Journal of Agriculture and Food Research 31 (2026) 103294

Contents lists available at ScienceDirect

Journal of Agriculture and Food Research

journal homepage: www.sciencedirect.com/journal/journal-of-agriculture-and-food-research

Genotypic and transcriptomic characterization of a *SAD* gene cluster in European hazelnut (*Corylus avellana* L.)

Andrea Ferrucci [a,*], Leonardo Caproni [b], Svenja Mager [b], Leontina Lipan [c], Agustí Romero-Aroca [c], Mercè Rovira [c], Claudio Todeschini [d], Stuart James Lucas [e], Fuqiang Cui [f,g,**], Matteo Dell'Acqua [b], Valerio Cristofori [a], Cristian Silvestri [a]

[a] Department of Agriculture and Forest Science (DAFNE), University of Tuscia, Via San Camillo De Lellis, s.n.c., Viterbo, 01100, Italy
[b] Institute of Plant Sciences, Scuola Superiore Sant'Anna, Piazza Martiri della Libertà, 33, Pisa, 56127, Italy
[c] Institute of Agrifood Research and Technology (IRTA), Fruit Production, Mas Bové, Ctra. Reus-El Morell, km 3,8, Tarragona, Constantí, 43120, Spain
[d] Hazelnut Company, Ferrero Trading Luxembourg, Route de Trèves, Senningerberg, 2633, Luxembourg
[e] Sabanci University Nanotechnology Research and Application Center (SUNUM), Istanbul, Turkey
[f] State Key Laboratory for Development and Utilization of Forest Food Resources, Zhejiang A&F University, Lin'an, Hangzhou, 311300, China
[g] Zhejiang International Science and Technology Cooperation Base for Plant Germplasm Resources Conservation and Utilization, Zhejiang A&F University, Hangzhou, 311300, China

ARTICLE INFO

Keywords:
Corylus avellana L.
Fatty acids
Genome-wide association study
Transcriptome profiling
Fatty acid desaturase

ABSTRACT

European hazelnut (*Corylus avellana* L.) is a nut crop whose kernels are rich in unsaturated fatty acids, a feature that has so far been only marginally studied from a genetic perspective in this species. In the current work, a preliminary genome-wide association study (GWAS) for fatty acid content in hazelnut kernels was conducted on a diversity panel of 89 individuals, using 40,113 single nucleotide polymorphisms (SNPs). The analysis reported significant marker-trait association for five and six SNPs with arachidic acid (C20:0) and linoleic acid (C18:2$\Delta^{9,12}$), respectively. A strong association with arachidic acid (*p value* = 2.86 x 10^{-11}) was observed for a SNP on chromosome 4 in proximity of a cluster including seven gene models putatively encoding for stearoyl-[acyl-carrier-protein] (ACP) Δ9-desaturase (SAD), the enzyme that introduces a first double bound in C18 fatty acids in plants. As a parallel approach, a genome-wide characterization of *fatty acid desaturases* (*FADs*) was performed identifying 27 members belonging to this gene family in the reference genome of this species. Comparative genomics analysis between European hazelnut and other Fagales highlighted the presence of a *SAD* gene cluster in the other species analysed except Persian walnut (*Juglans regia* L.). Finally, the characterization of *FAD* gene expression profiles in vegetative organs like leaves, petioles, nodes, internodes and roots of both plantlets grown in pot (*in vivo*) and in tissue culture (*in vitro*) was carried out. Differences among the organs in the expression patterns of paralogues of *FAD* genes were observed, with a general conservation of the transcriptional profiles between the two growth systems. Among the *SAD* genes arranged in cluster, *CaSAD.9* (*LOC132177089*) showed high expression levels [$\log_2$(TPM+1) ≥ 6.83] in all the organs of plantlets from both growth systems.

1. Introduction

The European hazelnut (*Corylus avellana* L.) is a shrubby plant species belonging to the family *Betulaceae*, order Fagales. It represents one of the most widely cultivated nut crops in the world, whose fruits are particularly appreciated for their flavour and high nutritional value, with over 90% of production being used by the food industry, particularly the confectionary sector [1]. Hazelnut consumption has been associated with health benefits such as antiatherogenic effects, thereby reducing the risk of cardiovascular diseases, and may also mitigate the risk of various neurodegenerative disorders [2–4]. This beneficial potential of hazelnuts mirrors a rich and balanced chemical composition, which includes significant amounts of minerals, vitamins, fibres, proteins and fats, the latter accounting for 60% of the kernel

* Corresponding author.
** Corresponding author. State Key Laboratory for Development and Utilization of Forest Food Resources, Zhejiang A&F University, Lin'an, Hangzhou, 311300, China.
E-mail addresses: andrea.ferrucci@unitus.it (A. Ferrucci), fuqiang.cui@zafu.edu.cn (F. Cui).

https://doi.org/10.1016/j.jafr.2026.103294
Received 2 July 2026; Received in revised form 9 September 2026; Accepted 10 September 2026
Available online 12 September 2026

weight [5]. Among lipid compounds, monounsaturated fatty acids (MUFAs) represent the richest fraction, with oleic acid accounting for 75-80% of total fatty acids in kernels. Polyunsaturated fatty acids (PUFAs) are mainly represented by linoleic acid, which ranges between 7 and 14%, while saturated fatty acids are present in lower amounts, with palmitic acid and stearic acid accounting for 5-7% and 1.6-3%, respectively. Over the years, several studies have been conducted to characterize the lipid profile of hazelnuts, highlighting a high variability in fatty acid composition between different cultivated varieties and thus suggesting a strong genotypic influence on this trait [5,6]. Since lipids are important components of the nutritional profile of hazelnuts, the study of genes involved in fatty acid metabolic processes is a fundamental aspect to consider when undertaking selection programmes aimed at improving the composition and balance of this class of metabolites.

The era of omics sciences enabled the generation of a large amount of data which can be exploited to analyse the genetic basis that underlies traits of interest at an increasingly affordable price. Genome-wide association studies (GWAS) implementing genotyping-by-sequencing (GBS) data represent a powerful tool to dissect the association of molecular markers with traits of interest [7]. Regarding woody crops, the establishment of field gene banks and the subsequent collection of phenotypic observations encounter several difficulties, mainly due to the larger size of trees and their longer life cycle compared to herbaceous crops [8]. Despite these limitations, GWAS based on GBS data have begun to be applied to the European hazelnut in recent years [9–11]. Genetic research in this species is further hindered due to a pronounced recalcitrance to *in vitro* tissue culture, like other Fagales, representing an obstacle to the application of genetic engineering techniques for both functional genomics studies and breeding [12].

Although fatty acids represent a conspicuous component of hazelnut kernels, the study of the genetic basis regulating the metabolism of this class of compounds in this species has been neglected, unlike the case with other oil-bearing tree crops such as oil palm [13,14], olive [15], macadamia [16] and walnut [17]. As unsaturated fatty acids are the most conspicuous proportion of this class of compounds in hazelnuts, a pivotal role is exerted by fatty acid desaturases (FADs), a family of oxygenase that catalyse the removal of two hydrogen atoms from a fatty acyl chain to introduce a carbon-carbon double bond, using molecular oxygen as a cofactor. These enzymes can be divided into two structurally unrelated groups: soluble and membrane-bound [18,19]. The first group is represented by stearoyl-acyl carrier protein (ACP) Δ9-desaturases (SADs), enzymes located in the stroma of plastids that convert stearoyl-ACP (C18:0) to oleoyl-ACP (C18:1Δ^9) by introducing a double bond between C-9 and C-10 of the acyl chain. Membrane-bound FADs include several members, further grouped into multiple subfamilies characterized by differences in substrate affinity and desaturation patterns [20]. This latter group includes Δ12 (ω-6) and Δ15-desaturase (ω-3), which, together with SADs, are responsible for the synthesis of the major C18 unsaturated fatty acids in plants and have been intensively studied to understand the mechanisms that regulate the balance between MUFAs and PUFAs to promote the accumulation of specific fatty acids [21]. In the present study, the association between single-nucleotide polymorphisms (SNPs) and the concentration of 10 fatty acids in kernels of European hazelnuts was investigated through a GWAS conducted on a diversity panel composed of 89 individuals for which both genotypic and phenotypic data were available. Genome-wide identification analysis of *FAD* genes was also carried out, together with comparative analysis of several *SAD* genes in species belonging to the Fagales order. At the same time, the expression profile of *FAD* genes in vegetative organs was analysed through transcriptome sequencing of leaves, petioles, nodes, internodes and roots of plantlets of the Italian cultivar 'Tonda Gentile Romana' grown both *in vivo* and *in vitro*.

2. Materials and methods

2.1. Plant material, genotyping and variant filtering

Plant material was grown in an experimental orchard located at the Institute of Agrifood Research and Technology (IRTA) Mas Bovè (Constantí, Tarragona, Northeastern Spain), where each genotype is grown in duplicate. A total of 158 genotypes, including 156 cultivars of the European hazelnut (*Corylus avellana* L.) representative of different part of the world, a wild-type individual of the ornamental *Corylus avellana* var. *contorta* and a wild-type individual of the Turkish hazelnut (*Corylus colurna* L.) were initially included in the analysis (Table S1). Genotypic data for these accessions had previously been generated as part of an independent study conducted by Salardi-Jost et al. [11], who used this collection to study genetic diversity, identify genomic loci under selection, and describe genome-wide association with nut traits like size, perimeter and calibre. This data consisted in paired-end Illumina reads obtained from DNA extracted from leaves that was genotyped-by-sequencing using a double-digestion restriction site associated DNA (ddRAD) method, with a protocol based on the *Nsp*I and *Mbo*I endonucleases. Demultiplexed reads were aligned to 'Tombul' reference genome [22] with bwa-mem [23], only alignments with MAPQ $\geq$ 5 were retained, and the resulting bam files were sorted and indexed with Samtools [24] before they were implemented in stacks [25] for SNP calling with gstacks. A vcf file was then obtained by running populations and SNPs were primarily filtered for a minimum percentage of individuals across populations required to process a locus (–min-samples-overall = 0.75), specifying a maximum observed heterozygosity required to process a nucleotide site at a locus (–max-obs-het = 0.8). Mean depth was measured using VCFtools (–site-mean-depth) and variants with depth lower than 10 and higher than 62 were filtered out (–min-meanDP 10 and –max-meanDP 62). In Tassel [26], further filtering was applied by removing samples with more than 20% missing variants and sites present in less than 90% of samples were removed with VCFtools [27]. Filtering based on the minimum proportion of site present resulted in 146 individuals being retained for further analysis; subsequently, sites with minor allele frequency (MAF) $<$ 0.05 were removed in Tassel. This dataset was converted to hapmap format and used as genotypic data input for further analysis.

2.2. Phenotyping

Fully ripe hazelnuts from 89 genotypes included in the diversity panel were harvested in 2017 from the same experimental orchard in which plant material was collected for genotypic characterization. The fatty acid profile of kernels was analysed according to Rovira et al. [28]. Briefly, fatty acid methyl esters (FAMEs) were prepared by transesterification with 0.5 M KOH, following the official method UNE-EN ISO 5509:2000. The FAME (1 mL) were separated using a gas chromatograph (HP 6890; Agilent Technologies, Barcelona, Spain) equipped with a flame ionisation detector (FID) and a 30 m long capillary column with a diameter of 0.25 mm (HP-Innowax, Agilent Technologies). Helium was employed as carrier gas, and the flow rate was set to 1 mL/min. The injector and detector temperatures were +220 and + 275 °C, respectively. The identification of FAME was based on retention time compared to that of a standard FAME mixture (Sigma-Aldrich, Madrid, Spain). For each genotype, the samples collected from two biological replicates were analysed separately and two technical replicates were performed on each of them. The values were expressed as relative fatty acid concentration (%), mediated and used for further analysis. Spearman's correlation among traits was calculated using the PerformanceAnalytics R's package [29]. Distribution of phenotypic data was visually inspected and those traits showing a strong deviation from normality (Shapiro-Wilk test p *value* $<$ 0.000001) were transformed in R using the BestNormalize package [30]. Methods chosen by the program

for transformation were orderNorm and Standardized Log_b(x + a) for margaric acid (C17:0) and eicosenoic acid (C20:1Δ^9), respectively. These data were stored in a text file and used as phenotypic data for further analysis.

2.3. Genome-wide association study and linkage-disequilibrium decay analysis

Both genotypic and phenotypic data were implemented in GAPIT v. 3.4.0 [31] using Fixed and random model Circulating Probability Unification (FarmCPU; [32]) model with 3 principal components to account for population structure. Linkage-disequilibrium (LD) was calculated for each chromosome using Plink [33] on all the SNPs retained after thinning to 25% (–ld-window 99999 –ld-window-kb 100000 -ld-window-r2 0 –thin 0.25 –seed 12345). The LD distances were categorized into 1 kb bins and the average r^2 values per bin were calculated with a precast python script (https://github.com/speciationgenomics/scripts/blob/master/ld_decay_calc.py). Theoretical LD values were estimated by fitting the binned LD data in a non-linear least squares model (nls) in R using the Hill and Weir formula [34] in a script adapted from Marroni et al. [35]. The LD decay was set as the distance corresponding to half of the maximum theoretical LD value and this distance was then scanned upstream and downstream for each SNP significantly associated with traits under study and the IDs of genes present within these intervals were retrieved. The Gene Ontology (GO) annotation of genes in the 'Tombul' reference genome was obtained from the GO Association File (gaf) available on NCBI (accessed on 14th February 2026) and an enrichment analysis of GO categories associated with genes in LD with significantly associated SNPs was performed in R using topGO [36].

2.4. Analysis of gene and protein structures

The gaf file available on NCBI for European hazelnut 'Tombul' reference genome (CavTom2PMs-1.0; [22]) was scanned for genes annotated with InterPro domains related to fatty acid desaturation activity (IPR005804, Fatty acid desaturase domain; IPR012171, Fatty acid desaturase; IPR021863, Fatty acid desaturase, N-terminal; IPR005067, Fatty acid desaturase, type 2; IPR052864, Chloroplast Fatty Acid Desaturase CarF; IPR053335, Fatty Acid Desaturase CarF-like; IPR001522, Fatty acid desaturase type 1, conserved site; IPR005803, Fatty acid desaturase type 2, conserved site), and the protein sequences of these gene models were retrieved. The members of FAD4 subfamily were identified by blasting the protein sequence of AtFAD4.1 (AT1G62190) to the reference genome of European hazelnut. The analysis of conserved motifs was done on the European hazelnut FAD protein sequences using MEME suite (https://meme-suite.org/meme/; accessed on 27th February 2026) with the following parameters: number of repetitions: any; maximum number of motifs: 15; optimum motif widths: between 4 and 300 residues. Pfam functional domains were analysed in CDD hit, and data were visualized in TBTools-II [37]. Subcellular localization prediction was done with DeepLoc 2.0, and only predictions with score $\geq$ 0.85 were considered [38]. The research of histidine-box domains was conducted by searching the protein sequence of FADs for canonical progressive sequences $HX_{3-4}H$, $HX_{2-3}H(H)$ and HX_2HH, where "X" represents any amino acid.

