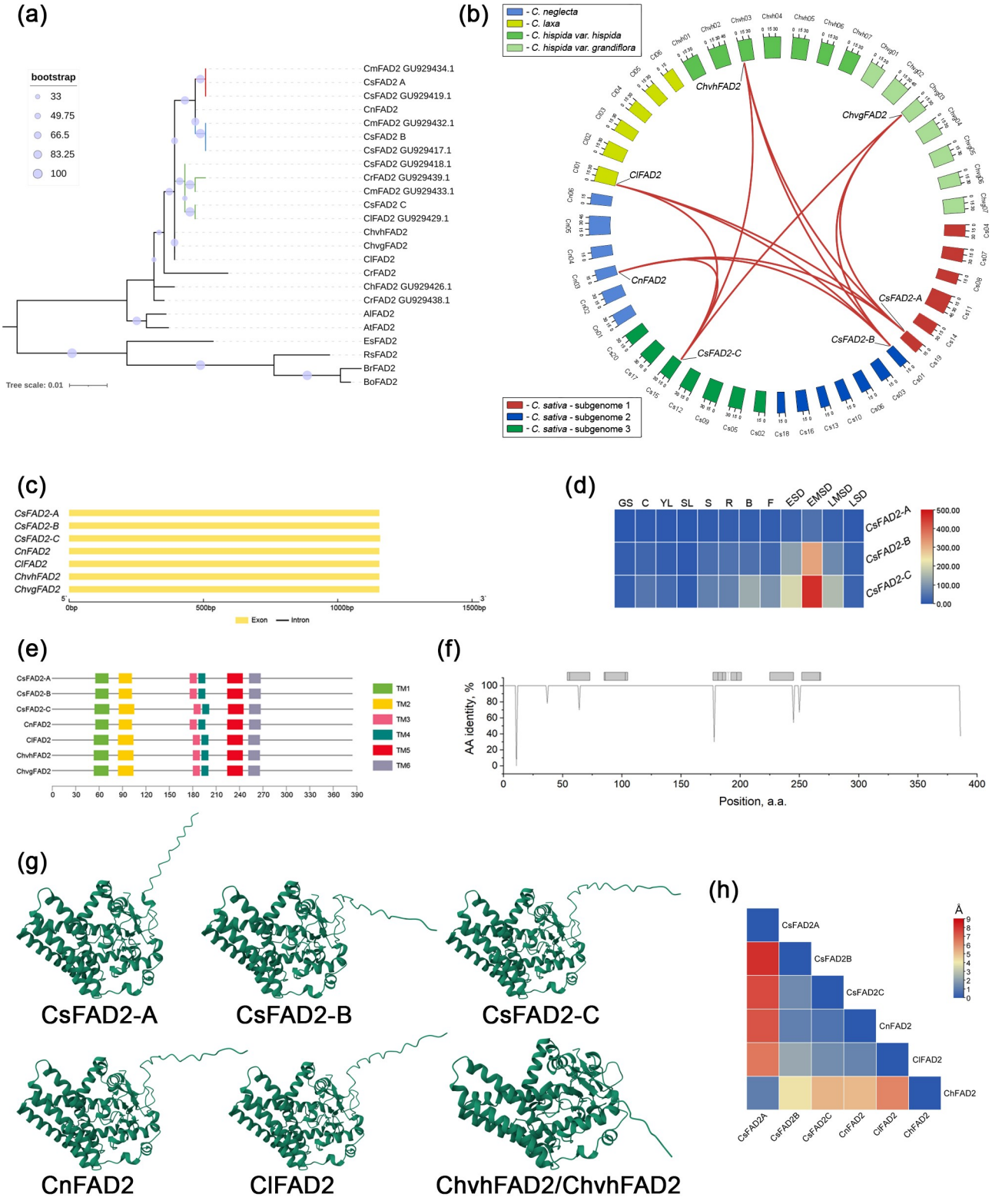

Fig. 1 Analysis of *FAD2* diversity within *Camelina* species: (**a**) – Maximum likelihood tree (*lnL*= -1292; rooted to AtFAD2) of FAD2 proteins of different *Camelina* species, constructed using both translation of the CDS of the identified genes and the previously reported sequences; (**b**) – synteny analysis of *FAD2* genes in allohexaploid *C. sativa* and its four diploid relatives; (**c**) – exon-intron structure of the identified *FAD2* genes; (**d**) – expression of *CsFAD2* from different subgenomes in various tissues; (**e**) – distribution of transmembrane domains within the putative protein product of the identified *FAD2* genes; (**f**) – sequence conservancy of *Camelina* FAD2 proteins with indication of the identified transmembrane domains; (**g**) – 3D structures of the identified FAD2 proteins in different *Camelina* species; (**h**) – RMSD values heatmap, showing degree of the structural differences between analysed FAD2 proteins

N-end. In particular, we found the number of transmembrane domains is constant and their location shows little variation (Fig. 1e). All FAD2 peptides of the analyzed *Camelina* species contained six transmembrane domains at mostly conserved positions. CsFAD2-C has all transmembrane domains shifted by one amino acid residue, since this protein has serine insertion at 11th position (Fig. 1f). Only six positions were variable (besides the insertion at position 11) within *Camelina* FAD2 proteins, three of which involved transmembrane region (pos. 64, 178, and 245). Substitutions at these positions may potentially affect the protein structure or function. Prediction and analysis of FAD2 proteins 3D structure suggested that *C. sativa* homeologs possess significant difference of FAD2 monomers with RMSD of 1.41–8.1Å (Fig. 1g, h). Noteworthy, CsFAD2-A protein demonstrated significant structural difference not only, compared to its homeologs, but also to its ortholog from parental species *C. neglecta* (7.29Å) (Fig. 1h). Taking in account that this homeolog is the least expressed one, it might be suggested that *CsFAD2-A* is undergoing gradual sequence divergence or disruption and might be pseudogenized or eliminated in the future. Despite the unequal expression of *FAD2* homeologs in *C. sativa*, their sequences and protein structure are highly conserved, suggesting that all three copies might be more or less important for the growth and development of this species.

Diversity of *FAD3* genes in *Camelina* species

Similarly to the analysis of *FAD2* genes we have performed the identification of *FAD3* genes in the genomes of the investigated *Camelina* species (Table 4). As it was shown for the described above desaturase, *FAD3* genes were also mostly represented by a single gene in each of the studied *Camelina* species, except allohexaploid *C. sativa*, in which three *FAD3* genes can be found. The genes *CsFAD3-A*, *CsFAD3-B*, *CsFAD3-C* are located in homologous regions (of ancestral block J) in chromosomes Cs07, Cs16 and Cs05, respectively. Absence of any additional copies in non-homologous regions or tandemly located suggests that all three genes were inherited from the parental species via the series of allopolyploidy events that *C. sativa* faced during evolution.

Table 4 Identified *FAD3* genes in five *Camelina* species

Name	Gene ID	Location	Strand
CsFAD3-A	104700502	7:6302189-6305695	+
CsFAD3-B	104749896	16:5592035-5595480	+
CsFAD3-C	104786676	5:12491769-12495209	+
CnFAD3	-	2:6094160-6097522	+
ClFAD3	-	4:26670423-26673736	-
ChvhFAD3	-	4:15654813-15658152	+
ChvgFAD3	-	4:15654813-15658152	+

The identified FAD3 showed also very high level of putative protein sequence conservation (~ 95.3%).

The phylogeny of the identified FAD3 was reconstructed as well (Fig. 2a). Unfortunately, no other studies reported about identification of complete *FAD3* gene sequences in the *Camelina* species other than investigated here. Only cloning of partial *FAD3* coding sequence has been reported [75], which was not included here. For FAD3, a clear grouping of orthologous proteins was observed. For instance, CsFAD3-C from H^7 subgenome and ChvhFAD3, ChvgFAD3 from parental species of H^7 subgenome were grouped in the common clade, which was not observed in the case of FAD2. Respectively, members of N^{6-7} genomic lineage, CsFAD3-A, CsFAD3-B and CnFAD3, were placed into the distinct clade.

Synteny analysis of the identified *FAD3* genes in *Camelina* species (Fig. 2b) showed that each of *CsFAD3* form four syntelogous pairs with *FAD3* genes of the other diploid species. Such non-selective syntelogs pairing may also be caused by the high conservancy rate of *FAD3* genes in different species. As it was in the case *FAD2*, the synteny analysis is more likely to show orthologous relations among the identified *FAD3* genes of *Camelina* sp. Moreover, the exon-intron structure of *FAD3* genes appears to be even more conservative: all genes consisted of a single exon (Fig. 2c). The genes are not variable in length (1164 b.p.), since they encode proteins of a similar length (387 a.a.). Relatively small amount of non-synonymous substitutions conditioned low values of *Ka/Ks* ratio, which was 0 for *CsFAD3-A* that has not changed, compared to its ancestral orhtolog, *CnFAD3*. For other homeologs, this value consisted 0.049–0.066. Similar to *CsFAD2*, *CsFAD3* genes were highly conserved (Table S2).

Since *CsFAD3* genes retain high sequence conservancy, their subfunctionalization might have been influenced by the divergence in expression patterns (Fig. 2d). In contrast to *FAD2* genes, the highest expression among *CsFAD3* was recorded for the genes from N^{6-7}-subgenomes, *CsFAD3-A* and *CsFAD3-B*. Both genes were almost equally expressed during early and early-mid seed development, exceeding *CsFAD3-C* by 1.3-1.4-fold. During the late-mid seed development *CsFAD3-A* demonstrated 1.3-fold higher expression, than its homeologs, while at the stage of late seed development all three genes were almost not expressed. At the other development stages, *CsFAD3-A* have not shown any significantly higher expression levels, compared to its homeologs.