2.5. Phylogenetics and comparative genomics analysis

Fatty acid desaturase protein sequences of European hazelnut were aligned with Muscle [39] and a maximum-likelihood phylogenetic tree was built with RAxML-NG [40] using the Jones-Taylor-Thornton (JTT) substitution model with 1000 bootstrap replicates. The resulting tree was visualized in iTOL [41]. The same workflow described in the previous paragraph was used to identify *FAD* genes in the reference genomes of Persian walnut cv. 'Chandler' (*Juglans regia* L.; Walnut_2.0; [42]), valley oak (*Quercus lobata* née; ValleyOak3.2; [43]) and *Arabidopsis thaliana* (L.) Heynh (TAIR10; [44]), while only genes annotated as *SAD* were retrieved in the reference genomes of European chestnut cv. 'Marrone di Chiusa Pesio' (*Castanea sativa* Mill.; ASM4071231v1; [45]), common alder (*Alnus glutinosa* (L.) Gaertn.; dhAlnGlut1.1; [46]) and pecan cv. 'Pawnee' (*C. illinoinensis*; C.illinoinensisPawnee_v1; [47]). The same workflow reported above was employed for sequence alignment, phylogenetic tree construction and visualization. Intragenomic analysis of synteny was done aligning the European hazelnut proteome against itself with blastp (-evalue $1e^{-10}$; -max_target_seqs 6). Multiple isoforms were first removed with agat_sp_keep_longest_isoform.pl and only the longest was retained [48]. The resulting alignment file was then inputted into MCScanX with default parameters [49] for building collinear blocks. Multiple species synteny analysis was performed aligning protein sequences of European hazelnut against those of other Fagales. Blastp (-evalue $1e^{-10}$; -max_target_seqs 5) was used to compare proteome files and the resulting alignments were then inputted into MCScanX [49] with -s parameter set to 20 to increase the minimum number of genes to call a collinear block. The IDs of coding sequences and pseudogenes present within the *SAD* gene cluster of the species analysed were retrieved from NCBI and the gene cluster was visualized with the R package geneviewer [50]. SAD protein sequences of Fagales species and *A. thaliana* were also aligned, and a phylogenetic tree was constructed using the parameters described above.

2.6. Transcriptome sequencing and expression profiling

Transcriptome data were obtained from both plantlets of cultivar 'Tonda Gentile Romana' grown in pot (*in vivo*) and derived from tissue culture (*in vitro*). The RNA was extracted from leaves, petioles, nodes, internodes and roots using a CTAB-based method according to Lucas et al. [22], reverse transcription and library preparation were carried out with SQK-PCB114.24 kit (Oxford Nanopore Technologies) according to the manufacturer's protocol. Then, 25 ng of barcoded cDNA from each sample were pooled together before ligation of sequencing adapters, and 150 fmol of this solution were loaded into a R10.4.1 flow cell on a MinION device and sequenced for 72 h. The RNA isolated from the organs of two *in vivo*-grown plantlets and from the plantlets contained in two 250 mL tissue culture jars (10 plantlets per jar) was pooled and treated as a single replicate for the respective growth system. The raw reads were demultiplexed and base called with Dorado 0.9.6, using the 'super' accuracy DNA basecalling model v5.0. The adapters were trimmed with Porechop_ABI [51] and the reads were oriented using restrander [52]. Reads with length $\geq$ 50 and with quality $\geq$ 7 were filtered with Nanofilt [53] for further analysis. The resulting fastq files were aligned to the *Corylus avellana* L. reference transcriptome [22] using minimap2 [54] in map-ont mode with -p 0 and -N 10 parameters. Gene expression was quantified as Transcripts Per Million (TPM) using NanoCount [55] with -n and -d −1 flags. The expression levels of *FAD* genes in different tissues were reported as mean $\log_2$(TPM+1) of three replicates. An explorative heatmap was constructed with the R package pheatmap [56]. Differences in the expression levels of *FAD* genes among the different organs were analysed with the R package DESeq2 [57]. A matrix of rounded raw_counts obtained with NanoCount for *in vivo* and *in vitro*-grown samples was filtered, removing genes with raw_counts < 1 in more than two replicates in all the organs analysed, and a likelihood ratio test (LRT) was done on these data. Subsequently, pairwise comparisons between organs were performed using the Wald test implemented in DESeq2 [57]. $\log_2$FC were shrunken using the lfcShrink function with the ashr method to obtain more accurate effect size estimates [58]. A descriptive principal component analysis (PCA) of the expression profiles among the organs of *FAD* genes expressed in both plantlets grew *in vivo* and *in vitro* was done with prcomp function in the stats package in R [59].

3. Results

3.1. Genome-wide association study for fatty acid content in hazelnut kernels

After filtering, 40,113 high-quality SNPs from 146 genotypes were used for LD analysis and estimation of LD decay. Among the 10 traits investigated only palmitic acid, vaccenic acid and α-linolenic acid showed a normal distribution according to the Shapiro-Wilk test (*p-value* > 0.05; Table S1). Most of the traits analysed showed a strong significant correlation with at least one other trait, according to Spearman's rank correlation coefficient (*p-value* < 0.001), with oleic acid and linoleic acid showing the most marked negative correlation (Spearman's $\rho = -0.96$; *p-value* < 0.001), while the strongest positive correlation was reported for palmitoleic acid and vaccenic acid (Spearman's $\rho = 0.80$; *p-value* < 0.001) (Fig. S1). The GWAS analysis was conducted on a subset of 89 genotypes for which both genotypic and phenotypic data were available, while the number of SNPs tested for association remained unchanged from that used for the LD analysis. The FarmCPU model used for GWAS analysis identified six SNPs associated with linoleic acid on chromosomes 1, 2, 4, 5, 8, and 11, while five SNPs were found to be associated with arachidic acid, one SNP on chromosomes 1, 2, and 7, and two SNPs on chromosome 4 (Fig. 1A; Table 1; Fig. S2). The visual inspection of quantile-quantile plots (Q-Q plots) did not show any substantial deviation from the expected distribution (Fig. 1A). Linkage disequilibrium decay values ranged from a minimum of 128,606 bp in chromosome 11 to a maximum of 416,335 bp in chromosome 3 (Table 2). The least number of genes within the LD decay-delimited range were found surrounding the SNP on chromosome 11 associated with linoleic acid (13 genes), while the highest number of genes in LD was reported for the SNP on chromosome 1 associated with arachidic acid (77 genes; Table 1; Table S2). The percentage of phenotypic variance explained (PVE) resulted to be generally low for SNPs associated with arachidic acid, being lower than 4%, with the only exception represented by the SNP on chromosome 7, which accounted for 33.25% (Table 1). A similar trend was observed for SNPs associated with linoleic acid which showed PVE lower than 2% in all the associated markers except for the SNPs on chromosome 2 and 11, reaching 11.34% and 11.74%, respectively (Table 1). The GO enrichment analysis of genes in LD with SNPs associated with the traits investigated indicated an over-representation of categories related to fatty acid metabolism, as genes annotated with “fatty acid metabolic process” (GO:0006631; *p-value* $= 3.20 \times 10^{-6}$) and “stearoyl-[ACP] desaturase activity” (GO:0045300; *p-value* $= 6.60 \times 10^{-7}$) GO categories were significantly enriched (Fig. 1B). At a genome-wide level, these two categories were represented by a total of 150 and 11 genes, respectively. Among the genes in LD with markers significantly associated with the traits studied, 11 were annotated with the former category, and five of these were also annotated with the latter. These were *LOC132171848* (*CaSAD.2*), which was in LD with the SNP on chromosome 2 associated with linoleic acid, *LOC132177385* (*CaSAD.3*), *LOC132179807* (*CaSAD.4*), *LOC132178572* (*CaSAD.5*) and *LOC132179592* (*CaSAD.6*), which were in LD with the SNP at position 22,707,256 of chromosome 4, which was associated with arachidic acid. Other gene models annotated with "fatty acid metabolic process" category (GO:0006631) were *LOC132162958* and *LOC132165299*, putatively encoding a peroxisomal long chain acyl-CoA synthetase and a peroxisomal adenine nucleotide carrier protein, respectively, which were both in LD with the SNP on chromosome 1 associated with linoleic acid. Two gene models, *LOC132170256* and *LOC132170424*, putatively encoding a mitochondrial acyl carrier protein and a chloroplastic biotin carboxyl carrier protein of acetyl-CoA carboxylase, respectively, were in LD with the SNP on chromosome 2 associated with linoleic acid. *LOC132183153*, which putatively encodes a 3-ketoacyl-CoA synthase, and *LOC132187310*, that putatively encodes a CoA ligase CCL8, were in LD with the SNP on chromosome 5 associated with linoleic acid, and the SNP on chromosome 7 associated with arachidic acid, respectively.

3.2. Genome-wide identification, structural and phylogenetic analysis of FADs

Twenty-three genes were annotated with InterPro categories related to fatty acid desaturation and four genes were identified as *fatty acid desaturase 4* (*FAD4*) in the European hazelnut cv. ‘Tombul’ reference genome ([22]; Table 3; Table S3). These genes were distributed among all the 11 chromosomes, except for chromosome 5, 8 and 11 (Fig. 2). The most prevalent group among this gene family was the *stearoyl-acyl carrier protein (ACP) Δ9-desaturase* (*SAD*), which accounted for 11 members (*CaSAD.1-CaSAD.11*). Among these genes encoding soluble plastidial desaturases, *CaSAD.3*, *CaSAD.4* and *CaSAD.5*, and *CaSAD.7*, *CaSAD.8* and *CaSAD.9* were arranged as tandem duplications on chromosome 4, forming a cluster expanding for about 220 kb that also included *CaSAD.6* (Fig. 2). Five members were identified in *Δ12 FAD* subfamily, one of these predicted to encode a chloroplastic enzyme, *CaFAD6*, while four members of this subfamily encoded microsomal enzymes (*CaFAD2.1--CaFAD2.4*). Among this latter group, intragenomic synteny analysis indicated *FAD2.1* on chromosome 9 and *FAD2.2* on chromosome 10 to be segmental duplicates, while *CaFAD2.2* and *CaFAD2.3* are tandem duplications. Multiple members were also observed within the *Δ15 FAD* subfamily, which included genes putatively encoding a microsomal enzyme, *CaFAD3*, and two plastidial enzymes, *CaFAD7/8.1* and *CaFAD7/8.2*. All the members of this subfamily were determined to be the product of segmental duplication events (Fig. 2). *FAD4* genes were present in four copies, *CaFAD4.1* and *CaFAD4.2* on chromosome 3, and *CaFAD4.3* and *CaFAD4.4* are tandem duplicates on chromosome 10. Two copies of *sphingolipid Δ8-desaturase*, *CaSLD1* and *CaSLD2*, were identified, while *palmitoyl-monogalactosyldiacylglycerol Δ7-desaturase* (*CaADS*) and *sphingolipid Δ4-desaturase* (*CaDES*) were both present in single copy.

The maximum-likelihood phylogenetic tree of European hazelnut FAD protein sequences clearly separated plastidial soluble from membrane-bound enzymes, with members of the same subfamily clustering together (Fig. 3A). Structural analysis of FAD protein sequences was set to identify up to 15 shared motifs, highlighting a strong conservation in sequence homology among members of the same subfamily (Fig. 3B; Table S4). All SAD members shared the presence of motifs 1 and 2, while motifs 6 and 7 were present in all the members of this subfamily except CaSAD.7, which putatively represents an N-terminal truncated copy. All the members of Δ15 FAD showed the presence of motifs 8, 9 and 10, and strong sequence homology was also observed among microsomal Δ12 FADs, with all the members sharing motifs 8, 9 and 11. Motif 8 was also found in plastidial Δ12 desaturase CaFAD6 and in both palmitoyl-monogalactosyldiacylglycerol Δ7-desaturase (CaADS) and sphingolipid Δ4-desaturase (CaDES). Both copies of sphingolipid Δ8-desaturase, CaSLD.1 and CaSLD.2, shared the presence of motifs 12 and 13. All four FAD4 orthologues had motif 14, while motif 15 was only found in CaFAD4.1 and CaFAD4.2. Analysis in CDD reported the presence of functional Pfam superfamily domains related to fatty acid desaturase activity in all the proteins analysed (Fig. 3C). The analysis of intron/exon structure revealed differences among members of *FAD* gene family and within members of the same subfamily (Fig. 3D). All *SAD* genes had three or four exons, with the only exception represented by *CaSAD.*10 that had only two exons. Except for *CaFAD2.2*, which had a single exon, the other three microsomal Δ12 desaturase encoding genes had a single intron spanning the 5’ UTR region. The plastidial Δ12 desaturase encoding gene *CaFAD6* had ten exons, while all genes encoding Δ15 FADs had eight exons. The two genes encoding *sphingolipid Δ8-desaturase*, *CaSLD.1* and *CaSLD.2*, had one and two exons, respectively, while the two single copy genes *CaADS* and *CaDES* had five and two exons, respectively. All *FAD4* encoding genes were single exon.

A distinctive structural domain of membrane-bound FADs is represented by histidine-boxes, which are generally found in the number of

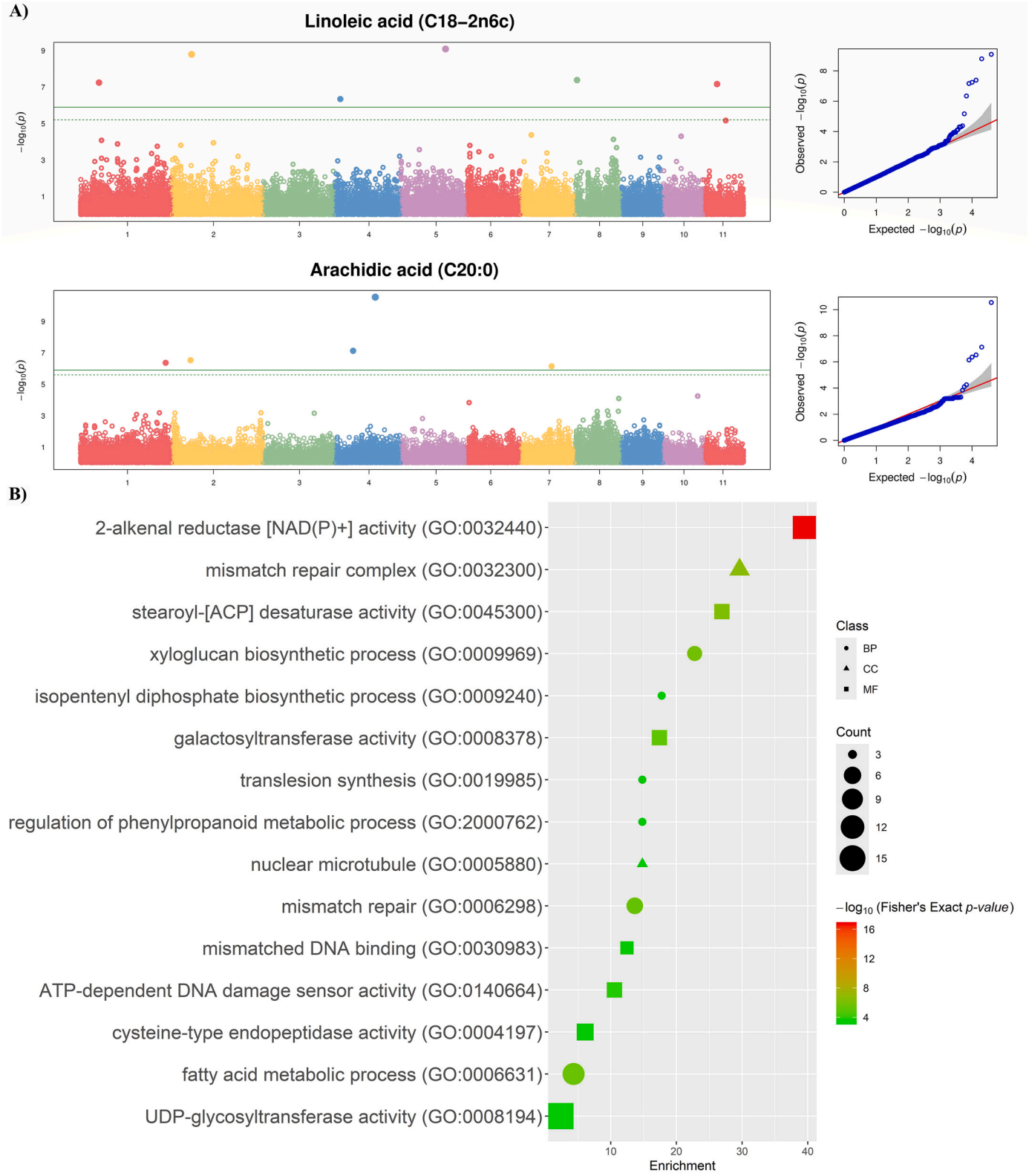

Fig. 1. Genome-wide association study for fatty acid content in kernels of European hazelnut and enrichment analysis of Gene Ontology (GO) categories. Manhattan plots displaying SNPs associated with linoleic acid (C18:1$\Delta^{9,12}$) and arachidic acid (C20:0) content in kernels of European hazelnut and quantile-quantile (Q-Q) plots (**A**). In Manhattan plots, SNPs are represented as coloured dot at their genomic position on the chromosomes (X-axis); the solid green horizontal lines represent the significance threshold based on the Bonferroni correction method, which was applied at a *p-value* of 0.05, whereas the dashed horizontal green line indicates a less stringent threshold based on false discovery rate (FDR) correction. In Q-Q plots, the expected (X-axis) versus the observed (Y-axis) $-\log_{10}$(*p-value*) is reported. Only results for traits with SNPs significantly associated are reported. GO enrichment of the genes in LD decay with SNPs associated with fatty acid concentration in kernels of European hazelnut (**B**). The level of enrichment is reported in the X-axis. The class and the count of the enriched GO category are also reported, together with the significance level of the enrichment, expressed as the -$\log_{10}$(*p-value*). Only categories with a *p-value* < 0.001 were reported. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1
List of SNPs associated with fatty acid composition in kernels of European hazelnut. Phenotypic trait associated, chromosome, physical position (bp), genomic region, significance level (*p-value*), phenotypic variance explained (PVE, %), allelic effect (the sign is in respect to the minor allele), the LD decay interval used to search for candidate genes and the number of candidate genes identified.

Trait	Chr	Position (bp)	Region	*p-value*	PVE (%)	Effect	LD window Start (bp)	LD window End (bp)	Genes
Linoleic acid (C18:2n6c)	1	10,390,970	Intergenic	5.64×10^{-8}	1.35×10^{-8}	1.5063	10,019,652	10,762,288	71
Linoleic acid (C18:2n6c)	2	10,986,120	Intergenic	1.59×10^{-9}	11.34	1.4010	10,664,912	11,307,328	47
Linoleic acid (C18:2n6c)	4	3,236,885	Exonic (*LOC132177264*)	4.50×10^{-7}	0	1.5247	2,898,725	3,575,046	77
Linoleic acid (C18:2n6c)	5	25,059,155	Intronic (*LOC132181991*)	8.09×10^{-10}	1.70	2.3490	24,833,883	25,284,427	40
Linoleic acid (C18:2n6c)	8	1,253,627	Intergenic	4.09×10^{-8}	1.29×10^{-8}	1.1626	1,042,280	1,464,974	42
Linoleic acid (C18:2n6c)	11	7,575,975	Intronic (*LOC132164987*)	6.78×10^{-8}	11.74	2.4501	7,447,369	7,704,581	13
Arachidic acid (C20:0)	1	47,498,530	Intergenic	4.26×10^{-7}	0	−0.0116	47,127,212	47,869,848	67
Arachidic acid (C20:0)	2	10,419,298	Intronic (*LOC132169035*)	2.93×10^{-7}	3.94	−0.0079	10,098,090	10,740,506	56
Arachidic acid (C20:0)	4	10,338,764	Intergenic	7.37×10^{-8}	1.63×10^{-8}	0.0095	10,000,604	10,676,925	28
Arachidic acid (C20:0)	4	22,707,256	Intergenic	2.86×10^{-11}	0	−0.0126	22,369,096	23,045,417	58
Arachidic acid (C20:0)	7	17,269,641	Intergenic	7.11×10^{-7}	33.25	0.0097	17,112,056	17,427,226	17

Table 2
Estimated LD decay distance for each chromosome of the European hazelnut 'Tombul' reference genome. It was assumed that LD decayed at the physical distance (Mb) at which its maximum estimated value (r^2) was halved. This distance was used to search for candidate genes upstream and downstream SNPs significantly associated with the traits analysed.

Chromosome	LD decay (bp)
1	371,318
2	321,208
3	416,335
4	338,161
5	225,272
6	192,737
7	157,585
8	211,347
9	179,720
10	277,167
11	128,606

three. Analysis of these domains revealed a high sequence conservation among members of the same subfamily (Table 4). Only two domains were present in sphingolipid Δ8-desaturases CaSLD.1 and CaSLD.2. Among FAD4 sequences, three domains were found in both CaFAD4.1 and CaFAD4.4, while only two domains were found in CaFAD4.2 and CaFAD4.3. Three domains were present in both CaADS and CaDES, and in all the members belonging to both Δ15 and Δ12 FAD subfamilies. Interestingly, a single histidine-box was also found in CaSAD.2 and CaSAD.10, while CaSAD.9 had two of these domains, with the most proximal being in common with that present in CaSAD.2.

Phylogenetic analysis was carried out comparing FAD protein sequences of European hazelnut with those from valley oak and Persian walnut, two other species belonging to the Fagales order, and *Arabidopsis thaliana*. Distinct functional subfamilies were separated with a very high level of bootstrapping support (≥97%), except for Δ15 desaturases whose separation from the other subfamilies was characterized by a lower confidence level (81%; Fig. 4).

3.3. Synteny analysis of the SAD gene cluster in Fagales and phylogenetic analysis

The analysis of distribution of genes encoding stearoyl-[acyl-carrier-protein] Δ9-desaturase among members of Fagales revealed differences related to the presence and organization of a *SAD* gene cluster in these species (Fig. 5). This cluster spanned 221,674 bp, from position 22,873,162 to 23,094,836 of chromosome 4 of European hazelnut, including 7 genes encoding stearoyl-[acyl-carrier-protein] Δ9-desaturase: *CaSAD.3-CaSAD.9*. Two pseudogenes annotated as *SAD* were also present in this region (*LOC132177386* and *LOC132178238*). The *SAD* gene cluster spanned 266,559 bp, from position 4,414,643 to 4,681,202 of chromosome 4 in European chestnut, including 6 *SAD* genes, namely *CsSAD.2-CsSAD.7* (Fig. 6). This cluster also included five *SAD* pseudogenes (*LOC142632122*, *LOC142632236*, *LOC142632337*, *LOC142632338* and *LOC142632340*). Valley oak also showed the presence of a 203,693 bp cluster extending from position 56,057,584 to 56,261,277 of chromosome 10, which included 6 *SAD* genes: *QlSAD.5-QlSAD.10*. In this region were also present four pseudogenes annotated as *SAD* (*LOC115962798*, *LOC115962799*, *LOC115965027* and *LOC115965031*). *SAD* genes organized in close contiguity were also observed on chromosome 1 of common alder, which, together with European hazelnut, is also a member of family *Betulaceae*. This region spanned 216,411 bp, from position 17,669,785 to 17,886,196 and included six genes encoding stearoyl-[acyl-carrier-protein] Δ9-desaturase: *AgSAD.1-AgSAD.6*. In pecan, two close *SAD* genes were observed on chromosome 1, *CiSAD.1* and *CiSAD.2*, which were about 250 kb from a third gene, *CiSAD.3*. This locus spanned 265,352 bp from position 46,971,140 to 47,236,492 and included two pseudogenes annotated as *SAD* (*LOC122278257* and *LOC122300577*), in close contiguity with *CiSAD.1* and *CiSAD.2*. Differently from pecan, no arrays of three or more *SAD* genes were observed in the orthologous region of Persian walnut, where only two *SAD* genes were identified.

Analysis of synteny between European hazelnut and other Fagales showed the presence of collinear blocks in the regions bordering the *SAD* gene cluster in these species (Fig. 6; Table S5). A syntenic block of 43 coding sequences linked the region upstream of the *SAD* gene cluster of European hazelnut with the sub-telomeric region of chromosome 4 in European chestnut, localized about 1.45 Mb upstream of the *SAD* gene cluster of this species. A similar scenario was observed analysing collinearity between European hazelnut and valley oak, in which a collinear block of 28 genes linked a region distanced 448 kb upstream from the *SAD* gene cluster of European hazelnut to a region localized 7.7 Mb downstream from the gene cluster of valley oak. A collinear block of 1224 genes linked chromosome 4 of European hazelnut to chromosome 1 of common alder, encompassing the *SAD* gene cluster in both species, although no direct collinear relationships were found in any of the *SAD* genes within these regions. Chromosome 1 of pecan shared a collinear block of 212 coding sequences with the region surrounding the *SAD* gene cluster of European hazelnut. In this context, *CiSAD.3* (XP_042944595.1) was collinear with *CaSAD.3* (XP_059445654.1), which was also collinear with *CiSAD.4* (XP_042967603.1) on chromosome 2, within a block of 162 coding sequences. A similar scenario was observed also when comparing European hazelnut to Persian walnut. A syntenic block of 209 coding sequences linked the area surrounding the *SAD* gene cluster of European hazelnut to a region on chromosome 7 of Persian walnut encompassing *JrSAD.4* (XP_018812159.1), which was collinear to *CaSAD.3* (XP_059445654.1). This gene was also collinear with *JrSAD.5*

Table 3
IDs of *FAD* genes and predicted features of their protein products. ER = endoplasmic reticulum; Pl = plastid.

Group	Gene name	Gene ID	Protein ID	Protein length (aa)	Isoelectric point	Molecular weight (kDa)	Subcellular localization
ADS	*CaADS*	*LOC132187499*	XP_059457809.1	397	9.91	45.21	Pl
DES	*CaDES*	*LOC132170731*	XP_059437799.1	324	7.55	37.63	ER
FAD4	*CaFAD4.1*	*LOC132174553*	XP_059442173.1	292	7.96	32.49	Pl
	CaFAD4.2	*LOC132176451*	XP_059444644.1	300	7.96	33.34	Pl
	CaFAD4.3	*LOC132163744*	XP_059430097.1	263	7.63	28.91	-
	CaFAD4.4	*LOC132163745*	XP_059430098.1	263	7.92	29.05	-
SLD	*CaSLD.1*	*LOC132184796*	XP_059454540.1	447	8.84	51.47	ER
	CaSLD.2	*LOC132187564*	XP_059457896.1	447	7.66	51.02	ER
Δ-12	*CaFAD2.1*	*LOC132191453*	XP_059462453.1	383	8.43	43.67	ER
	CaFAD2.2	*LOC132164733*	XP_059431272.1	382	8.72	44.09	ER
	CaFAD2.3	*LOC132164706*	XP_059431241.1	382	8.72	44.09	ER
	CaFAD2.4	*LOC132164512*	XP_059431015.1	382	9.07	44.03	ER
	CaFAD6	*LOC132183689*	XP_059453049.1	440	9.8	51.26	Pl
Δ-15	*CaFAD3*	*LOC132163977*	XP_059430352.1	386	7.91	44.39	ER
	CaFAD7/8.1	*LOC132186916*	XP_059457017.1	453	8.6	51.81	Pl
	CaFAD7/8.2	*LOC132161849*	XP_059428035.1	456	9.09	52.15	Pl
SAD	*CaSAD.1-1*	*LOC132187800*	XP_059458225.1	390	6.33	44.08	Pl
	CaSAD.1-2		XP_059458233.1	320	4.95	36.41	-
	CaSAD.2	*LOC132171848*	XP_059439243.1	394	6.96	44.8	Pl
	CaSAD.3	*LOC132177385*	XP_059445654.1	361	6.06	40.81	Pl
	CaSAD.4-1	*LOC132179807*	XP_059448571.1	386	6.45	44.01	Pl
	CaSAD.4-2		XP_059448572.1	386	6.45	44.01	Pl
	CaSAD.5	*LOC132178572*	XP_059446998.1	386	7.15	43.94	Pl
	CaSAD.6	*LOC132179592*	XP_059448314.1	390	6.96	43.92	Pl
	CaSAD.7	*LOC132178242*	XP_059446673.1	201	8.75	23.02	-
	CaSAD.8	*LOC132179896*	XP_059448672.1	386	7.17	44.18	Pl
	CaSAD.9	*LOC132177089*	XP_059445288.1	393	6.85	45.03	Pl
	CaSAD.10	*LOC132184406*	XP_059454016.1	394	6.36	44.6	Pl
	CaSAD.11	*LOC132187390*	XP_059457671.1	386	6.92	43.94	Pl

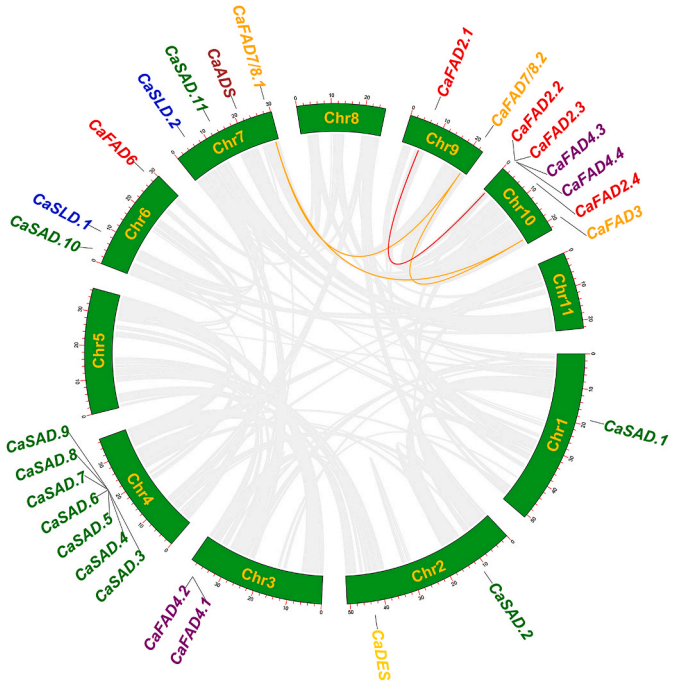

Fig. 2. Circular plot of European hazelnut 'Tombul' reference genome which indicates the genomic position of *FAD* genes along the chromosomes. At the centre of the plot the intragenomic analysis of synteny is illustrated. Chromosomes are depicted as green blocks and their size is indicated in Mb. *FAD* genes are reported in different colours according to their subfamily. Collinear relationships among *FAD* genes are highlighted. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

(XP_018839373.1) on chromosome 12, within a block containing 161 coding sequences.

On the contrary, the sequences encoded by these genes in pecan and Persian walnut shared a higher sequence homology with CaSAD.9, as highlighted by phylogenetic analysis of SAD protein sequences, and formed a cluster that also included CsSAD.2, QlSAD.10, AgSAD.6, all of which are encoded by genes located at the border of their respective *SAD* gene cluster. Other proteins encoded by gene models included in the *SAD* gene clusters of Fagales species formed two clusters, one of which containing sequences from *Fagaceae* species and the other including sequences from both *Betulaceae* and *Juglandaceae* (Fig. 7).