Investigation of 2 kbp upstream promoter region of the identified genes has revealed that *Camelina FAD3* genes demonstrate different composition of *cis*-acting elements in such regions (Fig. S3a). TATA-box (including AT ~ TATA-box) and CAAT-box elements appeared to be most abundant motifs, similarly to the promoter regions

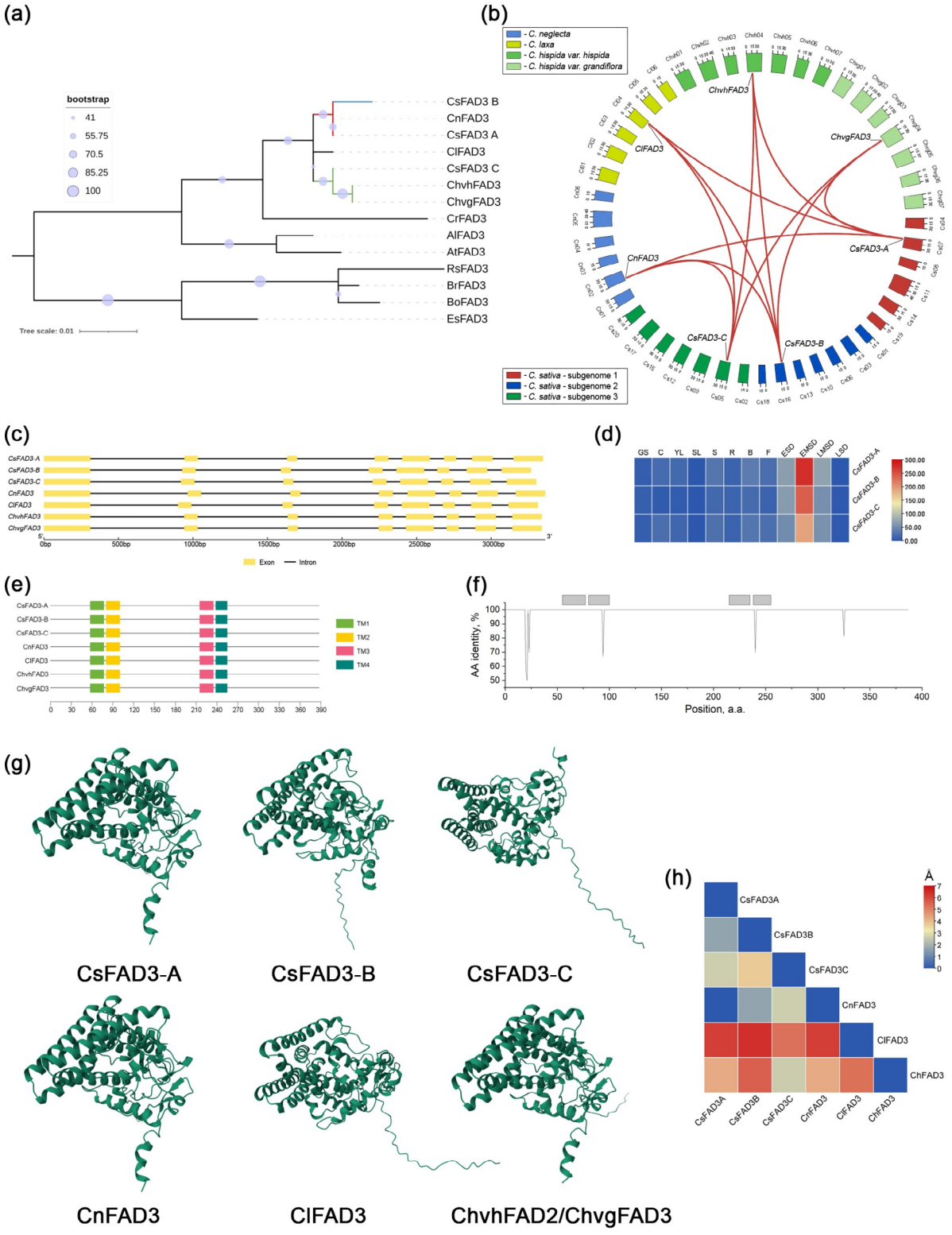

Fig. 2 The analysis of *FAD3* gene diversity within *Camelina* species: (**a**) – Maximum likelihood tree (*lnL*= -1275; rooted to AtFAD3) of FAD3 proteins of different *Camelina* species, constructed using both translation of the CDS of the identified genes; (**b**) – synteny analysis of *FAD3* genes in allohexaploid *C. sativa* and its four diploid relatives; (**c**) – exon-intron structure of the identified *FAD3* genes; (**d**) – expression of *CsFAD3* from different subgenomes in various tissues; (**e**) – distribution of transmembrane domains within the putative protein product of the identified *FAD3* genes; (**f**) – sequence conservancy of *Camelina* FAD3 proteins with indication of the identified transmembrane domains; (**g**) – 3D structures of the identified FAD3 proteins in different *Camelina* species; (**h**) – RMSD values heatmap, showing degree of the structural differences between analysed FAD3 proteins

of *FAD2* genes. However, the number of these elements tends to be higher in upstream regions of *FAD3* genes (up to 69 predicted TATA-box and up to 43 CAAT-box elements). The reason of such *cis*-activating elements distribution is unclear and has no direct correlation with gene expression in the case of *C. sativa FAD3* genes. Noteworthy, *CsFAD3-A* and *CsFAD3-B* had relatively higher number of MYC-associated motifs in the upstream regions (10 and 9 elements respectively). It is currently believed that distribution and activity of these sites might be associated with cold stress response and phytohormone signaling [76]. Similarly to the case of *FAD2* genes, the *cis*-element composition of *C. sativa FAD3* homeologs tends to be more similar to the wild species of the same genome type (e.g. promoters of *CsFAD3-A* and *CsFAD3-B* had similar *cis*-element number as *CnFAD3*, while *CsFAD3-C* was more alike *ChvgFAD3* or *ChvhFAD3*).

Prediction of possible TFBSs in *C. sativa FAD2* genes has shown presence of many predicted sites for DOF bindings (commonly, DOF3.4, DOF3.6, DOF5.8, etc.) (Fig. S3b, c, d). Noteworthy, the highest number of such sites were detect in the promoter region of the least expressed *CsFAD3-C* homeolog (Fig. 3d). Contrarily to other genes, the most expressed *CsFAD3-A* showed significant presence of DOF5.1 sites. Promoter regions of all three homeologs demonstrated presence of 6–10 predicted sites for BPC6 and 4–6 sites for CDF5 (Fig. S3b, c, d). *CsFAD3-A* and *CsFAD3-B* showed considerable amount of sites for AtHB-23, a homeodomain-leucine zipper I TF (Fig. S3b, c), which is believed to one of the crucial elements of the adaptation to increased salinity and root development [77]. Interestingly, only promoter regions of *CsFAD3-B* and *CsFAD3-C* contained significant number (6 and 9 sites respectively) of TFBSs for APETALA1 (AP1) (Fig. S3c, d). This TF is crucial for floral development [78]. Promoters of all three homeologs also contained several TFBSs for different types of AT-hook factors and homeodomain factors, which, however, may be related non-specific gene regulation.

Analysis of the conversed domains distribution revealed that all identified FAD3 proteins contain the same fatty acid desaturase domains, as FAD2 proteins: Fatty acid desaturase domain (PF00487) or Delta 12-FADS-like domain (cd0350) (Fig. S4). These domains were localized at the same positions in all FAD3 proteins. Moreover, all identified FAD3 polypeptides retained constant number of transmembrane domains with no differences in their positions (Fig. 2e). The amino acid sequences of the investigated proteins varied only at six distinct positions, two of which were in predicted transmembrane domains (TM): pos. 94 in TM2 and pos. 240 in TM4 (Fig. 2f). Other substitutions were located in N- and C-tails, exposed into cytosol (pos. 20, 21, 21, 325).

Comparison of FAD3 proteins 3D structure revealed slightly higher conservancy of these proteins within *Camelina* species (Fig. 2g, h). While FAD3 homeologs of *C. sativa* were slightly more diverse (compared to FAD2) – 1.68–3.76Å, these proteins retained conserved structure, if compared to those in parental species. For instance, CsFAD3-A was almost identical to CnFAD3 (RMSD -0.65×10^{-14}Å), as well as CsFAD3-B (1.68Å). CsFAD3-C have more different structure from its ortholog – ChFAD3 (2.58Å), but still could be considered enough structurally conserved. At the same time, ClFAD3 protein structure differed dramatically from other identified homologs (5.29–6.43Å), which may be conditioned by the relatively large evolutionary distance of *C. laxa* from other genus representatives. It is worth noting that overall differences in protein structure serve as rather evidence of evolutionary divergence of the proteins, rather than depict differences in functioning or activity of these enzymes. Considering the give above, it is highly unlikely that any of *C. sativa* FAD3 homeologs are currently undergoing pseudogenization or any kind of loss-of-function.

Diversity of *FAE1* genes in *Camelina* species

The results of the genome-wide search allowed identification of four *FAE1* genes in the genome of allohexaploid *C. sativa* and one gene per each of diploid species (*C. neglecta*, *C. laxa*, *C. hispida* var *grandiflora* and var. *hispida*) (Table 5). Two *FAE1* genes in the first subgenome of *C. sativa* were found to be tandemly located (GCA_000633955.1) on Chr11. However, we strongly believe that this could be the result of genome assembly artefact, since other flanking genes are also represented in two jointly located copies. If this is the result of a real tandem duplication, whole such genomic region should be duplicated and two identical loci with similar gene content might be located one after another.

In the case of GCA_000633955.1 assembly genes are duplicated individually (e.g. 3-ketoacyl-CoA synthase 17-like with two identical 104727572, 104721343 genes, etc.) and have 100% identical sequences and, thus, encode identical proteins. The RefSeq annotation also suggests that *CsFAE1-1-A* (104721341) and *CsFAE1-2-A* (104721342) genes produce indistinguishable transcripts (Csa11g007400). Finally, the search throughout a more recent genome assembly GCA_030686135.1 resulted in identification of only a single *CsFAE1-A* gene at the desired locus. Moreover, this is supported by the presence of only a single ortholog, *CnFAE1*, in the parental *C. neglecta* species in the homologous locus. Considering these facts in further analyses we treated *CsFAE1-1-A* and *CsFAE1-2-A* as a single gene, *CsFAE1-A*.