3.4. Expression profile and differential expression analysis of FAD genes in in vivo and in vitro-grown plantlets

A total of 20 *FAD* genes were found to be expressed in organs of plantlets grown *in vivo*, while only 16 of these genes were also identified as expressed in *in vitro*-grown plantlets (Fig. 8; Fig. S3). The expression profile of *FAD* genes resulted in two distinct macro-clusters in both growth systems studied, one of which included genes with expression levels > 4.40 $\log_2$(TPM+1) in all the organs of plantlets grown *in vivo*, and genes with expression levels > 2.50 $\log_2$(TPM+1) in all the organs of plantlets derived from *in vitro* tissue culture (Fig. 8). When comparing the expression profile of *FAD* genes in organs of plantlets grown *in vivo,* the cluster of highly expressed genes included *CaFAD2.1*, *CaSAD.10*, *CaSLD.1*, *CaSLD.2*, *CaFAD6* and *CaFAD7/8.2* (Fig. 8A). Except for this latter, that showed expression levels ranging from 3.76 in internodes to 6.55 $\log_2$(TPM+1) in leaves, these genes formed a distinct cluster also when comparing the expression profile of *FAD* genes in organs of *in vitro*-derived plantlets, showing expression levels > 5.23 $\log_2$(TPM+1) in all the organs analysed (Fig. 8B). Among these genes, differences in the expression levels among organs in both growth systems were highlighted by LRT and Wald test for *CaFAD6* and *CaFAD7/8.2*, while differences in organs of *in vitro*-grown plantlets were also observed for

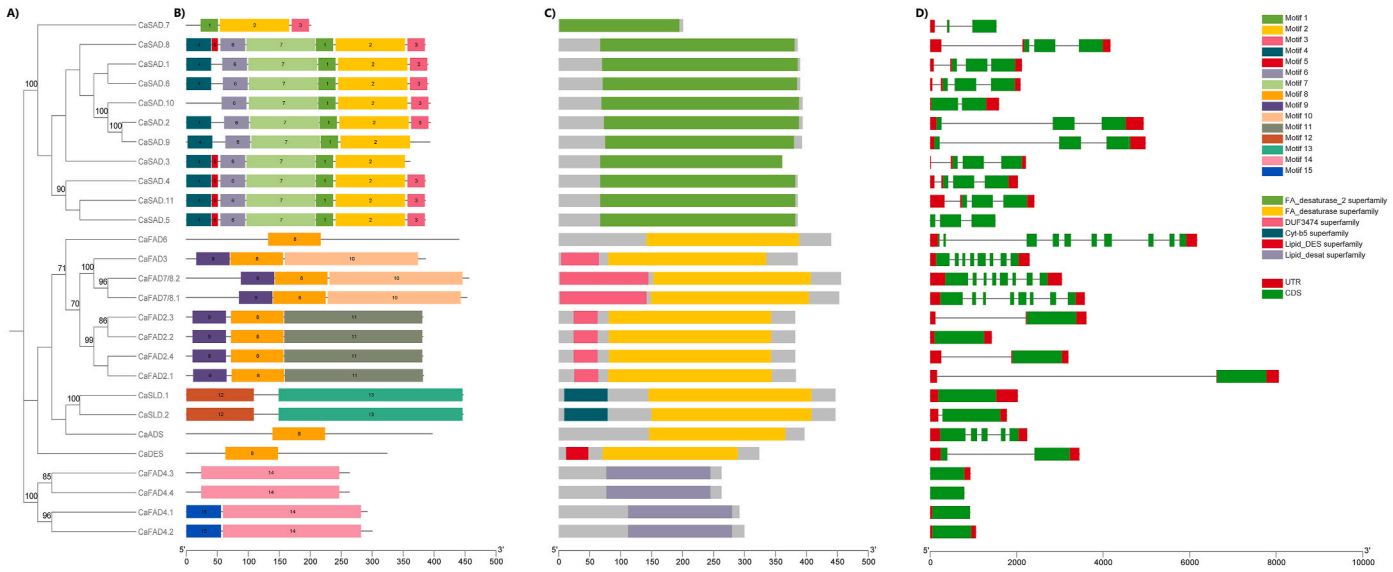

Fig. 3. Phylogenesis and structural analysis of FAD sequences in European hazelnut. Phylogenetic tree built employing the maximum-likelihood method and Jones-Taylor-Thornton (JTT) substitution model, with 1000 bootstraps (only bootstrap values ≥ 70% are reported; **A**), conserved motifs (**B**), Pfam domains (**C**) and intron-exon analysis (**D**) of FAD sequences.

Table 4
Conserved histidine-box domains of FAD proteins and their position.

Type	Protein	Histidine-box	Position	Histidine-box	Position	Histidine-box	Position
ADS	*CaADS*	HRNLSH	170-175	HRYHH	207-211	HNNHH	339-343
Δ-12	*CaFAD2.1*	HECGH	105-109	HGRHH	141-145	HVAHH	315-319
	CaFAD2.2	HECGH	104-108	HRRHH	140-144	HVAHH	314-318
	CaFAD2.3	HECGH	104-108	HRRHH	140-144	HVAHH	314-318
	CaFAD2.4	HECGH	104-108	HCRHH	140-144	HVAHH	314-318
	CaFAD6	HDCAH	164-168	HDQHH	200-204	HVPHH	360-364
Δ-15	*CaFAD3*	HDCGH	103-107	HRTHH	139-143	HVIHH	306-310
	CaFAD7/8.1	HDCGH	172-176	HRTHH	208-212	HVIHH	375-379
	CaFAD7/8.2	HDCGH	175-179	HRTHH	211-215	HVIHH	378-382
DES	*CaDES*	HELSH	95-99	HLEHH	132-136	HNEHH	263-267
FAD4	*CaFAD4.1*	HVRTCH	16-21	HAWAH	204-208	HAAHH	233-237
	CaFAD4.2	HVRTCH	16-21	HAWAH	204-208	HAAHH	233-237
	CaFAD4.3	-	-	HAWAH	169-173	HAAHH	198-202
	CaFAD4.4	-	-	HAWAH	169-173	HAAHH	198-202
SLD	*CaSLD.1*	HDSGH	158-162	HNAHH	195-199	-	-
	CaSLD.2	HDSGH	158-162	HNAHH	195-199	-	-
SAD	*CaSAD.2*	HVQVTH	56-61	-	-	-	-
	CaSAD.9	HVQVTH	58-63	HAKEH	242-246	-	-
	CaSAD.10	HQKTH	53-57	-	-	-	-

CaFAD2.1 and *CaSLD.1* (Table 5). *CaSAD.2* and *CaDES* showed a similar expression pattern in both growth systems, with expression levels that ranged from 3.63 to 5.61 $\log_2$(TPM+1) in *in vivo*-grown plantlets and from 2.77 to 5.33 $\log_2$(TPM+1) in *in vitro*-derived plantlets (Fig. 8). A similar expression pattern was also observed in *CaFAD7/8.1* and *CaFAD3* when comparing organs of *in vivo*-grown plantlets, with expression levels ranging from 1.66 to 6.58 $\log_2$(TPM+1) (Fig. 8A). This wide range of expression values mirrors differences in the expression levels among the organs analysed, particularly evident for *CaFAD7/8.1*, as highlighted by LRT and Wald test (Table 5). These two Δ15 desaturase encoding genes clustered together also with *CaADS*, which showed expression values ranging from 0.60 $\log_2$(TPM+1) in roots to 3.33 $\log_2$(TPM+1) in leaves of plantlets grown *in vivo* (Fig. 8A). Similarly, these three genes, together with *CaSAD.11*, formed a cluster also when comparing the expression profiles of *FAD* genes among organs of *in vitro*-derived plantlets, with expression values ranging from 0.47 to 4.76 $\log_2$(TPM+1) (Fig. 8B). Among organs of *in vivo*-grown plantlets, the expression profile of *CaSAD.11* clustered together with that of *CaSAD.10* and *CaSAD.6*, with expression levels ranging from 0.53 to 4.08 $\log_2$(TPM+1), and no expression found for *CaSAD.6* in either leaves or roots, whereas *CaSAD.10* was not expressed in petioles (Fig. 8A). When comparing the organs of *in vitro*-derived plantlets, the expression profile of this latter gene clustered separately, with expression levels ranging from 1.72 $\log_2$(TPM+1) in nodes to 6.21 $\log_2$(TPM+1) in roots and no expression detected in leaves and petioles (Fig. 8B). Differences in the expression levels of this gene among the organs of *in vitro*-derived plantlets were highlighted by LRT and Wald test (Table 5). *CaSAD.8* and *FAD4.2* clustered together when comparing the expression profiles of *FAD* genes in organs of *in vivo*-grown plantlets, with expression levels that went from 0.49 to 5.09 $\log_2$(TPM+1) and no expression detected in roots for *CaFAD4.2*, while *CaSAD.8* was not expressed in internodes and petioles (Fig. 8A). LRT and Wald test reported significant differences among the organs of *in vivo*-grown plantlets for the expression levels of both *CaSAD.8* and *CaFAD4.2* (Table 5). These two genes, together with

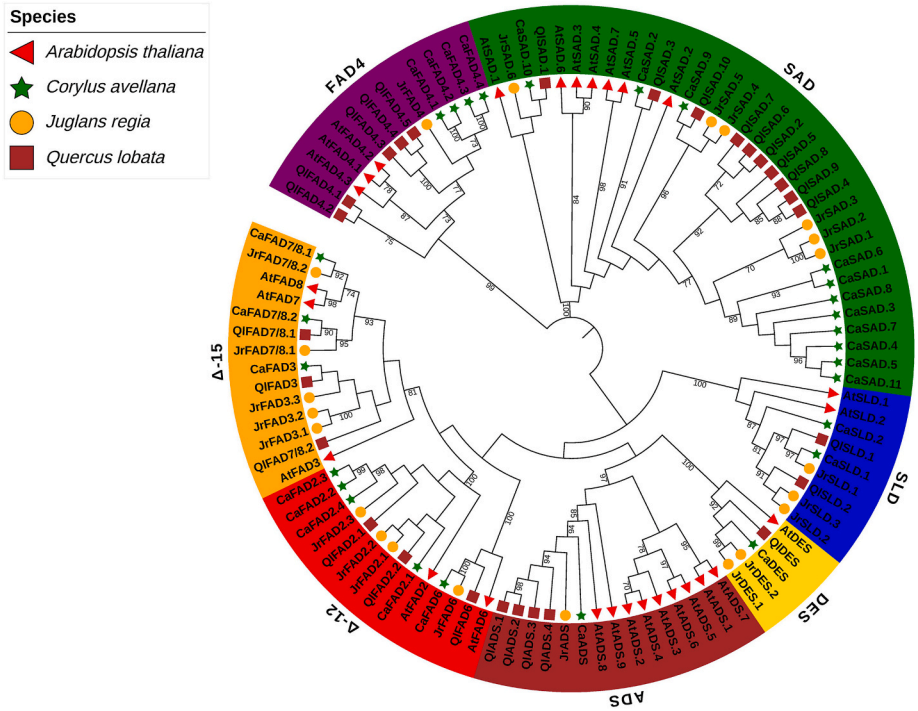

Fig. 4. Phylogenetic tree of FAD protein sequences from *Arabidopsis thaliana*, European hazelnut, Persian walnut and valley oak. The tree was built employing the maximum-likelihood method and Jones-Taylor-Thornton (JTT) substitution model, with 1000 bootstraps. Only bootstrap values ≥ 70% are reported. Clusters of proteins belonging to different subfamilies are highlighted by different colours. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

CaSAD.6, also formed a cluster when comparing expression profiles of *FAD* genes in organs of plantlets derived from *in vitro* tissue culture, with each of these genes being expressed in only two of the five organs analysed (Fig. 8B). In particular, *CaFAD4.2* was expressed in leaves [2.00 $\log_2$(TPM+1)] and roots [1.86 $\log_2$(TPM+1)], *CaSAD.8* was expressed in leaves [1.52 $\log_2$(TPM+1)] and internodes [0.59 $\log_2$(TPM+1)], and *CaSAD.6* was expressed in nodes [1.71 $\log_2$(TPM+1)] and internodes [1.45 $\log_2$(TPM+1)] (Fig. 8B). Other genes, belonging to the subfamilies of soluble plastidial Δ9 desaturase (*CaSAD.3*, *CaSAD.4-1*, *CaSAD.4-2* and *CaSAD.1*) and a microsomal Δ12 desaturase (*CaFAD2.3*), were detected only in transcriptomes of a few organs of plantlets grown *in vivo*, with expression levels that never exceeded 2.65 $\log_2$(TPM+1) (Fig. 8A).

4. Discussion

4.1. Genome-wide association study revealed marker-trait associations with fatty acid content in hazelnut kernels

In the present work, 89 individuals were used in a genome-wide association study to identify regions of the genome associated with the composition of 10 fatty acids in kernels of European hazelnut. The relatively low number of individuals included in the diversity panel presumably reduced the statistical power of the analysis, with the result that significantly associated markers were identified for only a few of the traits analysed (Fig. 1A; Table 1; Fig. S2; Table S2). A recent study by Baytar et al. [9] also used a relatively small population composed of 86 individuals, founding associations with several traits linked to yield traits in European hazelnut. More recently, Frary et al. [10] used a diversity panel comprising 96 accessions and identified 73 loci associated with 15 nutritional traits in this species, including certain fatty acids, namely palmitic acid (16:0), oleic acid (C18:1Δ^9) and linoleic acid (C18:2$\Delta^{9,12}$), as well as total oil content and oil stability.

In the present study, some of the markers associated with the traits analysed were in LD with gene models putatively involved in fatty acid metabolism, as highlighted by GO enrichment analysis (Fig. 1B). The SNP at position 10,986,120 on chromosome 2 was associated with linoleic acid content in hazelnut kernels (Fig. 1A; Table 1; Fig. S2; Table S2). This marker was in LD with 47 genes, three of which were annotated with the "fatty acid metabolic process" GO category (GO:0006631). Upstream of this SNP were *LOC132170256* and *LOC132170424*, located a short distance from each other (<20 kb) and encoding, respectively, a mitochondrial acyl carrier protein and a chloroplastic biotin carboxyl carrier protein of acetyl-CoA carboxylase. Downstream of this SNP was *CaSAD.2* (*LOC132171848*), whose protein product is a stearoyl-[acyl-carrier-protein] Δ9-desaturase, an enzyme that converts stearoyl-ACP (C18:0) to oleoyl-ACP (C18:1Δ^9; [60]). This represents the first desaturation step in the synthesis of unsaturated fatty acids and acts as a rate-limiting step for the production of oleic acid (C18:1 Δ9), which can then be further desaturated to linoleic acid (C18:2$\Delta^{9,12}$) in the endoplasmic reticulum by FAD2 and in plastids by FAD6 [61]. It was reported that the heterologous expression of two *SAD* genes of tung tree (*Vernicia fordii*) in a mutant line of *Arabidopsis thaliana* lacking both microsomal Δ15 desaturase and the fatty acid elongation 1-condensing enzyme (*fad3fae1*) resulted in a significant increase in linoleic acid content in seed oil, highlighting the key role of this gene in the synthesis of C18 unsaturated fatty acids [62].

A SNP associated with arachidic acid at position 22,707,256 of chromosome 4 presented 58 genes within the area delimited by the estimated LD decay value, among which there were four genes that also encode SAD: *CaSAD.3* (*LOC132177385*), *CaSAD.4* (*LOC132179807*), *CaSAD.5* (*LOC132178572*) and *CaSAD.6* (*LOC132179592*; Fig. 1A; Table 1; Fig. S2; Table S2). Although acyl-ACP desaturases are not directly involved in the biosynthesis of arachidic acid, a C20 saturated fatty acid, the strong correlations observed among fatty acids are consistent with the interconnected nature of fatty acid biosynthesis, whereby changes in the activity of enzymes involved in the synthesis of one fatty acid can affect the accumulation of other fatty acids derived

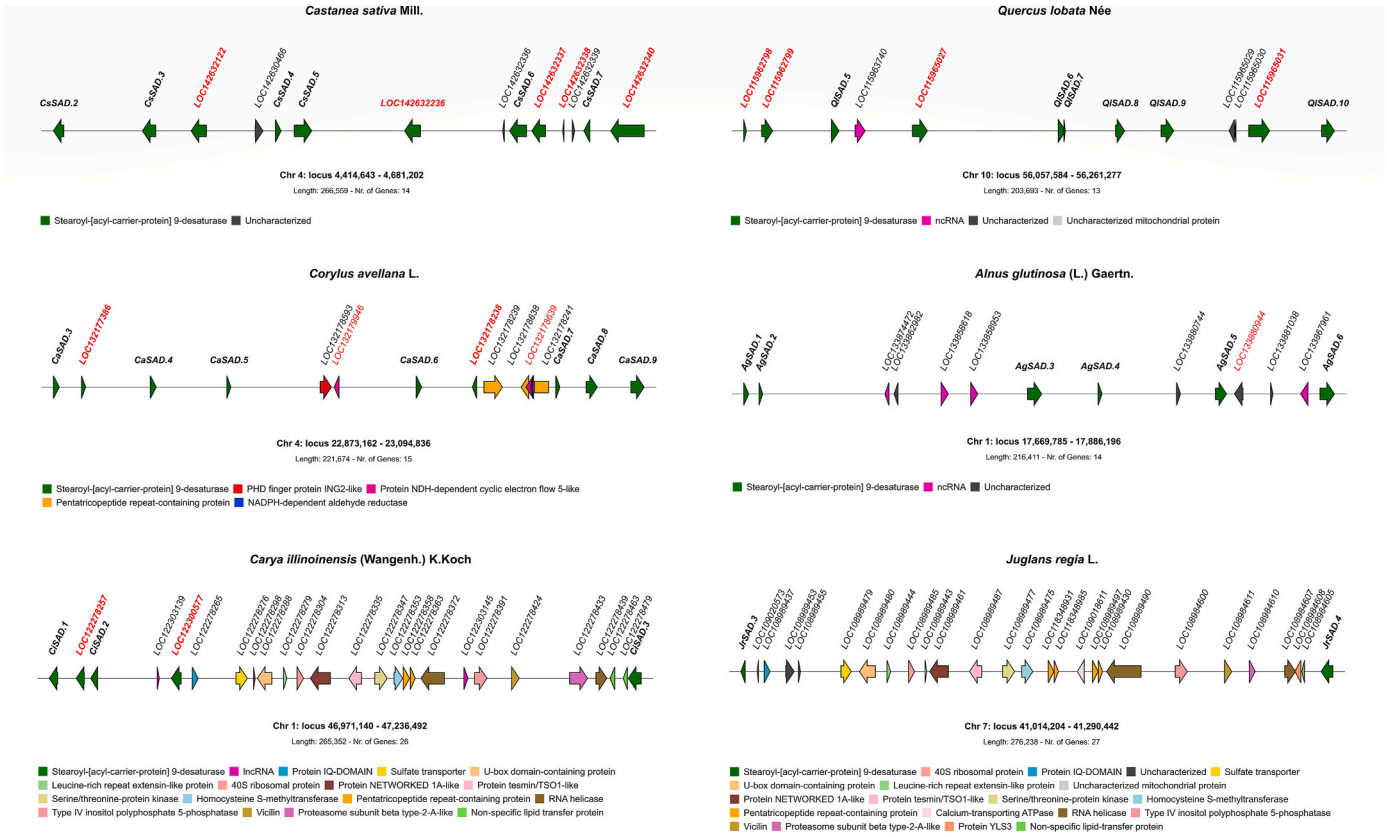

Fig. 5. Genomic architecture of *SAD* gene clusters in the reference genome of several species belonging to the order Fagales, namely European chestnut, valley oak, European hazelnut, common alder and pecan. For Persian walnut, in which the *SAD* gene subfamily has not expanded at this locus (≤2 genes), the orthologous genomic region corresponding to the pecan *SAD* gene cluster is shown. Gene models encoding SAD are indicated in bold, whereas sequences annotated as "pseudogenes" are indicated in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

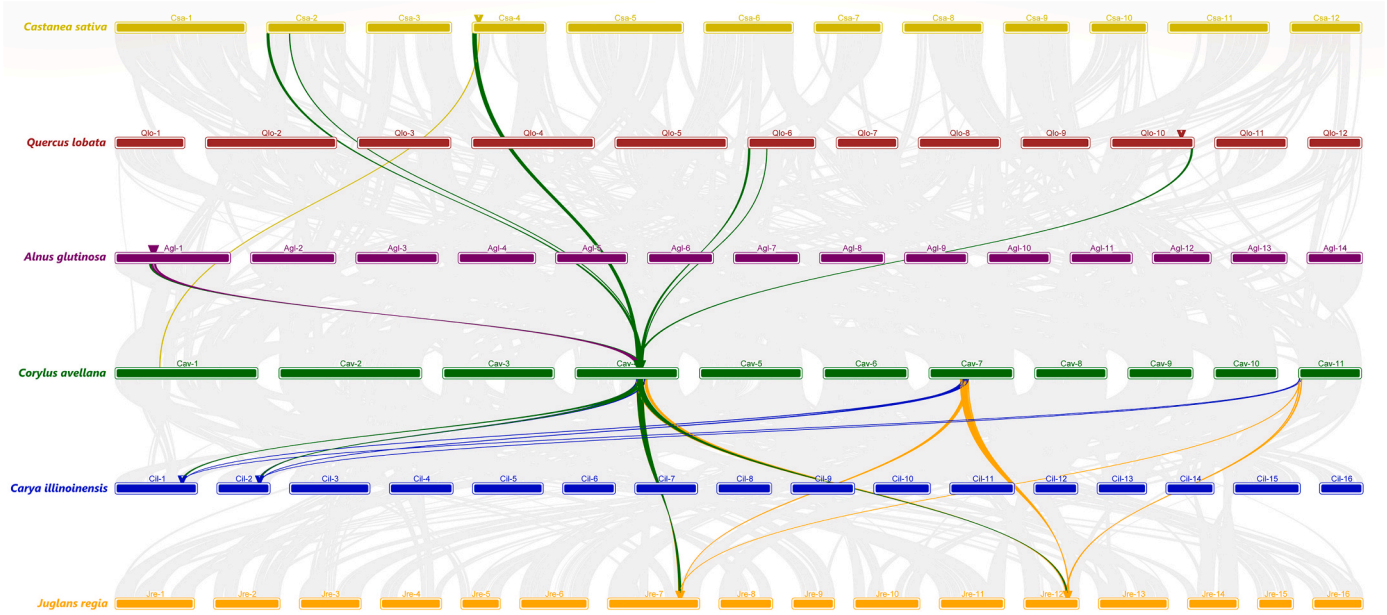

Fig. 6. Analysis of synteny between European hazelnut and other Fagales, namely European chestnut, valley oak, common alder, pecan and Persian walnut. Arrows indicate the position of coding sequences located 1 Mb upstream and downstream of the *SAD* gene cluster in *Fagaceae* (sweet chestnut and valley oak) and *Betulaceae* (common alder and European hazelnut), and genes surrounding *CiSAD.3* (chromosome 1) and *CiSAD.4* (chromosome 2) in pecan, and *JrSAD.4* (chromosome 7) and *JrSAD.6* (chromosome 12) in Persian walnut. Genes located in collinear blocks comprising ≥ 20 genes, situated in the regions indicated by the arrows, are highlighted in different colours depending on their respective species. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

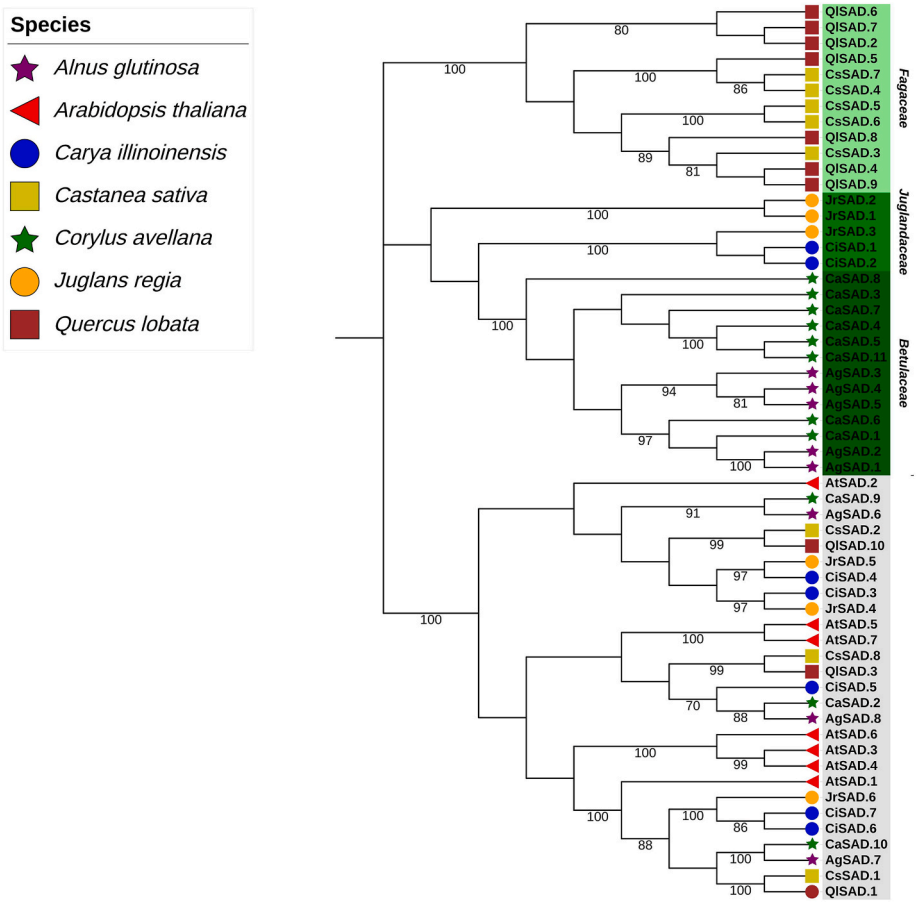

Fig. 7. Phylogenetic tree of SAD protein sequences from common alder, *Arabidopsis thaliana*, European chestnut, European hazelnut, Persian walnut and valley oak. The tree was built employing the maximum-likelihood method and Jones-Taylor-Thornton (JTT) substitution model, with 1000 bootstraps. Only bootstrap values ≥ 70% are reported. The sequences shown in green belong to clusters composed exclusively of proteins from Fagales species, with a different shade of colour corresponding to the taxonomic family of the species to which they belong. Sequences from a mixed cluster containing proteins from all analysed species are highlighted in grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

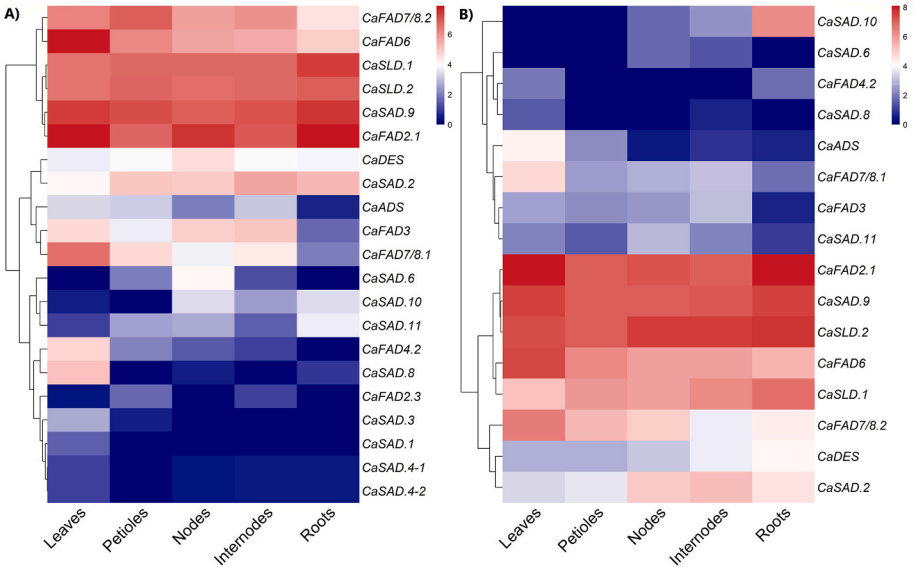

Fig. 8. Heatmap showing expression levels of *FAD* genes in vegetative organs of both plantlets of cultivar 'Tonda Gentile Romana' grown *in vivo* (**A**) and *in vitro* (**B**) with hierarchical clustering of genes. For each sample, the expression levels of *FAD* genes are reported as the average $\log_2$(TPM+1) of three replicates.

from the same metabolic pathway (Fig. S1). In particular, the positive correlation observed between stearic acid (C18:0) and arachidic acid (C20:0) (Spearman's $\rho = 0.64$; *p-value* < 0.001) suggests that variation in the C18:0 pool may be associated with changes in C20:0 accumulation

Table 5
Log_2FC values of *FAD* gene expression resulting from pairwise comparisons (Wald test) between vegetative organs of plantlets of cultivar "Tonda Gentile Romana" derived from *in vivo* and *in vitro* growth systems. The adj. *p-values* of both Wald test (in brackets) and Likelihood Ratio Test (LRT) are also reported. Only significant pairwise comparisons and LRTs were reported (adj. *p-value* < 0.05). In bold are indicated genes which showed pairwise comparisons and/or LRT adj. *p-values* < 0.001.