FAE1 genes of *Camelina* species have less conservation than we found for *FAD2* and *FAD3*. We established that

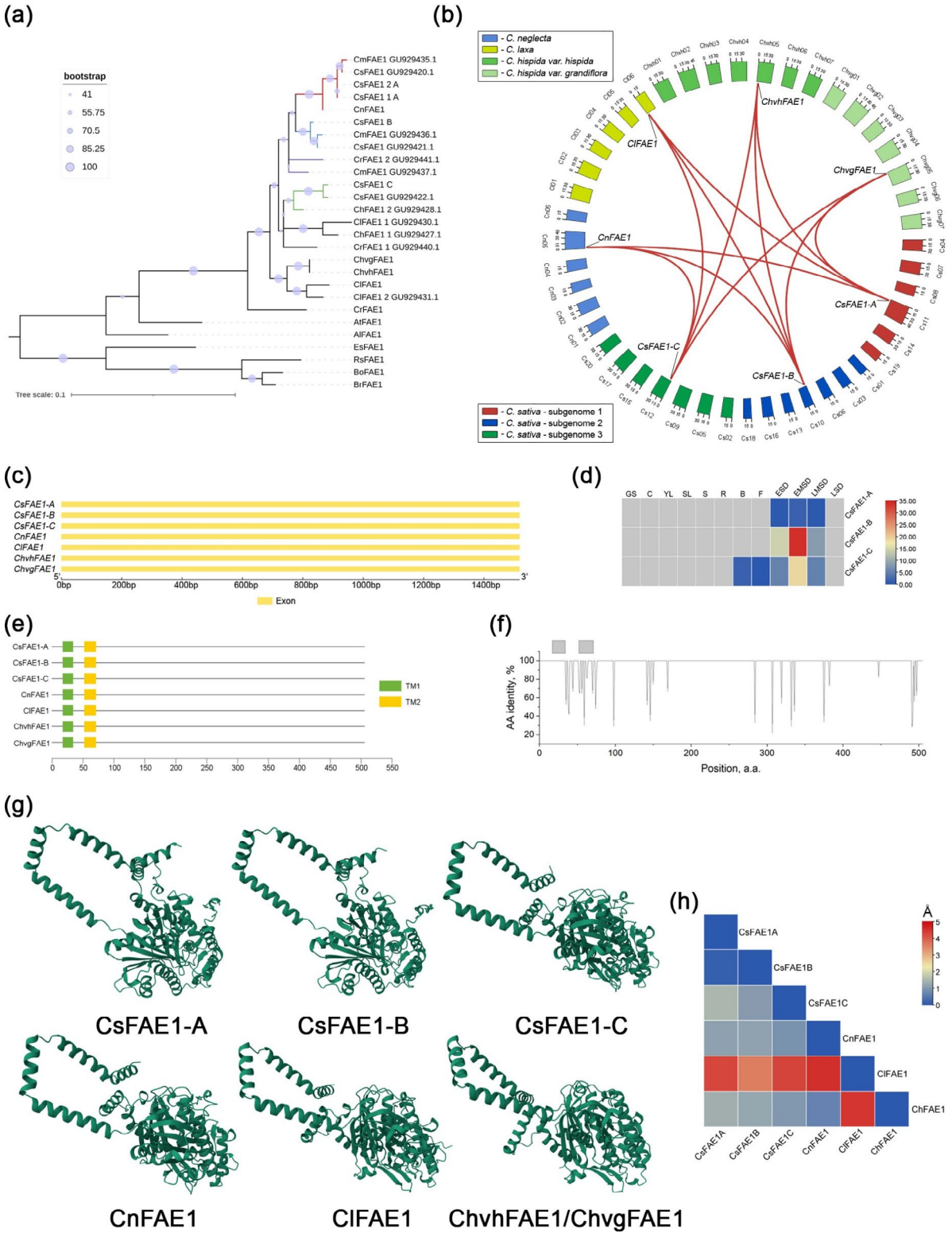

Fig. 3 The analysis of *FAE1* gene diversity within *Camelina* species: (**a**) – Maximum likelihood tree (*lnL*= -2312; rooted to AtFAE1) of FAE1 proteins of different *Camelina* species, constructed using both translation of the CDS of the identified genes and the previously reported sequences; (**b**) – synteny analysis of *FAE1* genes in allohexaploid *C. sativa* and its four diploid relatives; (**c**) – exon-intron structure of the identified *FAE1* genes; (**d**) – expression of *CsFAE1* from different subgenomes in various tissues; (**e**) – distribution of transmembrane domains within the putative protein product of the identified *FAE1* genes; (**f**) – sequence conservancy of *Camelina* FAE1 proteins with indication of the identified transmembrane domains; (**g**) – 3D structures of the identified FAE1 proteins in different *Camelina* species; (**h**) – RMSD values heatmap, showing degree of the structural differences between analysed FAE1 proteins

Table 5 Identified *FAE1* genes in five *Camelina* species

Name	Gene ID	Location	Strand
*CsFAE1-1-A**	104721341	11:3107019-3108726	-
*CsFAE1-2-A**	104721342	11:3110775-3112479	-
CsFAE1-B	104716684	10:2764395-2766099	-
CsFAE1-C	104729764	12:2883662-2885381	-
CnFAE1	-	5:2758201-2759718	-
ClFAE1	-	6:2574301-2575818	-
ChvhFAE1	-	5:3101298-3102815	-
ChvgFAE1	-	5:3101298-3102815	-

*Possibly genome assembly artefact: more recent *C. sativa* genome assembly (GCA_030686135.1, CP131541.1:2901200-2902717) contains only single *CsFAE1-A* gene, as well as the parental genome of *C. neglecta* does. From here and below referred as a single gene — *CsFAE1-A*

only 83.6% of all sites were invariable, possibly suggesting protein function diversification during the evolution of the genus. Preformed *FAE1* phylogeny reconstruction showed clear grouping of *CsFAE1* homeologs together with their orthologs from parental species (Fig. 3a). Representatives of the first *C. sativa* subgenome were placed in a joint clade with other *FAE1* of the N^6 genomic lineage (*CnFAE1*, previously reported *CsFAE1-A* and *CmFAE1-A*). Second copies of *FAE1*, coming from the genomes of *C. sativa* and *C. microcarpa* were separated into the sister clade of N^7 genome representatives.

Surprisingly, the third copy of *FAE1* from the genome of *C. microcarpa* and the *FAE1* of *C. rumelica* were placed in the basal braches to N^{6-7} group. This suggests that the third subgenome of *C. microcarpa* accession, used in the previous study [39], could be inherited not from the *C. hispida* species. In addition, *CsFAE1-C*, arising from the third subgenome of *C. sativa*, was expectedly placed together with its ortholog from *C. hispida*. Additional sequences of *C. hispida* infrataxa were shared the same clade with the *FAE1* of *C. laxa*. Higher sequence variability of *FAE1* genes, compared to above described desaturase genes, was also observed for homeologs from *C. sativa* genome. The *Ka/Ks* ratio for *CsFAE1-B* and *CsFAE1-C* was 0.093–0.116, while for *CsFAE1-A* this ratio reached 0.944 (Table S2). Such value close to 1 means that *CsFAE1-A* faced high mutation pressure after the allopolyploidy events that shaped *C. sativa*.

Synteny analysis showed formation of syntelogous pairs of *CsFAE1* genes with each of the genes from all four diploid *Camelina* genomes (Fig. 3b). As in the previous cases, the synteny analysis confirms orthology relations among *FAE1* genes of different *Camelina* species, since all the genes maintained in their origin loci and have not faced any translocations or duplications with the following gene loss. Exon-intron structure of the identified genes appears to be highly conserved, since all genes contain only one exon and preserve identical gene length (as well as they encode proteins of same length) (Fig. 3c).

All three *CsFAE1* genes are expressed at significantly lower levels, than two above described desaturases (Fig. 3d). The majority of the *CsFAE1* genes are not expressed elsewhere, except in seed during the early, early-mid and late-mid seed development stages. Only *CsFAE1-C* was expressed also in flower buds and flowers, but at very low levels (0.4–1.5 FPKM). Expression of exclusively this gene copy may indicate that it has special function and could be critical for the development of the mentioned plant tissues. *CsFAE1-B* was the most expressed during the early-mid seed development, exceeding *CsFAE1-C* by 1.8-fold and *CsFAE1-A* by 22.4-fold. During early seed development this gene was also expressed at significantly higher levels, 2.7-fold higher than *CsFAE1-C* and 28-fold higher that *CsFAE1-A*. Similarly, the dominance of *CsFAE1-B* was observed during late-mid seed development, as this gene was 1.6-fold and 14.9-fold higher expressed, than *CsFAE1-C* and *CsFAE1-A*, respectively.

Screening of 2 kbp upstream regions of the identified *FAE1* genes revealed presence of significant number of CAAT-box repeats (typically 31–41 per promoter region) and highly variable amount of TATA-box elements (Fig. S5a). While the majority of these genes contained 32–43 TATA-box motifs, *CsFAE1-C* had 53 repeats, while *ClFAE1* possessed as much as 101 TATA-box *cis*-elements. High number of MYB motifs (up to 8) was also detected in *CsFAE1-A* and *CnFAE1*, which is especially interesting, since *CsFAE1-A* has merely detectable expression (Fig. 3d).

Among the top ten most abundant TFBSs of *C. sativa FAE1* genes were primarily different types of DOF TFs, namely DOF1.5, DOF1.7, DOF2.2, DOF3.5, DOF3.6, DOF4.2, DOF5.1, DOF5.8, etc. (Fig. S5b, c, d). Genes *CsFAE1-A* and *CsFAE1-B* possessed presence of 5 sites for AP1, while *CsFAE1-C* had 12 of such motifs, related to flowering control. Interegstingly that *CsFAE1-C* also had 5 TFBSs for FLC (*Flowering Locus C*), a widely known TF that regulates flowering time in different species, including *C. sativa* [79]. At the same time, only the most expressed homeolog *CsFAE1-B* contained high number (seven) of predicted TFBSs for HDG1, HD-ZIP, homeodomain factor, which is believed to antagonistically control of cell proliferation [80]. Lastly, PISTILLATA (PI) sites associated with flowering regulation were detected in *CsFAE1-A* and *CsFAE1-C* [81], and several TFBSs of AHL25 (AT-hook TF) related to stem growth were detected in *CsFAE1-B*, which was not found to be expressed in stems or hypocotyls [82].

All identified FAE1 proteins were characterized by the presence of functional domains of fatty acid elongase at the conserved positions: FAE1/Type III polyketide synthase-like protein (PF08392) and or CHS_like (cd00831) (Fig. S6). The analysis of protein domain organization

also revealed presence of two transmembrane domains at positions 17–34 and 52–71, which were conserved for all FAE1 of the analyzed *Camelina* species (Fig. 3e). Both domains are located at the N-end of the peptide, while the large C-tail is expected to be exposed to the cytosol. In addition, amino acid substitutions were not evenly distributed across the whole length of the FAE1 proteins (Fig. 3f). While the TM1 of FAE1 tend to be conserved, the TM2 region contained seven variable positions, suggesting that almost a third of this domain is non-conserved. Other four variable positions were located in the region between TM1 and TM2, while the large C-tail carried 19 variable positions.

Structural conservancy on FAE1 homeologs in *C. sativa* was the highest among the investigated proteins with RMSD values in range of 1.05–1.47Å (Fig. 3g, h). It is noteworthy that all *C. sativa*, *C. neglecta* and *C. hispida* FAE1 proteins had highly similar structures (0.67-1,35Å). On contrary, the structure of ClFAE1 was the most different from other identified FAE1 genes (3.63–4.57Å) (Fig. 3h), which might explain lower levels of of gondoic (C20:1) and erucic (22:1) accumulation in seed lipids, described below. Considering the fact that *CsFAE1-A* is almost non-expressed in the genome of *C. sativa*, this gene might be undergoing pseudogenization due to its low expression and reduced functional importance.

Differential accumulation of fatty acids in seed lipids of *C. sativa* and its wild relatives

Composition of fatty acids, accumulated in seed lipids, represents a complex phenotype, shaped by the discussed above factors, like gene expression, sequence and structure of encoded protein product, etc. Combination of homeologous genes in polyploid organisms with complex evolutionary history, like members of *Camelina* genus, may differently affect accumulation of fatty acids in seeds of various species. Therefore, we also investigated differences in fatty acid content and composition in seed lipids of various *Camelina* species (Fig. 4, Table S3). The major component of seed oil of all *Camelina* species is α-linolenic (C18:3) acid, content of which ranged from 30 to 41% (Fig. 4a). However, the content of this particular fatty acid mostly was not significantly different within the analyzed species. Only *C. hipsida* var. *grandiflora* and *C. microcarpa* of Georgian origin showed significantly higher levels of the linolenic acid, compared to other species and *C. sativa*, in particular. Linoleic (C18:2), oleic (C18) and gondoic (C20:1) acids were less abundant, but still major components of the oil of *Camelina* species (Fig. 4a and b). The content of linoleic acid was also showed little or no difference for most of *Camelina* species, except *C. rumelica* and Georgian *C. microcarpa*, in which the amount of this fatty acid was significantly lower and ranged from 15 to 17%. The content of oleic acids was significantly different in the majority of species, besides *C. sativa* and *C. microcarpa* of German and Hungarian origin, and varied in the range of 12 to 19%. The maximal amount of oleic acid was identified in the seeds of *C. laxa*.

The relative content of palmitic (C16:0) and stearic (C18:0) acids was similar in the majority of the analyzed species and varied in their ranges of 7–8.5% and 2.3–3.8%, respectively (Fig. 4a). Similarly, arachidic (C20:0), eicosadienoic (C20:2), eicosatrienoic (C20:3), and behenic (C22:0) acids were present in minor quantities (≤2%) (Fig. 4b). The relative content of gondoic (C20:1) acid, one of the major in the *Camelina* seed lipids, varied from 10 to 16%. *C. microcarpa* and cultivated *C. sativa* and *C. alyssum* had lower levels of this fatty acid, typically ≤13%, similar to their wild relatives *C. hispida* and *C. laxa*. In addition, *C. neglecta* and *C. rumelica* seeds had higher levels of gondoic acid content, ~15%.

The most important VLCFA, erucic (C22:1) acid, was present in diverse quantities in different species, while its content significantly varied even within the same species, e.g. *C. microcarpa*. The content of this fatty acid in *C. sativa* seeds commonly consists about 2–3%, as well as in the majority of *C. microcarpa* accessions. In addition, *C. neglecta* seeds had nearly 2-fold higher relative amounts of erucic acid, ~4.7% on average, which was the highest amount of this fatty acid compared to other species. Conversely, *C. hispida* and *C. laxa* seeds had the lowest relative erucic acid content of this fatty acid (≤1.2%). This might be caused by peculiarities of *FAE1* gene regulatory elements, as it was discussed above (Fig. S5). Moreover, significant differences in ClFAE1 structure may also potentially be a reason for less effective erucic acid biosynthesis in *C. laxa*.

Since the description of the fatty acid profiles is a complicated task due to large number of diverse parameters (fatty acid content value), we have used specific coefficients that allow estimating the overall differences in desaturation/elongation pathways (Fig. 4c, d, e). The first noticeable trait of all *Camelina* species is prevalence of oleic acid desaturation pathway (DR – 0.6–0.71) above the elongation (ER – 0.14–0.23) (Fig. 4c). The highest rates of oleic acid elongation are inherent for *C. neglecta*, which could be explained by its highest content of VLCFAs. The lowest rates of oleic acid elongation were recorded in *C. hispida*, *C. laxa* and some of the replicates of *C. sativa* and *C. microcarpa* during the FA analysis repetition. More interesting, that along with the low ER rates *C. laxa* showed also low values of desaturation pathway share in FA biosynthesis.

A more detailed analysis of desaturation pathway (Fig. 4d) suggests that *C. laxa* stands apart from other *Camelina* species. While all of the analyzed species had oleic acid desaturation values (ODR) higher than 0.75 and

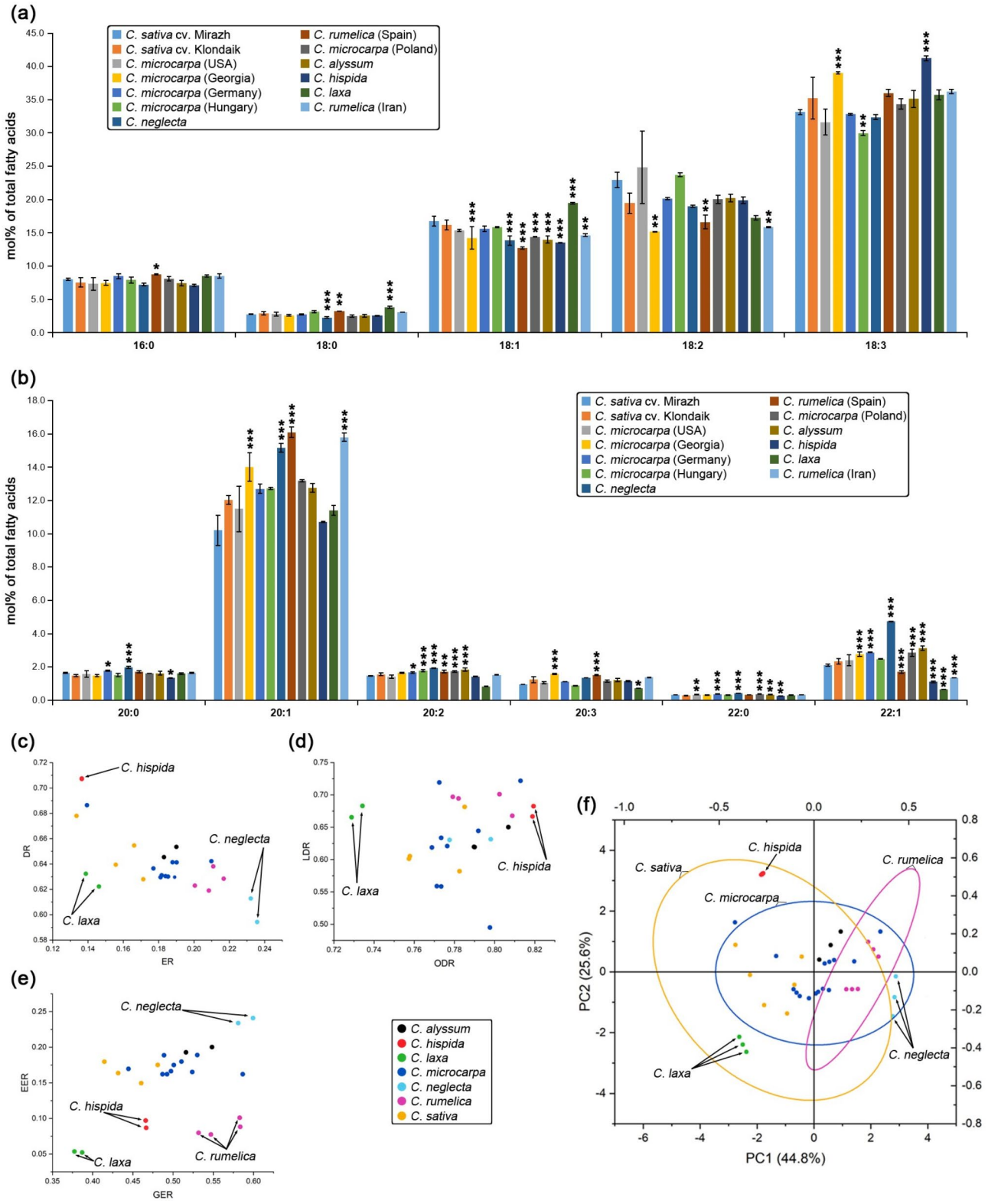

Fig. 4 Differences in fatty acid profiles of seed lipids of *Camelina* species: mean content of fatty acids (± STD) with chain length C16-C18 (**a**) and chain length C20-C22 (**b**); scatter plots showing relations between observed values of ER and DR coefficients (**c**), ODR and LDR (**d**), and GER and EER (**e**); and PCA plot showing interspecific differences in fatty acid composition of seed lipids (**f**). Content of a particular fatty acid, significantly different from *C. sativa*, is denoted with * - if significant at $p < 0.05$; ** - significant at $p < 0.01$; *** - significantly different at $p < 0.001$

up to 0.82, *C. laxa* showed the lowest rate of oleic acid conversion via the desaturation pathway (0.73), but still high rates of linoleic acid conversion. This may indicate that *C. laxa* could have decreased FAD2 activity, since the biosynthesis of linoleic acid out of oleic acid is decreased, but its conversion rate to linolenic acid remains the same. The other species were not so remarkably different by ODR and LDR values. The analysis of elongation pathway showed a significant distinction among the species (Fig. 4e). *C. neglecta* showed maximal values of both GER and EER values (0.59 and 0.24, respectively), suggesting its overall high activity of FAE1 enzyme, since it mediates both stages of C22:1 biosynthesis from C18:1. On contrary, *C. laxa* showed the minimal values for both coefficients (GER – 0.38; EER – 0.05), which is consistent with the observed lowest level of erucic acid in the seed of this species. *C. hispida* and *C. rumelica* have higher rates of elongation pathway activity, but still lower than those of *C. sativa*, *C. microcarpa* and *C. alyssum*.

The values range of the estimated coefficients for polyploid *C. sativa*, *C. alyssum*, *C. microcarpa* and *C. rumelica* always falls in the range between diploid *C. hispida* and *C. neglecta*. This could be well explained by the evolutionary history of these allopolyploids, since almost all of them inherited at least one subgenome from *C. hispida* and one or two from *C. neglecta* [13]. Therefore, an intermediate phenotype (in terms of fatty acid accumulation) may be observed for *C. sativa* species, despite the described above differences in expression of homeologs, inherited for distinct parental species. It is likely that, besides the analyzed *FAD2*, *FAD3* and *FAE1*, a complex interplay of other enzymes of fatty acid biosynthesis and accumulation may be involved into shaping the observed phenotypes of different species.

Similarly, semi-distinct type of FA biosynthesis in *C. laxa* is also well explained by its basal status for all *Camelina* genus [13]. The performed PCA analysis demonstrates well this effect (Fig. 4f). The variation ranges of polyploid species fell between *C. hispida* and *C. neglecta*. Moreover, allohexaploid species (*C. sativa*, *C. microcarpa*, *C. alyssum*), which have two *C. neglecta*-type subgenomes (N^{6-7}) tend to have FA biosynthesis more similar to *C. neglecta*. At the same time, *C. laxa* stands apart from this polyploid species complex and their parental taxa.