Growth system	Gene	Leaves/ Petioles	Leaves/ Nodes	Leaves/ Internodes	Leaves/ Roots	Petioles/ Nodes	Petioles/ Internodes	Petioles/ Roots	Nodes/ Internodes	Nodes/ Roots	Internodes/ Roots	LRT
In vivo	***CaFAD6***	**1.41 (2.19 x 10^{-4})**	**1.9 (4.70 x 10^{-7})**	**2.08 (3.72 x 10^{-7})**	**3.03 (7.81 x 10^{-13})**	-	-	1.25 (3.75 x 10^{-3})	-	-	-	**(4.98 x 10^{-16})**
	CaFAD7/ 8.1	1.51 (5.08 x 10^{-3})	**1.97 (2.82 x 10^{-4})**	1.82 (1.74 x 10^{-3})	**4.14 (1.66 x 10^{-7})**	-	-	1.69 (1.70 x 10-2)	-	-	-	**(7.87 x 10^{-10})**
	CaSAD.8	**3.24 (5.41 x 10^{-4})**	2.53 (3.52 x 10^{-3})	3.35 (1.83 x 10^{-3})	3.12 (9.44 x 10^{-3})	-	-	-	-	-	-	**(1.82 x 10^{-7})**
	CaFAD7/ 8.2	-	-	-	**1.76 (1.44 x 10^{-4})**	-	-	**2.44 (1.12 x 10^{-7})**	-	1.17 (5.88 x 10^{-3})	1.53 (1.14 x 10^{-3})	**(7.79 x 10^{-7})**
	CaFAD4.2	-	1.43 (4.14 x 10^{-2})	-	**4.59 (8.05 x 10^{-4})**	-	-	-	-	-	-	**(6.71 x 10^{-4})**
	CaFAD3	-	-	-	2.21 (1.18 x 10^{-2})	-	-	-	-	2.63 (3.86 x 10^{-3})	2.84 (2.92 x 10^{-3})	(3.34 x 10^{-3})
	CaSAD.10	-	−1.39 (4.59 x 10^{-2})	-	-	-	-	−2.79 (1.80 x 10^{-2})	-	-	-	(5.56 x 10^{-3})
	CaSAD.6	-	−2.02 (9.80 x 10^{-3})	-	-	-	-	-	-	3.09 (1.45 x 10^{-2})	-	(7.10 x 10^{-3})
	CaSAD.2	-	-	−1.33 (9.97 x 10^{-3})	-	-	-	-	-	-	-	-
In vitro	***CaFAD6***	0.85 (1.28 x 10^{-2})	**1.28 (9.93 x 10^{-5})**	**1.38 (6.48 x 10^{-5})**	**1.87 (2.29 x 10^{-7})**	-	-	-	-	-	-	**(3.62 x 10^{-8})**
	CaFAD2.1	0.82 (2.25 x 10^{-3})	0.64 (1.43 x 10^{-2})	**0.88 (6.41 x 10^{-4})**	-	-	-	−0.82 (1.00 x 10^{-3})	-	−0.58 (1.18 x 10^{-2})	**−0.94 (4.55 x 10^{-4})**	**(2.28 x 10^{-6})**
	CaSAD.10	-	-	-	**−5.11 (5.06 x 10^{-4})**	-	-	**−5.03 (6.40 x 10^{-4})**	-	−1.84 (4.11 x 10^{-2})	-	**(1.74 x 10^{-5})**
	CaFAD7/ 8.2	-	0.98 (3.58 x 10^{-2})	**1.92 (5.37 x 10^{-4})**	**1.86 (8.31 x 10^{-4})**	-	-	-	-	-	-	**(2.20 x 10^{-4})**
	CaSLD.1	-	-	−0.81 (3.36 x 10^{-2})	**−1.35 (3.76 x 10^{-4})**	-	-	-	-	-	-	(3.63 x 10^{-3})
	CaADS	-	2.15 (1.69 x 10^{-2})	-	2.54 (2.36 x 10^{-2})	-	-	-	-	-	-	(3.77 x 10^{-3})
	CaSAD.2	-	−1.2 (2.92 x 10^{-2})	−1.49 (1.10 x 10^{-2})	-	-	-	-	-	-	-	(1.64 x 10^{-2})
	CaFAD7/ 8.1	-	-	-	1.89 (2.04 x 10^{-2})	-	-	-	-	-	-	(3.33 x 10^{-2})
	CaSLD.2	-	-	-	-	-	-	−0.47 (4.95 x 10^{-2})	-	-	-	-

(Fig. S1). Stearoyl-ACP desaturases catalyse the desaturation of C18:0-ACP to C18:1-ACP in plastids. Therefore, variation in their activity could affect the partitioning of stearic acid (C18:0) between desaturation and alternative metabolic pathways [63,64]. Stearic acid (C18:0) that is not desaturated may contribute to the pool of fatty acids exported from plastids and subsequently activated to acyl-CoAs, which can enter the very-long-chain fatty acid biosynthetic pathway in the endoplasmic reticulum [65]. Thus, although acyl-ACP desaturases are not directly involved in C20:0 biosynthesis, the marker-trait association identified by GWAS may reflect an indirect effect of altered C18:0 availability and metabolic partitioning, potentially contributing to variation in arachidic acid (C20:0) content. However, this interpretation remains hypothetical, as no functional validation was performed in the present study.

The results obtained from the preliminary GWAS analysis presented in this study suggest the identification of marker-trait associations for fatty acid content in hazelnut kernels. However, expanding the diversity panel and collecting data spanning multiple seasons will be essential for confirming these results over time, enabling the identification of additional loci potentially involved in determining the fatty acid composition of hazelnut kernels, a complex trait strongly influenced by environmental factors, and the development of molecular markers suitable for use in plant breeding.

4.2. Most SAD genes are organized into a single cluster in several Fagales species

Genome-wide identification of genes encoding fatty acid desaturase pinpointed 27 members of this gene family in the reference genome of European hazelnut cv. 'Tombul'. This number is in line with that found in other species belonging to the order Fagales, suggesting a general conservation of genes involved in unsaturated fatty acid biosynthesis, although with differences in the number of members belonging to different subfamilies (Fig. 4; Table S3). A common feature among the analysed species belonging to the families of *Fagaceae* and *Betulaceae* was the presence of a genomic region spanning between 200 and 270 kb containing several copies of *SAD* genes arranged in close contiguity, forming a cluster of tandem and proximal paralogues. Although being characterized by a different genetic organization and being composed of only five genes annotated as *SAD*, two of which reported as pseudogenes, a similar structure was also found in pecan, while the orthologous locus in Persian walnut had only two *SAD* genes separated by almost 280 kb (Fig. 5).

SAD genes are frequently found to be arranged in small tandem arrays in plant genomes, as observed on chromosome 3 of *Arabidopsis thaliana*, where three paralogues of this gene, *AtSAD.3* (*AT3G02610*), *AtSAD.4* (*AT3G02620*) and *AtSAD.5* (*AT3G02630*) are present as tandem duplicates [66]. Three contiguous *SAD* paralogues were also reported in tigernut (*Cyperus esculentus* L.; [67]) and wheat (*Triticum aestivum* L.), with this latter species having arrays of three tandem copies on chromosomes 2, 5 and 2 of genome A, B and D, respectively [68]. Zare et al. [69] described the expansion of *SAD* genes in the genome of *Salvia hispanica*, which resulted from tandem duplication events, and linked this genetic feature to high levels of α-linolenic acid in the seeds of this species. Expansion of *SAD* genes was also reported in the woody plant *Akebia trifoliata* subsp. *australis*, which carried 12 copies of this Δ9-desaturase encoding gene [70].

Species belonging to the taxonomic order of Fagales were characterized by an early evolutionary history marked by several gene and whole genome duplication events that resulted in high phenotypic diversification [71]. Within this clade, families like *Betulaceae*, *Fagaceae* and *Juglandaceae* include some of the most ecologically and economically important species that develop seeds containing large amounts of oil with different acidic profiles, with the most marked differences related to oleic and linoleic acid content [72,73]. Among these crops, European hazelnut has the highest concentration of oleic acid (C18:1Δ^{9}), which represents more than 75% of total fatty acids in its oil, while linoleic acid (C18:2$\Delta^{9,12}$) accounts for about 7-14% [74,75]. A similar profile is also showed by several *Quercus* species, whose acorn oil is particularly rich in oleic acid (65%), followed by linoleic acid (17–37%; [76]). Among *Juglandaceae*, oils obtained from pecan nuts are also rich in oleic acid (49–77%), which is followed by linoleic acid (13–40%). In contrast, walnut oil shows an inverse profile, as linoleic acid is the most representative fatty acid, ranging from 45% to 68%, followed by oleic acids, in a range of 12-35% [77]. PUFAs also result to be the most prevalent oils in chestnut, which belongs to the family of *Fagaceae*, and in which linoleic acid accounts for 38-50%, followed by oleic acid, ranging from 20 to 37%, although its nuts present a conspicuously lower fat content (2-5%) compared to other Fagales [78,79].

The present study highlighted some conservation of collinear relationships between the regions surrounding the *SAD* gene cluster in European hazelnut and those flanking this cluster in European chestnut, valley oak and common alder, although direct collinearity between *SAD* genes was observed only when comparing European hazelnut with members of *Juglandaceae*, among which only pecan showed a clustered arrangement of *SAD* genes. Both *JrSAD.4* and *JrSAD.5*, respectively on chromosome 7 and 12 of Persian walnut, and *CiSAD.3* and *CiSAD.4*, respectively on chromosome 1 and 2 of pecan, showed a collinear relationship with *CaSAD.3*, the genes that borders the *SAD* gene cluster on chromosome 4 of European hazelnut in its most proximal side (Fig. 6). Differently, phylogenetic analysis of SAD protein sequences showed that JrSAD.4, JrSAD.5, CiSAD.3 and CiSAD.4 are more closely related to CaSAD.9, whose coding sequence is in the most distal portion of the *SAD* gene cluster of European hazelnut (Fig. 5; Fig. 7). These protein sequences have also been grouped together with QlSAD.10, CsSAD.2 and AgSAD.6, whose coding sequences borders the *SAD* gene cluster in their respective species, and are putatively orthologs of *AtSAD.2* (Fig. 7). All the other SAD sequences encoded by genes included in the clusters of Fagales species formed two groups, one including sequences of *Fagaceae* and one with sequences of *Betulaceae* and *Juglandaceae* (Fig. 7). This, together with the differences in the organization and number of genes within the *SAD* gene clusters of these species, would suggest a common origin in this taxonomic order for this loci, which have undergone independent evolution by expanding through tandem and proximal duplications in each lineage, although further analysis will be needed to confirm this hypothesis.

4.3. Expression profile of FAD genes in vegetative organs is generally conserved between in vivo and in vitro growth systems

Although an overall reduction in the expression levels of *FAD* genes was observed in samples grown *in vitro* with respect to those grown *in vivo*, the expression profile of *FAD* genes in vegetative organs is generally conserved between the two growth systems investigated (Fig. 8; Fig. S3). Transcriptome analysis also indicated the expression in vegetative organs of several genes included in the *SAD* gene cluster on chromosome 4 (Fig. 8). Among these genes, only *CaSAD.9* showed a constitutive expression profile with high expression levels [$\log_2(\text{TPM}+1) \geq 6.83$] in all the organs of plantlets from both growth systems, while other genes were expressed only in a few organs at lower levels [$\log_2(\text{TPM}+1) \leq 5.09$] (Fig. 8). Differences in the expression levels of contiguous paralogous genes have often been observed in plants, and these phenomena have been proposed to result from the incomplete duplication of promoter regions and coding sequences, together with homology-dependent gene silencing [80]. Among genes encoding soluble plastidial Δ9-desaturases, *CaSAD.2* and *CaSAD.11* also showed a constitutive expression profile, even though at lower expression levels compared to *CaSAD.9* [$\log_2(\text{TPM}+1) \leq 5.61$] (Fig. 8). Phylogenetic analysis indicated a high sequence homology between CaSAD.9 and AtSAD.2 (AT2G43710), whose encoding gene showed a ubiquitous expression pattern in *Arabidopsis thaliana*, resulting to be the most expressed *SAD* gene during seed development (Fig. 7; [81]). Orthologs of this gene were also the most expressed *SADs* in four different oilseeds, namely rapeseed (*Brassica napus*), castor (*Ricinus communis*), burning bush (*Euonymus alatus*), and nasturtium (*Tropaeolum majus*; [82]. High expression levels (FPKM > 140) were registered in developing kernels of walnut for *JrSAD.4* (*LOC108984606*) and *JrSAD.5* (*LOC109005061*), two putative orthologs of *CaSAD.9* (Fig. 7; [83]). Huang et al. [84] identified 5 *SAD* paralogues in hickory (*Carya catayensis*) and reported *CcSAD-2*, a putative ortholog of *AtSAD.1* (*AT1G43800*), as the most expressed *SAD* gene during all the phases of embryo development, followed by *CcSAD-5*, a putative ortholog of *AtSAD.2* (*AT2G43710*).

Among the four microsomal *Δ12 FAD* genes of European hazelnut, only *CaFAD2.1* was constitutively expressed at high levels [$\log_2(\text{TPM}+1) \geq 6.78$] in the organs of plantlets derived from both growth systems (Fig. 8). A similar scenario was observed in another species with a high-oleate content, the wild olive (*Olea europaea* L. var. *sylvestris*), which has five *FAD2* paralogs whose expression was hypothesized to be regulated by a small interfering RNA (siRNA) capable of binding to the 5′ UTR of these genes during fruit ripening, with the exception of *FAD2-3* that has an insertion in the putative siRNA binding site [85]. Furthermore, *CaFAD2.1* has a long intron in the 5' UTR region (>6000 bp) that could be essential in controlling the expression of this gene, similarly to what was observed in *Brassica napus FAD2A5-1* gene (Fig. 3D; [86]). The importance of an intronic sequence in regulating

gene expression of a *FAD2* gene was also pointed out in cultivated olive (*Olea europaea* L. subsp. *europaea* var. *europaea*; [87]).

Similar expression profiles were shared by all the *Δ15 FAD* genes, with expression levels generally higher in aerial tissues (Fig. 8; Fig. S3). This is consistent with what has been observed in transgenic *Arabidopsis* plants expressing plastidial *Δ15 FAD7* and *FAD8* fused to green fluorescent protein (GFP), where a different promoter activity for the two genes was highlighted, with *FAD7* showing higher expression levels compared to *FAD8*, and a higher expression level in leaves with respect to roots [88].

FAD genes, which are responsible for the desaturation of fatty acids that form cellular membranes, conferring fluidity and general stability, are directly involved in cold stress tolerance [89,90]. This aspect is of fundamental importance from an agronomic point of view since European hazelnut is a plant species characterized by an early phenology, with pollination occurring in the winter season, between late December and early February in the Northern hemisphere, and budbreak generally takes place between the second half of February and the first half of March, when frosts represent a serious threat to the vegetative development of this crop [91].

Since the data presented in this study do not consider the expression profile of *FAD* genes in kernels, which constitute the edible part of the European hazelnut, it is not possible to establish a direct link between the results of the association mapping to those of the transcriptomic analysis. A further metabolic validation would also be necessary to link the transcriptome profile of *FAD* genes in vegetative organs with their fatty acid composition. However, the constitutive expression pattern exhibited by some of these genes, particularly those expressed at higher levels compared to other paralogues, makes them interesting candidates warranting closer attention in future studies on *FAD* genes in European hazelnut.

5. Future perspectives

Although association mapping studies on the European hazelnut have become increasingly common in recent years, the application of genetic studies like these to this species is still limited compared to other woody crops. Association mapping studies on this species should include a larger number of genotypes and data spanning multiple seasons to increase the statistical power of the analysis and account for environmental effects that may influence the distribution of certain phenotypes. Although the collection of data over multiple seasons is feasible, *ex situ* field germplasm collections that include a large number of accessions are uncommon for this woody crop. To this end, greater collaboration should be encouraged among the institutions that manage hazelnut germplasm banks. The preliminary genome-wide association study reported in this work identified some significant marker-trait associations for some of the traits analysed. Since some of the genes in LD with the significantly associated markers are annotated as putative *SADs*, their role in fatty acid metabolism in plants is well documented. Nevertheless, genetic engineering studies aimed at modifying the expression levels of these genes would be essential to assess how this would affect the fatty acid profile in this species. On the other hand, greater efforts are also needed to improve the regenerative potential of the European hazelnut through *in vitro* tissue culture, overcoming its recalcitrance, which represents the main obstacle to the application of biotechnological tools in this species.