Discussion

Origin of *FAD2*, *FAD3* and *FAE1* panel could be well explained by*Camelina* genus evolution

In general, the studied (sub)genomes of *Camelina* species contained one copy of *FAD2*, *FAD3*, and *FAE1*. The investigated desaturases FAD2 and FAD3 had higher sequence conservancy rates, than FAE1, however all these genes were still highly conserved in terms of sequence diversity or genomic organization. Comparative genomics analysis did not reveal any duplications during *Camelina* evolution. Allohexaploid genome of *C. sativa* retained all three copies of either *FAD2*, *FAD3*, or *FAE1*, which were inherited diploid parental species. No evidence of pseudogenisation or significant sequence divergence effects were detected for these homeologous gene triplets. Considering the results of earlier studies, aimed on the identification of *FAD2* and *FAE1* genes in *Camelina* sp [39]. , allohexploid *C. microcarpa* and allotetraploid *C. rumelica* seem to preserve full sets of these desaturase/elongase genes, which were inherited during these species origin.

Evolution of the upstream promoter regions and gene expression may be two key components influencing homeologous copies subfunctionalization. Here we observed that the organization of gene upstream region and *cis*-acting elements composition *FAD2*, *FAD3*, or *FAE1* homeologs often reflected promoter organization of a parental gene in wild species. Despite the performed analyses appear to be bioinformatics prediction of possible TFBSs, requiring further experimental validation [48, 83]; some notable differences in upstream promoter sequence organization of homeologous genes may be elucidated. For instance, expressionally active homeologs usually contained TFBSs of TFs, associated with flowering regulation. The highest expressed *CsFAD2-C* homeolog contained numerous sites for RAMOSA1-like TFs (Fig. 1d, S1d), possibly controlling inflorescence architecture, seed size [73, 74]. At the same time, almost non-expressed *CsFAD2-A* had not signs of presence of such TFBSs (Fig. 1d, S1b). While such differences were not so obvious in *FAD3* homeologs, *FAE1* genes showed similar divergence of upstream promoter region. *CsFAE1-C* was the only homeolog showing combination of numerous TFBSs for different flowering-controlling TFs, like FLC (*Flowering Locus C*), PI (*PISTILLATA*) and prevailing by sites number AP1 (*APETALA1*) [78, 79, 81]. Apart from moderate expression rates of *CsFAE1-C* it might be suggested that this gene is an example of subfunctionalization via gaining tissue-specificity, which could be a possible evolution pathway of homeologous gene triplets in *C. sativa* [84]. It is important to consider whatever homeologus gene speciation took place, while developing strategies for gene silencing or editing in *C. sativa*.

Additional analyses of the 3D structure of the identified proteins allowed revealing additional differences among homeologous FAD2, FAD3 or FAE1 of *C. sativa*, which may appear not as obvious during the conventional sequence comparison. For instance, CsFAD2-A protein demonstrated significant changes in protein structural, compared to both its homeologs and orthologs, even from parental species *C. neglecta* (Fig. 1h). This fact coupled with significantly decreased expression

(up to 10-fold decreased from levels of homeologs) suggests that CsFAD2-A might be undergoing gradual pseudogenization, which was not detectable based on the sequence comparison solely or calculating *Ka/Ks* values. The homeologs of FAD3 and FAE1 showed no significant structural changes, compared to their orthologs in parental species. However, the analysis showed that FAD2, FAD3 and FAE1 proteins of *C. laxa* are the least conformationally similar to the majority of their orthologs within *Camelina* species. Such differences in conformation of these enzymes are consistent with the observed distinctions of FA profile of *C. laxa*. Decreased fatty acid elongation and desaturation rates in *C. laxa* may be conditioned by various factors, but protein structural difference could have also contributed to the observed phenotype of the species. However, the underlying reasons of this difference in FA accumulation in wild *Camelina* species may be the subject of a different study.

In parallel, a previous study revealed the presence of two (possibly paralogous) copies of *FAE1* in *C. laxa* and *C. hispida* [39]. However, the results of our genomic search have not confirmed such findings, since the investigated genomes of *C. laxa*, *C. hipida* var. *grandiflora* and var. *hispida* contained only a single *FAE1* gene each. It is very likely, that identified gene duplicates in these species could indeed be allelic variants of these genes [39]. The procedure used in the study could result in extraction of two distinct allelic variants, if the organism is hereterozygous. Moreover, it has been shown that *C. laxa* and *C. hispida* show higher genetic heterogeneity even within a particular line, which is rarely observed for higher ploidy species, like *C. sativa* or *C. microcarpa* [85].

Currently it is known that there are at least three distinct *C. microcarpa* cytotypes: Type 1 ($2n=40$), Type 2 ($2n=38$) and tetraploid cytotype ($2n=26$), also called *C. intermedia* [15]. Each of these cytotypes is believed to have different genome composition. Type 1 inherited two *C. neglecta*-type genomes (N^6 and N^7) and one *C. hispida*-type (H^7), while Type 2 might have inherited three *C. neglecta*-like subgenomes ($N^6N^7N^6$) [13]. Only *C. microcarpa* Type 1 is believed to be a direct ancestor of the cultivated *C. sativa* [19]. The *CmFAD2* and *CmFAE1* genes investigated here most likely belong to Type 2 *C. microcarpa* with altered third genome. In case of *FAD2* phylogeny reconstruction, no reliable grouping of H^7 (sub)genome sequences was obtained (Fig. 1a). The third *CmFAD2* homeolog (GU929433.1) shared clade with one of CrFAD2 homeologs (GU929439.1), while origin of the latter in unclear (it could be either from N^6 or from H^7 subgenome component of *C. rumelica*). In the case of *FAE1* phylogeny reconstruction, a more reliable topology was reconstructed (Fig. 3a). While the sequences of N^7 (sub)genome origin were grouped separately, all FAE1 sequences of *C. microcarpa* were placed into N^{6-7} (sub) genome clade: two with their orthologs from *C. sativa* and *C. neglecta*, and one (GU929437.1) was placed as a basal branch for the group. All this suggests that the accession of *C. microcarpa*, used by Hutcheon et al. [39]. in their study for the sequencing, might belong to the Type 2 of *C. microcarpa*, which has a distinct genome organization [13, 17].

Diversity of *FAD2*, *FAD3* and *FAE1* can be exploited for *C. sativa* improvement

It has been shown that different *Camelina* species have unique FA biosynthesis-related traits, despite the general feature of relatively high content polyunsaturated FAs (Fig. 4). This is consistent with the results of other previous investigations [5, 11, 14]. In many cases, the content of a particular fatty acid was not significantly different among various *Camelina* species. However, there were notable differences between several species.

For example, *C. neglecta* showed a significant increase in erucic (C22:1) acid biosynthesis, compared to other *Camelina* species and lower levels of oleic (C18:1) and linolenic (C18:2) acids (Fig. 4a, b). The same was previously shown in the research that compared FA profiles of seed lipids of *C. sativa* and *C. neglecta* [41]. It was supposed that the lower accumulation of erucic acid accumulation in *C. sativa* might be caused by expression differences between the subgenomes [41], especially taking in account the transcriptional dominance of the third (H^7) subgenome [17]. Here we showed that the expression of *CsFAE1* homeologs (main gene regulating erucic acid biosynthesis) significantly differed. In particular, *CsFAE1-A*, the gene from N^6 subgenome that was inherited directly from *C. neglecta*, showed the lowest expression among all three *CsFAE1* homeologs in all investigated tissues (Fig. 3d).

The similar effect was observed for differences in FA composition between *C. sativa* and *C. hispida* (Fig. 4a, b), which may be explained by the suppression or increased expression of FAD2/FAD3 genes from H^7 subgenome (Figs. 1d and 2d). *C. sativa* demonstrates similar amounts of linoleic (C18:2) acid to *C. hispida*, having *CsFAD2-C* expressed at the higher level, than other homeologs. However, at the same time *C. sativa* shows significantly lower amounts of linolenic (C18:3) acid, when the *CsFAD3-C* gene from H^7 subgenome is being suppressed, if compared with *CsFAD3-A* and *CsFAD3-B*. *C. hispida* the highest content of α-linolenic acid among all investigated *Camelina* species (42% on average). These examples show that the third subgenome of *C. sativa* may exhibit not in all cases [17]. Similar patterns of differential expression of *FAD2*, *FAD3* and *FAE1* homeologs in *C. sativa* during mid-stage of seed maturation (called early-mid stage in our study) were also observed earlier [86].

Another notable example is *C. laxa*, which has the lowest accumulation rates of erucic acid, compared to other *Camelina* species (Fig. 4b). The content of this FA in *C. laxa* was 3.3-3.6-fold lower than in *C. sativa* and 7.4-fold lower than in *C. neglecta*. Additionally, *C. rumelica* and one of its parental species, *C. neglecta*, both showed the highest content of gondoic acid (15.16–16.07%, Table S3), compared to other *Camelina* sp. The same differences among wild *Camelina* species were reported, except *C. neglecta*, since authors had not analyzed it [14]. Previous studies also reported that *C. sativa* has the highest total content of FA among other wild relatives [14]. However, the accumulation of seed lipids is controlled by other genes, which were not investigated in the present study, but may the subject of future research. Genetic manipulation of these genes may also help alter the seed lipid accumulation in the cultivated false flax [33].

Previously, number of mutations in the investigated *FAD2*, *FAD3* and *FAE1* in *C. sativa* were reported that are causing the alteration of FA accumulation in seeds [87]. The most significant effect is caused by deleterious mutations, leading to the protein truncation. However, several critical positions were identified, which may affect the efficiency of these enzymes functioning [87]. For instance, it has been shown that G150E mutation in *CsFAD2-B* (*fad2a* in original publication), which is located to the functionally important His-box motif, leads to decrease of linoleic and linolenic acids content. Similarly, the mutation of *CsFAD3-B* (*fad3a*) in TM2, close to His-box (G101S), leads to the reduction of linolenic acid content in *C. sativa* seeds [87]. It is worth noting that both positions in FAD2 and FAD3 were found to be conserved among *Camelina* species (Figs. 1f and 2f). The mutation P141L of *CsFAE1-A* (*fae1a*), involving conserved a.a. position among both *Camelina* sp. and *A. thaliana* (however, located in non-conserved region, Fig. 3f), also was previously shown to result in lack of function of this enzyme, leading to decrease if VLCFA content [87].

The targeting of a particular *FAD*/*FAE* gene copy for gene editing may be the most efficient strategy, especially taking in account differential expression of the homeologs in *C. sativa*. It was shown that CRISPR/Cas9-mediated mutations in *CsFAE1-B* lead to the most significant decrease in VLCFA synthesis, compared to other homeologs [34]. Here we show that *CsFAE1-B* is the highest expressed homeolog in *C. sativa* (Fig. 3d), what conditions its significant role in biosynthesis of gondoic and erucic acids in this species. It was also demonstrated that all three *CsFAD2* homeologs could be simultaneously targeted by CRISPR/Cas9 [35–37], even at multiple sites due to high conservancy of *CsFAD2* sequences (Fig. 1f). The authors of these studies targeted regions that encode either transmembrane domains or adjacent regions. However, the majority of the generated mutants appeared to be FAD2-knockouts, resulting from the frameshift mutations.