This study reported, for the first time in the European hazelnut, the presence of a cluster of *SAD* genes on chromosome 4 of this species. Comparative genomic analysis revealed the arrangement of multiple copies of *SAD* genes in close proximity in other members of the order Fagales as well, while synteny analysis indicated the conservation of collinear relationships among the genes surrounding these clusters, suggesting a common ancestral origin for these loci. However, further analysis will be necessary to reconstruct the evolutionary history of these genomic regions across the various lineages within this taxonomic order. In particular, the analysis considered only a limited number of species, for which only the reference genome was included in the analysis. Further studies should consider a larger number of species, also evaluating multiple genomes for each species, where available.

The present study also provided information on the expression profile of *FAD* genes in the vegetative organs of plantlets derived from two different growth systems, identifying several genes that were constitutively expressed in these organs at higher levels than their paralogs. Given the crucial role played by *FAD* genes in the response to cold stress, a condition that often negatively affects the yield of this species, an interesting area for future research could be the analysis of the expression profile of these genes in response to cold stress conditions. Furthermore, the expression profile of these genes should also be evaluated in seeds harvested at different stages of development, since the kernel is the edible part of the European hazelnut and fatty acids account for a conspicuous proportion of its weight.

6. Conclusion

The present study adopted a multidisciplinary approach to characterize some of the genetic basis that regulates fatty acid content in European hazelnut, one of the most cultivated nut crops in the world. Although the genome-wide association study for fatty acid content in kernels reported some statistically significant associations, further analysis that will include a wider diversity panel and data collected in multiple seasons would be necessary to confirm the reliability of the results obtained in the present study. Among the markers significantly associated with the traits analysed, a SNP was in proximity to a ~220 kb cluster of seven gene models putatively encoding stearoyl-[acyl-carrier-protein] Δ9-desaturases. Comparative genomics revealed the presence of contiguous *SAD* genes arranged into a cluster also in other members of the order Fagales. Phylogenetic analysis of the SAD proteins revealed that all sequences encoded by the genes within the *SAD* clusters formed two distinct groups based on the taxonomic family of the species to which they belonged, with the sole exception of the sequences encoded by one of the two genes flanking the gene cluster, which formed a distinct group and are putative orthologs of *AtSAD.2* (*AT2G43710*). These analyses may suggest the presence of this locus in a common ancestor, and the subsequent independent expansion of the members of this subfamily of genes through tandem and proximal duplication events, although additional analyses are required to unravel the evolutionary history of this locus in Fagales species. Finally, transcriptome sequencing of several organs including leaves, petioles, nodes, internodes and roots of plantlets derived from *in vivo* and *in vitro* growth systems enabled the study of the expression profiles of *FAD* genes in these vegetative organs. These data enabled the identification of genes constitutively expressed at higher levels compared to other paralogues and provided insights into the general expression pattern of *FAD* genes across vegetative organs. Further transcriptome analysis is needed to evaluate the expression levels of *FAD* genes during the development of kernels, which are the edible part of hazelnut, and to provide support to the results obtained by the association mapping analysis. Furthermore, biochemical validation of the transcriptome profile will also be necessary to link the results obtained by the expression analysis with the fatty acid composition in the organs analysed. In conclusion, the present study represents one of the first attempts to shed light on the genetic architecture underlying fatty acid content in European hazelnut, providing a foundation for future research in this nut crop.

CRediT authorship contribution statement

Andrea Ferrucci: Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Leonardo Caproni:** Software, Methodology, Data curation. **Svenja Mager:** Software, Methodology, Data curation. **Leontina Lipan:** Resources,

Methodology, Data curation. **Agustí Romero-Aroca:** Resources, Methodology, Data curation. **Mercè Rovira:** Writing – review & editing, Resources, Methodology, Data curation. **Claudio Todeschini:** Funding acquisition, Data curation. **Stuart James Lucas:** Writing – review & editing, Software, Methodology, Data curation. **Fuqiang Cui:** Writing – review & editing, Project administration, Funding acquisition. **Matteo Dell'Acqua:** Writing – review & editing, Software, Methodology, Data curation. **Valerio Cristofori:** Writing – review & editing, Project administration, Funding acquisition. **Cristian Silvestri:** Writing – review & editing, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was funded by the State Key Laboratory of Subtropical Silviculture (SKLSS-KF2024-07) and carried out within the framework of the Ministry of University and Research (MUR) initiative "Department of Excellence" (Law 232/2016) DAFNE Project 2023-27 "Digital, Intelligent, Green and Sustainable acronym: D.I.Ver.So). All the bioinformatics calculations and analyses were carried out at the DAFNE HPC scientific computing centre of the University of Tuscia.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jafr.2026.103294.

Data availability

Data will be made available on request.

References

[1] A. Romero-Aroca, M. Rovira, V. Cristofori, C. Silvestri, Hazelnut kernel size and industrial aptitude, Agriculture 11 (2021) 1115, https://doi.org/10.3390/agriculture11111115.
[2] A. Orem, F.B. Yucesan, C. Orem, B. Akcan, B.V. Kural, C. Alasalvar, F. Shahidi, Hazelnut-enriched diet improves cardiovascular risk biomarkers beyond a lipid-lowering effect in hypercholesterolemic subjects, J. Clin. Lipidol. 7 (2013) 123–131, https://doi.org/10.1016/j.jacl.2012.10.005.
[3] S. Perna, A. Giacosa, G. Bonitta, C. Bologna, A. Isu, D. Guido, M. Rondanelli, Effects of hazelnut consumption on blood lipids and body weight: a systematic review and Bayesian meta-analysis, Nutrients 8 (2016) 747, https://doi.org/10.3390/nu8120747.
[4] S. Talebi, F. Khodagholi, Z. Bahaeddin, M. Ansari Dezfouli, A. Zeinaddini-Meymand, S. Berchi Kankam, F. Foolad, F. Alijaniha, F. Fayazi Piranghar, Does hazelnut consumption affect brain health and function against neurodegenerative diseases? Nutr. Neurosci. 27 (2024) 1008–1024, https://doi.org/10.1080/1028415X.2023.2296164.
[5] V. Cristofori, S. Ferramondo, G. Bertazza, C. Bignami, Nut and kernel traits and chemical composition of hazelnut (*Corylus avellana* L.) cultivars, J. Sci. Food Agric. 88 (2008) 1091–1098, https://doi.org/10.1002/jsfa.3203.
[6] A.İ. Köksal, N. Artik, A. Şimşek, N. Güneş, Nutrient composition of hazelnut (Corylus avellana L.) varieties cultivated in Turkey, Food Chem. 99 (2006) 509–515, https://doi.org/10.1016/j.foodchem.2005.08.013.
[7] L. Tibbs Cortes, Z. Zhang, J. Yu, Status and prospects of genome-wide association studies in plants, Plant Genome 14 (2021) e20077, https://doi.org/10.1002/tpg2.20077.
[8] Q. Du, W. Lu, M. Quan, L. Xiao, F. Song, P. Li, D. Zhou, J. Xie, L. Wang, D. Zhang, Genome-wide association studies to improve wood properties: challenges and prospects, Front. Plant Sci. 9 (2018) 1912, https://doi.org/10.3389/fpls.2018.01912.
[9] A.A. Baytar, E.G. Yanar, A. Frary, S. Doğanlar, Association mapping and candidate gene identification for yield traits in European hazelnut (*Corylus avellana* L.), Plant Direct 8 (2024) e625, https://doi.org/10.1002/pld3.625.
[10] A. Frary, E.G. Yanar, A.A. Baytar, N.T. Eken, F. Keskin, S.E. Kınal, İ. Taşdemir, S. Doğanlar, Nutritional quality in hazelnut (Corylus avellana L.) germplasm and genome wide identification of associated loci, Tree Genet. Genomes 22 (2026) 15, https://doi.org/10.1007/s11295-026-01740-1.
[11] M. Salardi-Jost, S. Mager, G. Dini, M.W. Macharia, M. Rovira, V. Cristofori, M. E. Pé, A. Cavallero, C. Todeschini, G. Genova, M. Dell'Acqua, Genomic characterization of a global Corylus avellana L. collection: insights into hazelnut ancestry and genetic determinants of production traits, BMC Plant Biol. 25 (2025) 1738, https://doi.org/10.1186/s12870-025-07596-2.
[12] C. Silvestri, V. Cristofori, M. Ceccarelli, M.E. Caceres, J. Escribà-Lacuesta, E. Rugini, Adventitious shoot organogenesis from leaf and petiole explants of European hazelnut, Plant Cell Tissue Organ Cult. 126 (2016) 59–65, https://doi.org/10.1007/s11240-016-0976-7.
[13] W. Xia, T. Luo, Y. Dou, W. Zhang, A.S. Mason, D. Huang, X. Huang, W. Tang, J. Wang, C. Zhang, Y. Xiao, Identification and validation of candidate genes involved in fatty acid content in oil palm by genome-wide association analysis, Front. Plant Sci. 10 (2019) 1263, https://doi.org/10.3389/fpls.2019.01263.
[14] Y. Zheng, C. Chen, Y. Liang, R. Sun, L. Gao, T. Liu, D. Li, Genome-wide association analysis of the lipid and fatty acid metabolism regulatory network in the mesocarp of oil palm (Elaeis guineensis Jacq.) based on small noncoding RNA sequencing, Tree Physiol. 39 (2019) 356–371, https://doi.org/10.1093/treephys/tpy091.
[15] E. Niu, S. Gao, W. Hu, C. Zhang, D. Liu, G. Shen, S. Zhu, Genome-wide identification and functional differentiation of fatty acid desaturase genes in Olea europaea L, Plants 11 (2022) 1415, https://doi.org/10.3390/plants11111415.
[16] Y. Niu, G. Li, S. Ni, X. He, C. Zheng, Z. Liu, L. Gong, G. Kong, W. Li, J. Liu, The chromosome-scale reference genome of Macadamia tetraphylla provides insights into fatty acid biosynthesis, Front. Genet. 13 (2022) 835363, https://doi.org/10.3389/fgene.2022.835363.
[17] K. Liu, S. Zhao, S. Wang, H. Wang, Z. Zhang, Identification and analysis of the FAD gene family in walnuts (Juglans regia L.) based on transcriptome data, BMC Genom. 21 (2020) 299, https://doi.org/10.1186/s12864-020-6692-z.
[18] D. López Alonso, F. García-Maroto, J. Rodríguez-Ruiz, J.A. Garrido, M.A. Vilches, Evolution of the membrane-bound fatty acid desaturases, Biochem. Systemat. Ecol. 31 (2003) 1111–1124, https://doi.org/10.1016/S0305-1978(03)00041-3.
[19] M.T. Nakamura, T.Y. Nara, Structure, function, and dietary regulation of Δ6, Δ5, and Δ9 desaturases, Annu. Rev. Nutr. 24 (2004) 345–376, https://doi.org/10.1146/annurev.nutr.24.121803.063211.
[20] M.A. Berestovoy, O.S. Pavlenko, I.V. Goldenkova-Pavlova, Plant fatty acid desaturases: role in the life of plants and biotechnological potential, Biol Bull Rev 10 (2020) 127–139, https://doi.org/10.1134/S2079086420020024.
[21] E.M. Dvorianinova, O.L. Zinovieva, E.N. Pushkova, D.A. Zhernova, T.A. Rozhmina, L.V. Povkhova, R.O. Novakovskiy, E.A. Sigova, A.A. Turba, E.V. Borkhert, G. S. Krasnov, C. Ruan, A.A. Dmitriev, N.V. Melnikova, Key FAD2, FAD3, and SAD genes involved in the fatty acid synthesis in flax identified based on genomic and transcriptomic data, IJMS 24 (2023) 14885, https://doi.org/10.3390/ijms241914885.
[22] S.J. Lucas, K. Kahraman, B. Avşar, R.J.A. Buggs, I. Bilge, A chromosome-scale genome assembly of European hazel (*Corylus avellana* L.) reveals targets for crop improvement, Plant J. 105 (2021) 1413–1430, https://doi.org/10.1111/tpj.15099.
[23] H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. https://doi.org/10.48550/arXiv.1303.3997, 2013.
[24] P. Danecek, J.K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M.O. Pollard, A. Whitwham, T. Keane, S.A. McCarthy, R.M. Davies, H. Li, Twelve years of SAMtools and BCFtools, GigaScience 10 (2021) giab008, https://doi.org/10.1093/gigascience/giab008.
[25] N.C. Rochette, A.G. Rivera-Colón, J.M. Catchen, Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics, Mol. Ecol. 28 (2019) 4737–4754, https://doi.org/10.1111/mec.15253.
[26] P.J. Bradbury, Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, E.S. Buckler, TASSEL: software for association mapping of complex traits in diverse samples, Bioinformatics 23 (2007) 2633–2635, https://doi.org/10.1093/bioinformatics/btm308.
[27] P. Danecek, A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R. E. Handsaker, G. Lunter, G.T. Marth, S.T. Sherry, G. McVean, R. Durbin, 1000 Genomes Project Analysis Group, The variant call format and VCFtools, Bioinformatics 27 (2011) 2156–2158, https://doi.org/10.1093/bioinformatics/btr330.
[28] M. Rovira, J.F. Hermoso, J. Rufat, V. Cristofori, C. Silvestri, A. Romero, Agronomical and physiological behavior of Spanish hazelnut selection "Negret-N9" grafted on non-suckering rootstocks, Front. Plant Sci. 12 (2022) 813902, https://doi.org/10.3389/fpls.2021.813902.
[29] B.G. Peterson, P. Carl, PerformanceAnalytics: econometric tools for performance and risk analysis. https://doi.org/10.32614/CRAN.package.PerformanceAnalytics, 2007.
[30] R.A. Peterson, Finding optimal normalizing transformations via bestNormalize, The R Journal 13 (2021) 310, https://doi.org/10.32614/RJ-2021-041.
[31] J. Wang, Z. Zhang, GAPIT version 3: boosting power and accuracy for genomic association and prediction, Genom. Proteom. Bioinform. 19 (2021) 629–640, https://doi.org/10.1016/j.gpb.2021.08.005.
[32] X. Liu, M. Huang, B. Fan, E.S. Buckler, Z. Zhang, Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies, PLoS Genet. 12 (2016) e1005767, https://doi.org/10.1371/journal.pgen.1005767.
[33] C.C. Chang, C.C. Chow, L.C. Tellier, S. Vattikuti, S.M. Purcell, J.J. Lee, Second-generation PLINK: rising to the challenge of larger and richer datasets, GigaScience 4 (2015), https://doi.org/10.1186/s13742-015-0047-8 s13742-015-0047–8.
[34] W.G. Hill, B.S. Weir, Variances and covariances of squared linkage disequilibria in finite populations, Theor. Popul. Biol. 33 (1988) 54–78, https://doi.org/10.1016/0040-5809(88)90004-4.