It was shown that gene dosage regulation allows altering FA composition of seed lipids in such way that a desirable rate of oleic acid accumulation can be achieved [35, 36]. Moreover, increase in number of substitutions and indels affected similarly to accumulation of frameshift mutations in *CsFAD2* gene copies. However, homozygous triple *fad2*-knockouts showed drastic developmental defects, with strong aberrant phenotype and growth delay [36]. More optimal strategy is targeting lower number of *CsFAD2* copies, which is partially complicated by high sequence conservancy of the homeologs [35]. In this case only the most expressed copies of *CsFAD2*, *CsFAD3* and *CsFAE1* homeologs might be targeted, retaining the least expressed, which could allow avoiding unfavorable phenotypic effects in the mutant progeny.

Introduction of *FAD2*, *FAD3* or *FAE1* alleles to *C. sativa* from wild germplasm may be of interest for breeding, aimed on the improvement of this emerging crop. However, low crossability of *C. sativa* with lower ploidy relatives may limit the utility of this approach [13]. Therefore, transgenesis or genome editing remains the most considerable option for FA composition redesign in false flax. However, the challenges of achieving efficient and precise edits in allopolyploid genome of *C. sativa* remain significant. Despite that, high similarity of *C. sativa* subgenomes with such relatives as tetraploid *C. intermedia* [15] or diploid *C. neglecta* [16, 40–42] allows use of this species as models for *C. sativa* biotechnology research. Additionally, increase of *C. sativa* genetic diversity via hybridization with *C. microcarpa* is also viewed as a highly promising breeding approach [13, 14, 18, 19].

Proposed strategy for practical genome editing applications and future perspectives

Taking into consideration the observed gene diversity, insights for future genome editing strategy can be elucidated. Especially accounting the differential expression and structural features of homeologous copies of *FAD2*, *FAD3*, and *FAE1*, these genes possess a great interest and the potential for precise genetic manipulation. Among such approach could be:

i. **Selective Editing (Knockout) of Highly Expressed Gene Copies**: CRISPR/Cas9 can be utilized to target the highest expressed homeologs, in order to achieve the desired changes in FA content without negatively affecting vital organism functions, thus, precisely regulate gene dosage. For, instance, such homeolog as *CsFAE1-B* might be targeted to reduce VLCFA synthesis without affecting growth or development.

Retaining less-expressed homeologs would mitigate deleterious phenotypic effects.

ii. **Promoter and TFBSs Engineering**: Based on the observed upstream regulatory differences, editing or engineering such promoter regions could enhance or reduce specific gene expression in desired tissues. For instance, introduction of disruption sites, regulating seed-specific expression, could be a possible way to regulated specific FA accumulation in seed lipids, without inducing large-scale phenotype abnormalities. Noteworthy, not all TFBSs might be amenable for CRISPR/Cas9 editing due to absence of PAM-site motif (-NGG), required for successful target site recognition cleavage.

iii. **Gene Silencing via RNA Interference (RNAi), miRNA or Antisense Oligonucleotides (ASOs)**: as in previous examples the desired homeolog (likely the most expressed) may be targeted for specific silencing or partial knock-down, achieving desired FA composition, minimizing unintended phenotypic effects. However, high sequence similarity of homeologous genes may appear a significant obstacle or could cause undesired off-targets.

iv. **Introgression of Alleles from Wild Germplasm**: Incorporating advantageous allelic variants from wild relatives, especially from such distant species, like *C. laxa*, may have significant effect on fatty acid composition. Optimally, this can be achieved via transgenesis, considering the low crossability of *C. sativa* with its lower ploidy relatives.

v. **Adaptive Protein Engineering**: In silico techniques may be applied, in order to designe new FAD2, FAD3 or FAE1 structural variants with increased or reduced efficiency of FA desaturation/elongation. Rational design or directed evolution approaches could be applied for enzyme optimization. Later, nucleotide sequences, encoding such designer proteins, could be created and introduced in *C. sativa* via transgenic approaches.

The proposed approaches, grounded on the performed genomic, analyses cover modern genome editing and metabolic engineering techniques, which may be applied to optimize FA accumulation in *C. sativa*, providing the roadmap for improvement of this prominent oilseed crop.

Conclusions

Present study reports a comprehensive and detailed analysis of *FAD2*, *FAD3* and *FAE1* gene diversity in five *Camelina* species, in particular: hexaploid *C. sativa* and four diploids, namely *C. neglecta*, *C. laxa*, *C. hispida* var *hispida* and var. *grandiflora*. The analyzed genes retained high sequence conservancy rate and retained in triplets in allohexaploid *C. sativa*. Subfunctionalization of these gene homeologs in *C. sativa* most likely was conditioned by the divergence of expression patterns, which is potentially related to observed distinctions in upstream promoters organization. Notable, variation of FA composition in wild different species and in *C. sativa* is believed to be conditioned by several factors, including possible gene regulatory elements differences and FAD2, FAD3 and FAE1 protein conformation. These differences in FA accumulation highlight the potential of natural diversity of wild *FAD2*, *FAD3* and *FAE1* alleles that might be introgressed to *C. sativa*, in order to boost genetic heterogeneity of cultivated false flax. The described above findings provide a basis for a variety of strategies for future *Camelina* research and breeding. Gene editing of particular *FAD2*, *FAD3*, and *FAE1* copies in *C. sativa*, which could exhibit high expression, could enable precise regulation of gene dosage of the mentioned genes and development plants with desirable seed lipids FA profiles, potentially avoiding negative phenotypic effects. Alternatively, editing of upstream regulatory elements of *FAD2*, *FAD3*, and *FAE1* may provide other possible ways for accurate inactivation expression of these enzymes, thus potentially enabling targeted manipulation of FA accumulation in seeds. Finally, wild relatives gene diversity offers possibility for direct transgenesis, avoiding direct interspecific hybridization with low efficiency. These approaches together provide insights for advancing *C. sativa* as a high-performing oilseed crop, addressing both economic and sustainability goals.

Abbreviations

FA	Fatty acid
TF	Transcription factor
TFBS	Transcription factor binding site

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12896-024-00936-4.

Supplementary Material 1

Acknowledgements

Not applicable.

Author contributions

RYB, VYH, TJN participated in the research design and data collection and data analysis. RYB participated in research design and draft manuscript preparation. EBC and YBB performed the research design, manuscript writing and editing and supervised the research project. All authors read and approved the final manuscript.

Funding

This work was supported by the Civilian Research and Development Foundation (CRDF Global) (2018–2019) [Grant No.: 63881/63882] (for RYB, TJN, EBC, YBB); Camelina genomics research in YBB lab was supported by the project of National Academy of Sciences of Ukraine [State registration No. 0123U102104 under Program 6541230] (for VYH, YBB). EBC was funded by the US Department of Energy, Office of Science, Office of Biological

& Environmental Research, under Award Number DE-SC0023142 and US Department of Agriculture-National Institute of Food and Agriculture, grant no. 2019-67021-29946.

Data availability

Whole genome sequences information for *C. sativa* (GCA_000633955.1), *C. neglecta* (GCA_023864065.1), *C. laxa* (GCA_024034495.1), *C. hispida* var. *hispida* (GCA_023657505.1) and *C. hispida* var. *grandiflora* (GCA_023864115.1) were obtained from the NCBI Genome database (https://www.ncbi.nlm.nih.gov/datasets/genome/). All accession numbers of the sequences, used for phylogeny reconstruction are listed within Table S1. The transcriptomics data of *C. sativa* (cv. DH55) were obtained from BAR ePlant database (https://bar.utoronto.ca/). The datasets supporting the conclusions of this study are included in the article and in additional files. All accessions of used plant genotypes are listed within Table 2. The data used for gene upstream region analysis are available at PlantCARE (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/) and JASPAR-2024 (https://jaspar.elixir.no/) databases.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Institute of Food Biotechnology and Genomics of National Academy of Sciences of Ukraine, 2a Baidy-Vyshnevetskoho str., Kyiv 04123, Ukraine
[2]Center for Plant Science Innovation & Department of Biochemistry, University of Nebraska-Lincoln, E318 Beadle Center, 1901 Vine Street, Lincoln, NE 68588, USA

Received: 7 August 2024 / Accepted: 12 December 2024
Published online: 18 December 2024

References

1. Iskandarov U, Silva JE, Kim HJ, Andersson M, Cahoon RE, Mockaitis K, Cahoon EB. A specialized diacylglycerol acyltransferase contributes to the extreme medium-chain fatty acid content of *Cuphea* seed oil. Plant Physiol. 2017;174:97–109. https://doi.org/10.1104/pp.16.01894.
2. Yuan L, Li R. Metabolic engineering a model oilseed *Camelina sativa* for the sustainable production of high-value designed oils. Front Plant Sci. 2020;11:11. https://doi.org/10.3389/fpls.2020.00011.
3. Iskandarov U, Kim HJ, Cahoon EB. Camelina: an emerging oilseed platform for advanced biofuels and bio-based materials. In: McCann MC, Buckeridge MS, Carpita NC, editors. Plants and bioenergy. New York, NY: Springer 2014:131–40. https://doi.org/10.1007/978-1-4614-9329-7_8.
4. Resurreccion EP, Roostaei J, Martin MJ, Maglinao RL, Zhang Y, Kumar S. The case for camelina-derived aviation biofuel: Sustainability underpinnings from a holistic assessment approach. Ind Crop Prod. 2021;170:113777. https://doi.org/10.1016/j.indcrop.2021.113777.
5. Zanetti F, Alberghini B, Jeromela AM, Grahovac N, Rajkovic D, Kiprovski B, Monti A. Camelina, an ancient oilseed crop actively contributing to the rural renaissance in Europe. A review. Agron Sustain Dev. 2021;41:2. https://doi.org/10.1007/s13593-020-00663-y.
6. Ghidoli M, Ponzoni E, Araniti F, Miglio D, Pilu R. Genetic improvement of *Camelina sativa* (L.) Crantz: opportunities and challenges. Plants. 2023;12:570. https://doi.org/10.3390/plants12030570.
7. Liu X, Brost J, Hutcheon C, Guilfoil R, Wilson AK, Leung S, Shewmaker CK, Rooke S, Nguyen T, Kiser J, De Rocher J. Transformation of the oilseed crop *Camelina sativa* by Agrobacterium-mediated floral dip and simple large-scale screening of transformants. Vitro Cell Developml Biol – Plant. 2012;48:462–8. https://doi.org/10.1007/s11627-012-9459-7.
8. Yemets AI, Boychuk YN, Shysha EN, Rakhmetov DB, Blume YB. Establishment of in vitro culture, plant regeneration, and genetic transformation of *Camelina sativa*. Cytol Genet. 2013;47(3):138–44. https://doi.org/10.3103/S0095452713030031.
9. Zubr J. Oil-seed crop: *Camelina sativa*. Ind Crop Prod. 1997;6:113–9. https://doi.org/10.1016/S0926-6690(96)00203-8.
10. Manca A, Pecchia P, Mapelli S, Masella P, Galasso I. Evaluation of genetic diversity in a *Camelina sativa* (L.) Crantz collection using microsatellite markers and biochemical traits. Genet Resour Crop Evol. 2012;60:1223–6. https://doi.org/10.1007/s10722-012-9913-8.
11. Vollmann J, Eynck C. Camelina as a sustainable oilseed crop: Contributions of plant breeding and genetic engineering. Biotechnol J. 2015;10:525–35. https://doi.org/10.1002/biot.201400200.
12. Luo Z, Brock J, Dyer JM, Kutchan T, Schachtman D, Augustin M, Ge Y, Fahlgren N, Abdel-Haleem H. Genetic diversity and population structure of a *Camelina sativa* spring panel. Front Plant Sci. 2019;10:184. https://doi.org/10.3389/fpls.2019.00184.
13. Blume RY, Kalendar RN, Guo L, Cahoon EB, Blume YB. Overcoming genetic paucity of *Camelina sativa*: possibilities for interspecific hybridization conditioned by the genus evolution pathway. Front Plant Sci. 2023;14:1259431. https://doi.org/10.3389/fpls.2023.1259431.
14. Brock JR, Scott T, Lee AY, Mosyakin SL, Olsen KM. Interactions between genetics and environment shape *Camelina* seed oil composition. BMC Plant Biol. 2020;20:423. https://doi.org/10.1186/s12870-020-02641-8.
15. Mandáková T, Lysak MA. The identification of the missing maternal genome of the allohexaploid camelina (*Camelina sativa*). Plant J. 2022;112:622–9. https://doi.org/10.1111/tpj.15931.
16. Mándaková T, Pouch M, Brock JR, Al-Shehbaz IA, Lysak MA. Origin and evolution of diploid and allopolyploid *Camelina* genomes were accompanied by chromosome shattering. Plant Cell. 2019;31(11):2596–612. https://doi.org/10.1105/tpc.19.00366.
17. Chaudhary R, Koh CS, Kagale S, Tang L, Wu SW, Lv Z, Mason AS, Sharpe AG, Diederichsen A, Parkin IAP. Assessing diversity in the *Camelina* genus provides insights into the genome structure of *Camelina sativa*. Genes Genomes Genet. 2020;G3(4):1297–308. https://doi.org/10.1534/g3.119.400957.
18. Brock JR, Mandáková T, McKain M, Lysak MA, Olsen KM. Chloroplast phylogenomics in *Camelina* (Brassicaceae) reveals multiple origins of polyploid species and the maternal lineage of *C. sativa*. Hortic Res. 2022a;9:uhab050. https://doi.org/10.1093/hortre/uhab050.
19. Brock JR, Ritchey MM, Olsen KM. Molecular and archaeological evidence on the geographical origin of domestication for *Camelina sativa*. Am J Bot. 2022b;109(7):1177–90. https://doi.org/10.1002/ajb2.16027.
20. Sakharova VH, Blume RY, Rabokon AN, Pirko YV, Blume YB. Efficiency of genetic diversity assessment of little-pod false flax (*Camelina microcarpa* Andrz. ex DC.) in Ukraine using SSR- and TBP-marker systems. Rep Natl Acad Sci Ukraine. 2023;4:85–94. https://doi.org/10.15407/dopovidi2023.04.093.
21. Mansour MP, Shrestha P, Belide S, Petrie JR, Nichols PD, Singh SP. Characterization of oilseed lipids from DHA-poducing *Camelina sativa*: A new transformed land plant containing long-chain omega-3 oils. Nutrients. 2014;6:776–89. https://doi.org/10.3390/nu6020776.
22. Petrie JR, Shrestha P, Belide S, Kennedy Y, Lester G, Liu Q, Divi UK, Mulder RJ, Mansour MP, Nichols PD, Singh SP. Metabolic engineering *Camelina sativa* with fish oil-like levels of DHA. PLoS ONE. 2014;9(1):e85061. https://doi.org/10.1371/journal.pone.0085061.
23. Ruiz-Lopez N, Haslam RP, Napier JA, Sayanova O. Successful high-level accumulation of fish oil omega-3 long-chain polyunsaturated fatty acids in a transgenic oilseed crop. Plant J. 2014;77:198–208. https://doi.org/10.1111/tpj.12378.
24. Betancor MB, Li K, Sprague M, Bardal T, Sayanova O, Usher S, Han L, Måsøval K, Torrissen O, Napier JA, Tocher DR, Olsen RE. An oil containing EPA and DHA from transgenic *Camelina sativa* to replace marine fish oil in feeds for Atlantic salmon (*Salmo salar* L.): Effects on intestinal transcriptome, histology, tissue fatty acid profiles and plasma biochemistry. PLoS ONE. 2017;12(4):e0175415. https://doi.org/10.1371/journal.pone.0175415.
25. Betancor MB, MacEwan A, Sprague M, Gong X, Montero D, Han L, Napier JA, Norambuena F, Izquierdo M, Tocher DR. Oils from transgenic *Camelina sativa* as a source of EPA and DHA in feeds for European sea bass (*Dicentrarchus labrax* L). Aquaculture. 2021;530:735759. https://doi.org/10.1016/j.aquaculture.2020.735759.
26. Tjellström H, Strawsine M, Silva J, Cahoon EB, Ohlrogge JB. Disruption of plastid acyl:acyl carrier protein synthetases increases medium chain fatty acid accumulation in seeds of transgenic *Arabidopsis*. FEBS Lett. 2013;587:936–42. https://doi.org/10.1016/j.febslet.2013.02.021.

27. Kim HJ, Silva JE, Iskandarov U, Andersson M, Cahoon RE, Mockaitis K, Cahoon EB. Structurally divergent lysophosphatidic acid acyltransferases with high selectivity for saturated medium chain fatty acids from *Cuphea* seeds. Plant J. 2015a;84:1021–33. https://doi.org/10.1111/tpj.13063.
28. Kim HJ, Silva JE, Vu HS, Mockaitis K, Nam JW, Cahoon EB. Toward production of jet fuel functionality in oilseeds: identification of FatB acyl-acyl carrier protein thioesterases and evaluation of combinatorial expression strategies in *Camelina* seeds. J Exp Bot. 2015b;66:4251–65. https://doi.org/10.1093/jxb/erv225.
29. Bansal S, Kim HJ, Na GN, Hamilton ME, Cahoon EB, Lu C, Durrett TP. Towards the synthetic design of camelina oil enriched in tailored acetyl-triacylglycerols with medium-chain fatty acids. J Exp Bot. 2018;69(18):4395–402. https://doi.org/10.1093/jxb/ery225.
30. Augustin JM, Brock JR, Augustin MM, Wellinghoff RL, Shipp M, Higashi Y, Kumssa TT, Cahoon EB, Kutchan TM. Field performance of terpene-producing *Camelina sativa*. Ind Crop Prod. 2019;136:50–8. https://doi.org/10.1016/j.indcrop.2019.04.061.
31. Hölzl G, Rezaeva BR, Kumlehn J, Dörmann P. Ablation of glucosinolate accumulation in the oil crop *Camelina sativa* by targeted mutagenesis of genes encoding the transporters GTR1 and GTR2 and regulators of biosynthesis MYB28 and MYB29. Plant Biotechnol J. 2023;21:189–201. https://doi.org/10.1111/pbi.13936.
32. Lyzenga WJ, Harrington M, Bekkaoui D, Wigness M, Hegedus DD, Rozwadowski KL. CRISPR/Cas9 editing of three CRUCIFERIN C homoeologues alters the seed protein profile in *Camelina sativa*. BMC Plant Biol. 2019;19:292. https://doi.org/10.1186/s12870-019-1873-0.
33. Aznar-Moreno JA, Durrett TP. Simultaneous targeting of multiple gene homeologs to alter seed oil production in *Camelina sativa*. Plant Cell Physiol. 2017;58:1260–7. https://doi.org/10.1093/pcp/pcx058.
34. Ozseyhan ME, Kang J, Mu X, Lu C. Mutagenesis of the *FAE1* genes significantly changes fatty acid composition in seeds of *Camelina sativa*. Plant Physiol Biochem. 2018;123:1–7. https://doi.org/10.1016/j.plaphy.2017.11.021.
35. Jiang WZ, Henry IM, Lynagh PG, Comai L, Cahoon EB, Weeks DP. Significant enhancement of fatty acid composition in seeds of the allohexaploid, *Camelina sativa*, using CRISPR/Cas9 gene editing. Plant Biotechnol J. 2017;15:648–57. https://doi.org/10.1111/pbi.12663.
36. Morineau C, Bellec Y, Tellier F, Gissot L, Kelemen Z, Nogué F, Faure J-D. Selective gene dosage by CRISPR-Cas9 genome editing in hexaploid *Camelina sativa*. Plant Biotechnol J. 2017;15:729–39. https://doi.org/10.1111/pbi.12671.
37. Lee K-R, Jeon I, Yu H, Kim S-G, Kim H-S, Ahn S-J, Lee J, Lee S-K, Kim HU. Increasing monounsaturated fatty acid contents in hexaploid *Camelina sativa* seed oil by *FAD2* gene knockout using CRISPR-Cas9. Front Plant Sci. 2021;12:702930. https://doi.org/10.3389/fpls.2021.702930.
38. Kagale S, Koh C, Nixon J, Bollina V, Clarke WE, Tuteja R, Spillane C, Robinson SJ, Links MG, Clarke C, Higgins EE, Huebert T, Sharpe AG, Parkin IAP. The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. Nat Commun. 2014;5:3706. https://doi.org/10.1038/ncomms4706.
39. Hutcheon C, Ditt RF, Beilstein M, Comai L, Schroeder J, Goldstein E, Shewmaker CK, Nguyen T, De Rocher J, Kiser J. Polyploid genome of *Camelina sativa* revealed by isolation of fatty acid synthesis genes. BMC Plant Biol. 2010;10:233. https://doi.org/10.1186/1471-2229-10-233.
40. Martin SL, Lujan Toro B, James T, Sauder CA, Laforest M. Insights from the genomes of 4 diploid *Camelina* spp. G3: Genes, Genomes, Genetics. 2022;12(12):jkac182. https://doi.org/10.1093/g3journal/jkac182
41. Chaudhary R, Koh CS, Perumal S, Jin L, Higgins EE, Kagale S, Smith MA, Sharpe AG, Parkin IAP. Sequencing of *Camelina neglecta*, a diploid progenitor of the hexaploid oilseed *Camelina sativa*. Plant Biotechnol J. 2023;21:521–35. https://doi.org/10.1111/pbi.13968.
42. Wang S, Blume RY, Zhou Z-W, Lu S, Nazarenus TJ, Blume YB, Xie W, Cahoon EB, Chen L-L, Guo L. Chromosome-level assembly and analysis of *Camelina neglecta* – a novel diploid model for camelina biotechnology research. Biotechnol Biofuels Bioprod. 2024;17:17. https://doi.org/10.1186/s13068-024-02466-9.
43. Hatje K, Keller O, Hammesfahr B, Pillmann H, Waack S, Kollmar M. Cross-species protein sequence and gene structure prediction with fine-tuned Webscipio 2.0 and Scipio. BMC Res Notes. 2011;4:265. https://doi.org/10.1186/1756-0500-4-265.
44. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acid Res. 2004;32(5):1792–7. https://doi.org/10.1093/nar/gkh340.
45. Hu B. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2015;31(8):1296–7. https://doi.org/10.1093/bioinformatics/btu817.
46. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S. PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. Nucleic Acids Res. 2002;30(1):325–7. https://doi.org/10.1093/nar/30.1.325.
47. Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, Liu Y, Feng J, Chen H, He Y, Xia R. TBtools-II: A one for all, all for one bioinformatics platform for biological big-data mining. Mol Plant. 2023;16:1733–42. https://doi.org/10.1016/j.molp.2023.09.010.
48. Rauluseviciute I, Riudavets-Puig R, Blanc-Mathieu R, Castro-Mondragon JA, Ferenc K, Kumar V, Lemma RB, Lucas J, Chèneby J, Baranasic D, Khan A, Fornes O, Gundersen S, Johansen M, Hovig E, Lenhard B, Sandelin A, Wasserman WW, Parcy F, Mathelier A. JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2024;52(D1):D174–82. https://doi.org/10.1093/nar/gkad1059.
49. Korhonen JH, Palin K, Taipale J, Ukkonen E. Fast motif matching revisited: high-order PWMs, SNPs and indels. Bioinformatics. 2017;33(4):514–21. https://doi.org/10.1093/bioinformatics/btw683.
50. Kagale S, Nixon J, Khedikar Y, Pasha A, Provart NJ, Clarke WE, Bollina V, Robinson SJ, Coutu C, Hegedus DD, Sharpe AG, Parkin IAP. The developmental transcriptome atlas of the biofuel crop *Camelina sativa*. Plant J. 2016;88:879–94. https://doi.org/10.1111/tpj.13302.
51. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, Gough J, Haft DH, Letunić I, Marchler-Bauer A, Mi H, Natale DA, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A. InterPro in 2022. Nucleic Acids Res. 2023;51(D1):D418–27. https://doi.org/10.1093/nar/gkac993.
52. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021;49(D1):D412–9. https://doi.org/10.1093/nar/gkaa913.
53. Wang J, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ, Marchler-Bauer A. The conserved domain database in 2023. Nucleic Acids Res. 2023;51(D1):D384–8. https://doi.org/10.1093/nar/gkac1096.
54. Hallgren J, Tsirigos KD, Pedersen MD, Almagro Armenteros JJ, Marcatili P, Nielsen H, Krogh A, Winther O. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv. 2022. https://doi.org/10.1101/2022.04.08.487609. Accessed 06 Aug 2024.
55. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305(3):567–80. https://doi.org/10.1006/jmbi.2000.4315.
56. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23:2947–8. https://doi.org/10.1093/bioinformatics/btm404.
57. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nat Methods. 2022;19:679–82. https://doi.org/10.1038/s41592-022-01488-1.
58. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009;25(11):1422–3. https://doi.org/10.1093/bioinformatics/btp163.
59. Sehnal D, Bittrich S, Deshpande M, Svobodová R, Berka K, Bazgier V, Velankar S, Burley SK, Koča J, Rose AS. Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. Nucleic Acids Res. 2021;49(W1):W431–7. https://doi.org/10.1093/nar/gkab314.
60. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat Meth. 2017;14:587–9. https://doi.org/10.1038/nmeth.4285.
61. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol Biol Evol. 2015;32:268–74. https://doi.org/10.1093/molbev/msu300.
62. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 2016;44(W1):W232–5. https://doi.org/10.1093/nar/gkw256.
63. Hoang DT, Chernomor O, von Haeseler A, Minhnand BQ, Vinh LS. UFBoot2: Improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35(2):518–22. https://doi.org/10.1093/molbev/msx281.

64. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acid Res. 2024;52(W1):W78–82. https://doi.org/10.1093/nar/gkae268.
65. Stanke M, Schöffmann O, Morgenstern B, Waack S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics. 2006;7(1):62. https://doi.org/10.1186/1471-2105-7-62.
66. Brock JR, Mandáková T, Lysak MA, Al-Shehbaz IA. *Camelina neglecta* (*Brassicaceae*, *Camelineae*), a new diploid species. Europe PhytoKeys. 2019;115:51–7. https://doi.org/10.3897/phytokeys.115.31704.
67. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49. https://doi.org/10.1093/nar/gkr1293.
68. Cahoon EB, Dietrich CR, Meyer K, Damude HG, Dyer JM, Kinney AJ. Conjugated fatty acids accumulate to high levels in phospholipids of metabolically engineered soybean and *Arabidopsis* seeds. Phytochemistry. 2006;67:1166–76. https://doi.org/10.1016/j.phytochem.2006.04.013.
69. Velasco L, Goffman FD, Becker HC. Variability for the fatty acid composition of the seed oil in a germplasm collection of the genus *Brassica*. Genet Resour Crop Evol. 1998;45:371–82. https://doi.org/10.1023/A:100862862.
70. Pleines S, Friedt W. Breeding for improved C18-fatty acid composition in rapeseed (*Brassica napus* L). Fat Sci Technol. 1988;90:167–71. https://doi.org/10.1002/lipi.19880900502.
71. Martín G, Veciana N, Boix M, Rovira A, Henriques R, Monte E. The photoperiodic response of hypocotyl elongation involves regulation of CDF1 and CDF5 activity. Physiol Plant. 2020;169(3):480–90. https://doi.org/10.1111/ppl.13119.
72. Petrella R, Caselli F, Roig-Villanova I, Vignati V, Chiara M, Ezquer I, Tadini L, Kater MM, Gregis V. BPC transcription factors and a Polycomb Group protein confine the expression of the ovule identity gene *SEEDSTICK* in *Arabidopsis*. Plant J. 2020;102(3):582–99. https://doi.org/10.1111/tpj.14673.
73. Vollbrecht E, Springer PS, Goh L, Buckler ES 4th, Martienssen R. Architecture of floral branch systems in maize and related grasses. Nature. 2005;436(7054):1119–26. https://doi.org/10.1038/nature03892.
74. Landoni M, Cassani E, Pilu R. *Arabidopsis thaliana* plants overexpressing *Ramosa1* maize gene show an increase in organ size due to cell expansion. Sex Plant Reprod. 2007;20:191–8. https://doi.org/10.1007/s00497-007-0056-6.
75. Rodríguez-Rodríguez MF, Salas JJ, Venegas-Calerón M, Garcés R, Martínez-Force E. Molecular cloning and characterization of the genes encoding a microsomal oleate Δ12 desaturase (*CsFAD2*) and linoleate Δ15 desaturase (*CsFAD3*) from *Camelina sativa*. Ind Crop Prod. 2016;89:405 – 15. https://doi.org/10.1016/j.indcrop.2016.05.038
76. Ohta M, Sato A, Renhu N, Yamamoto T, Oka N, Zhu J-K, Tada Y, Suzaki T, Miura K. *MYC*-type transcription factors, *MYC67* and *MYC70*, interact with *ICE1* and negatively regulate cold tolerance in *Arabidopsis*. Sci Rep. 2018;8:11622. https://doi.org/10.1038/s41598-018-29722-x.
77. Spies FP, Perotti MF, Cho Y, Jo CI, Hong JC, Chan RL. A complex tissue-specific interplay between the Arabidopsis transcription factors AtMYB68, AtHB23, and AtPHL1 modulates primary and lateral root development and adaptation to salinity. Plant J. 2023;115(4):952–66. https://doi.org/10.1111/tpj.16273.
78. Irish VF, Sussex IM. Function of the *apetala-1* gene during *Arabidopsis* floral development. Plant Cell. 1990;2(8):741–53. https://doi.org/10.1105/tpc.2.8.741.
79. Chao WS, Wang H, Horvath DP, Anderson JV. Selection of endogenous reference genes for qRT-PCR analysis in *Camelina sativa* and identification of *FLOWERING LOCUS C* allele-specific markers to differentiate summer- and winter-biotypes. Ind Crop Prod. 2019;129:495–502. https://doi.org/10.1016/j.indcrop.2018.12.017.
80. Horstman A, Fukuoka H, Muino JM, Nitsch L, Guo C, Passarinho P, Sanchez-Perez G, Immink R, Angenent G, Boutilier K. AIL and HDG proteins act antagonistically to control cell proliferation. Development. 2015;142(3):454–64. https://doi.org/10.1242/dev.117168.
81. Riechmann JL, Wang M, Meyerowitz EM. DNA-binding properties of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. Nucleic Acids Res. 1996;24(16):3134–41. https://doi.org/10.1093/nar/24.16.3134.
82. Zhao J, Favero DS, Peng H, Neff MM. *Arabidopsis thaliana* AHL family modulates hypocotyl growth redundantly by interacting with each other via the PPC/DUF296 domain. Proc Natl Acad Sci USA. 2013;110(48):E4688–97. https://doi.org/10.1073/pnas.1219277110.
83. Hernandez-Garcia CM, Finer JJ. Identification and validation of promoters and *cis*-acting regulatory elements. Plant Sci. 2014;217–18:109–19 https://doi.org/10.1016/j.plantsci.2013.12.007.
84. Blume RY, Rabokon AM, Pydiura M, Yemets AI, Pirko YV, Blume YB. Genome-wide identification and evolution of the tubulin gene family in *Camelina sativa*. BMC Genomics. 2024;25:599. https://doi.org/10.1186/s12864-024-10503-y.
85. Galasso I, Manca A, Braglia L, Ponzoni E, Breviario D. Genomic fingerprinting of *Camelina* species using cTBP as molecular marker. Am J Plant Sci. 2015;6:1184–200. https://doi.org/10.4236/ajps.2015.68122.
86. Brock J. The evolutionary history of *Camelina* Crantz (Brassicaceae) and domestication of the biofuel crop, *C. sativa* (L.) Crantz. https://openscholarship.wustl.edu/art_sci_etds/2482 (2021). Accessed 06 Aug 2024.
87. Neumann NG, Nazarenus TJ, Aznar-Moreno JA, Rodriguez-Aponte SA, Mejias Veintidos VA, Comai L, Durrett TP, Cahoon EB. Generation of camelina mid-oleic acid seed oil by identification and stacking of fatty acid biosynthetic mutants. Ind Crop Prod. 2021;159:113074. https://doi.org/10.1016/j.indcrop.2020.113074.

Publisher's note