[35] F. Marroni, S. Pinosio, G. Zaina, F. Fogolari, N. Felice, F. Cattonaro, M. Morgante, Nucleotide diversity and linkage disequilibrium in Populus nigra cinnamyl alcohol dehydrogenase (CAD4) gene, Tree Genet. Genomes 7 (2011) 1011–1023, https://doi.org/10.1007/s11295-011-0391-5.
[36] A. Alexa, J. Rahnenführer, topGO. https://doi.org/10.18129/B9.BIOC.TOPGO, 2017.
[37] C. Chen, Y. Wu, J. Li, X. Wang, Z. Zeng, J. Xu, Y. Liu, J. Feng, H. Chen, Y. He, R. Xia, TBtools-II: a "one for all, all for one" bioinformatics platform for biological big-data mining, Mol. Plant 16 (2023) 1733–1742, https://doi.org/10.1016/j.molp.2023.09.010.
[38] V. Thumuluri, J.J. Almagro Armenteros, A.R. Johansen, H. Nielsen, O. Winther, DeepLoc 2.0: multi-label subcellular localization prediction using protein language models, Nucleic Acids Res. 50 (2022) W228–W234, https://doi.org/10.1093/nar/gkac278.
[39] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797, https://doi.org/10.1093/nar/gkh340.
[40] A.M. Kozlov, D. Darriba, T. Flouri, B. Morel, A. Stamatakis, RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference, Bioinformatics 35 (2019) 4453–4455, https://doi.org/10.1093/bioinformatics/btz305.
[41] I. Letunic, P. Bork, Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool, Nucleic Acids Res. 52 (2024) W78–W82, https://doi.org/10.1093/nar/gkae268.
[42] A. Marrano, M. Britton, P.A. Zaini, A.V. Zimin, R.E. Workman, D. Puiu, L. Bianco, E.A.D. Pierro, B.J. Allen, S. Chakraborty, M. Troggio, C.A. Leslie, W. Timp, A. Dandekar, S.L. Salzberg, D.B. Neale, High-quality chromosome-scale assembly of the walnut (*Juglans regia* L.) reference genome, GigaScience 9 (2020) giaa050, https://doi.org/10.1093/gigascience/giaa050.
[43] V.L. Sork, S.T. Fitz-Gibbon, D. Puiu, M. Crepeau, P.F. Gugger, R. Sherman, K. Stevens, C.H. Langley, M. Pellegrini, S.L. Salzberg, First draft assembly and annotation of the genome of a California endemic oak *Quercus lobata* née (Fagaceae), G3 Genes|Genomes|Genetics 6 (2016) 3485–3495, https://doi.org/10.1534/g3.116.030411.
[44] L. Reiser, E. Bakker, S. Subramaniam, X. Chen, S. Sawant, K. Khosa, T. Prithvi, T. Z. Berardini, The arabidopsis information resource in 2024, Genetics 227 (2024) iyae027, https://doi.org/10.1093/genetics/iyae027.
[45] L. Bianco, P. Fontana, A. Marchesini, S. Torre, M. Moser, S. Piazza, S. Alessandri, V. Pavese, P. Pollegioni, C. Vernesi, M. Malnoy, D. Torello Marinoni, S. Murolo, L. Dondini, C. Mattioni, R. Botta, F. Sebastiani, D. Micheletti, L. Palmieri, The de novo, chromosome-level genome assembly of the sweet chestnut (Castanea sativa Mill.) Cv. Marrone Di Chiusa Pesio, BMC Genom. Data 25 (2024) 64, https://doi.org/10.1186/s12863-024-01245-7.
[46] M.J.M. Christenhusz, Z. Goodwin, D.G. Bell, C.A. Martin, Royal botanic gardens kew genome acquisition lab, royal botanic garden Edinburgh genome acquisition lab, Darwin tree of life barcoding collective, plant genome sizing collective, wellcome sanger institute tree of life management, samples and laboratory team, wellcome sanger institute scientific operations: sequencing operations, wellcome sanger institute tree of life core informatics team, tree of life core informatics collective, Darwin tree of life consortium, Genome Sequence Common Alder Alnus glutinosa (L.) Gaertn. (Betulaceae). Wellcome Open Res 9 (2024) 570, https://doi.org/10.12688/wellcomeopenres.23137.1.
[47] J.T. Lovell, N.B. Bentley, G. Bhattarai, J.W. Jenkins, A. Sreedasyam, Y. Alarcon, C. Bock, L.B. Boston, J. Carlson, K. Cervantes, K. Clermont, S. Duke, N. Krom, K. Kubenka, S. Mamidi, C.P. Mattison, M.J. Monteros, C. Pisani, C. Plott, S. Rajasekar, H.S. Rhein, C. Rohla, M. Song, R.St Hilaire, S. Shu, L. Wells, J. Webber, R.J. Heerema, P.E. Klein, P. Conner, X. Wang, L.J. Grauke, J. Grimwood, J. Schmutz, J.J. Randall, Four chromosome scale genomes and a pan-genome annotation to accelerate pecan tree breeding, Nat. Commun. 12 (2021) 4125, https://doi.org/10.1038/s41467-021-24328-w.
[48] Jacques Dainat, Darío Hereñú, Dr K.D. Murray, Ed Davis, Ivan Ugrin, Kathryn Crouch, Lucile Sol, Nuno Agostinho, pascal-git, Zachary Zollman, tayyrov, NBISweden/AGAT: AGAT-v1.4.1. https://doi.org/10.5281/ZENODO.13799920, 2024.
[49] Y. Wang, H. Tang, J.D. DeBarry, X. Tan, J. Li, X. Wang, T. -h. Lee, H. Jin, B. Marler, H. Guo, J.C. Kissinger, A.H. Paterson, MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity, Nucleic Acids Res. 40 (2012), https://doi.org/10.1093/nar/gkr1293 e49–e49.
[50] N. Van Der Velden, Geneviewer: Gene Cluster Visualizations, 2025, https://doi.org/10.32614/CRAN.package.geneviewer.
[51] Q. Bonenfant, L. Noé, H. Touzet, Porechop_ABI: discovering unknown adapters in Oxford nanopore technology sequencing reads for downstream trimming, Bioinform. Adv. 3 (2023) vbac085, https://doi.org/10.1093/bioadv/vbac085.
[52] J. Schuster, M.E. Ritchie, Q. Gouil, Restrander: rapid orientation and artefact removal for long-read cDNA data, NAR Genom. Bioinf. 5 (2023) lqad108, https://doi.org/10.1093/nargab/lqad108.
[53] W. De Coster, S. D'Hert, D.T. Schultz, M. Cruts, C. Van Broeckhoven, NanoPack: visualizing and processing long-read sequencing data, Bioinformatics 34 (2018) 2666–2669, https://doi.org/10.1093/bioinformatics/bty149.
[54] H. Li, Minimap2: pairwise alignment for nucleotide sequences, Bioinformatics 34 (2018) 3094–3100, https://doi.org/10.1093/bioinformatics/bty191.
[55] J. Gleeson, A. Leger, Y.D.J. Prawer, T.A. Lane, P.J. Harrison, W. Haerty, M.B. Clark, Accurate expression quantification from nanopore direct RNA sequencing with NanoCount, Nucleic Acids Res. 50 (2022), https://doi.org/10.1093/nar/gkab1129 e19–e19.
[56] R. Kolde, Pheatmap: Pretty Heatmaps, 2010, https://doi.org/10.32614/CRAN.package.pheatmap.
[57] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550, https://doi.org/10.1186/s13059-014-0550-8.
[58] M. Stephens, False discovery rates: a new deal, Biostat kxw041 (2016), https://doi.org/10.1093/biostatistics/kxw041.
[59] R Core Team, _R: a Language and Environment for Statistical Computing_, R Foundation for Statistical Computing, Vienna, Austria, 2023. https://www.R-project.org/.
[60] J. Shanklin, C. Somerville, Stearoyl-acyl-carrier-protein desaturase from higher plants is structurally unrelated to the animal and fungal homologs, Proc. Natl. Acad. Sci. USA 88 (1991) 2510–2514, https://doi.org/10.1073/pnas.88.6.2510.
[61] M.L. Hernández, M.D. Sicardo, M. Alfonso, J.M. Martínez-Rivas, Transcriptional regulation of stearoyl-acyl carrier protein desaturase genes in response to abiotic stresses leads to changes in the unsaturated fatty acids composition of olive mesocarp, Front. Plant Sci. 10 (2019) 251, https://doi.org/10.3389/fpls.2019.00251.
[62] J. Chen, J. Gao, Lingling Zhang, Lin Zhang, Tung tree stearoyl-acyl carrier protein Δ9 desaturase improves oil content and cold resistance of arabidopsis and Saccharomyces cerevisiae, Front. Plant Sci. 14 (2023) 1144853, https://doi.org/10.3389/fpls.2023.1144853.
[63] L. De Jaeger, J. Springer, E.J.H. Wolbert, D.E. Martens, G. Eggink, R.H. Wijffels, Gene silencing of stearoyl-ACP desaturase enhances the stearic acid content in Chlamydomonas reinhardtii, Bioresour. Technol. 245 (2017) 1616–1626, https://doi.org/10.1016/j.biortech.2017.06.128.
[64] P. Zhang, J.W. Burton, R.G. Upchurch, E. Whittle, J. Shanklin, R.E. Dewey, Mutations in a Δ^9 –Stearoyl-ACP-Desaturase gene are associated with enhanced stearic acid levels in soybean seeds, Crop Sci. 48 (2008) 2305–2313, https://doi.org/10.2135/cropsci2008.02.0084.
[65] M. Batsale, D. Bahammou, L. Fouillen, S. Mongrand, J. Joubès, F. Domergue, Biosynthesis and functions of very-long-chain fatty acids in the responses of plants to abiotic and biotic stresses, Cells 10 (2021) 1284, https://doi.org/10.3390/cells10061284.
[66] K. Cho, C.M. O'Neill, S.-J. Kwon, T.-J. Yang, A.M. Smooker, F. Fraser, I. Bancroft, Sequence-level comparative analysis of the *Brassica napus* genome around two stearoyl-ACP desaturase loci, Plant J. 61 (2010) 591–599, https://doi.org/10.1111/j.1365-313X.2009.04084.x.
[67] Z. Zou, X. Fu, C. Li, X. Yi, J. Huang, Y. Zhao, Insights into the stearoyl-acyl carrier protein desaturase (SAD) family in tigernut (Cyperus esculentus L.), an oil-bearing tuber plant, Plants 14 (2025) 584, https://doi.org/10.3390/plants14040584.
[68] Z. Hajiahmadi, A. Abedi, H. Wei, W. Sun, H. Ruan, Q. Zhuge, A. Movahedi, Identification, evolution, expression, and docking studies of fatty acid desaturase genes in wheat (Triticum aestivum L.), BMC Genom. 21 (2020) 778, https://doi.org/10.1186/s12864-020-07199-1.
[69] T. Zare, J.F. Paril, E.M. Barnett, P. Kaur, R. Appels, B. Ebert, U. Roessner, A. Fournier-Level, Comparative genomics points to tandem duplications of *SAD* gene clusters as drivers of increased α-linolenic (ω-3) content in *S. hispanica* seeds, Plant Genome 17 (2024) e20430, https://doi.org/10.1002/tpg2.20430.
[70] Hui Huang, J. Liang, Q. Tan, L. Ou, X. Li, C. Zhong, Huilin Huang, I.M. Møller, X. Wu, S. Song, Insights into triterpene synthesis and unsaturated fatty-acid accumulation provided by chromosomal-level genome analysis of Akebia trifoliata subsp, Australis. Hortic Res. 8 (2021) 33, https://doi.org/10.1038/s41438-020-00458-y.
[71] Y.-Y. Yang, G.W. Stull, X.-J. Qu, L. Zhao, Y. Hu, Z.-H. Wang, H. Ma, D.-Z. Li, S. A. Smith, T.-S. Yi, Genome duplications, genomic conflict, and rapid phenotypic evolution characterize the Cretaceous radiation of Fagales. https://doi.org/10.1101/2023.06.11.544004, 2023.
[72] L.S. Maguire, S.M. O'Sullivan, K. Galvin, T.P. O'Connor, N.M. O'Brien, Fatty acid profile, tocopherol, squalene and phytosterol content of walnuts, almonds, peanuts, hazelnuts and the macadamia nut, Int. J. Food Sci. Nutr. 55 (2004) 171–178, https://doi.org/10.1080/09637480410001725175.
[73] E. Ryan, K. Galvin, T.P. O'Connor, A.R. Maguire, N.M. O'Brien, Fatty acid profile, tocopherol, squalene and phytosterol content of Brazil, pecan, pine, pistachio and cashew nuts, Int. J. Food Sci. Nutr. 57 (2006) 219–228, https://doi.org/10.1080/09637480600768077.
[74] S.M. Ghazani, A.G. Marangoni, Healthy fats and oils, in: Reference Module in Food Science, Elsevier, 2016 B9780081005965001001, https://doi.org/10.1016/B978-0-08-100596-5.00100-1.
[75] J. Sun, P. Hu, C. Lyu, J. Tian, X. Meng, H. Tan, W. Dong, Comprehensive lipidomics analysis of the lipids in hazelnut oil during storage, Food Chem. 378 (2022) 132050, https://doi.org/10.1016/j.foodchem.2022.132050.
[76] M. Taib, L. Bouyazza, B. Lyoussi, Acorn oil: chemistry and functionality, J. Food Qual. (2020) 1–11, https://doi.org/10.1155/2020/8898370.
[77] H. Goodarzi, D. Hassani, L. Pourhosseini, S. Kalantari, A. Lashgari, Total lipid and fatty acids components of some Persian walnut (Juglans regia) cultivars, Sci. Hortic. 321 (2023) 112252, https://doi.org/10.1016/j.scienta.2023.112252.
[78] O.P. Borges, J. Soeiro Carvalho, P. Reis Correia, A. Paula Silva, Lipid and fatty acid profiles of Castanea sativa Mill. Chestnuts of 17 native Portuguese cultivars, J. Food Compos. Anal. 20 (2007) 80–89, https://doi.org/10.1016/j.jfca.2006.07.008.
[79] C.T. Ciucure, E.-I. Geana, C. Sandru, O. Tita, M. Botu, Phytochemical and nutritional profile composition in fruits of different sweet chestnut (Castanea sativa Mill.) cultivars grown in Romania, Separations 9 (2022) 66, https://doi.org/10.3390/separations9030066.

[80] J. Kim, S.-H. Shiu, S. Thoma, W.-H. Li, S.E. Patterson, Patterns of expansion and expression divergence in the plant polygalacturonase gene family, Genome Biol. 7 (2006) R87, https://doi.org/10.1186/gb-2006-7-9-r87.
[81] S. Kazaz, G. Barthole, F. Domergue, H. Ettaki, A. To, D. Vasselon, D. De Vos, K. Belcram, L. Lepiniec, S. Baud, Differential activation of partially redundant Δ9 Stearoyl-ACP desaturase genes is critical for Omega-9 monounsaturated fatty acid biosynthesis during seed development in arabidopsis, Plant Cell 32 (2020) 3613–3637, https://doi.org/10.1105/tpc.20.00554.
[82] M.A. Troncoso-Ponce, A. Kilaru, X. Cao, T.P. Durrett, J. Fan, J.K. Jensen, N. A. Thrower, M. Pauly, C. Wilkerson, J.B. Ohlrogge, Comparative deep transcriptional profiling of four developing oilseeds, Plant J. 68 (2011) 1014–1027, https://doi.org/10.1111/j.1365-313X.2011.04751.x.
[83] W. Shi, D. Zhang, Z. Ma, Transcriptome analysis of genes involved in fatty acid and lipid biosynthesis in developing walnut (Juglans regia L.) seed kernels from Qinghai Plateau, Plants 11 (2022) 3207, https://doi.org/10.3390/plants11233207.
[84] J. Huang, T. Zhang, Q. Zhang, M. Chen, Z. Wang, B. Zheng, G. Xia, X. Yang, C. Huang, Y. Huang, The mechanism of high contents of oil and oleic acid revealed by transcriptomic and lipidomic analysis during embryogenesis in Carya cathayensis Sarg, BMC Genom. 17 (2016) 113, https://doi.org/10.1186/s12864-016-2434-7.
[85] T. Unver, Z. Wu, L. Sterck, M. Turktas, R. Lohaus, Z. Li, M. Yang, L. He, T. Deng, F. J. Escalante, C. Llorens, F.J. Roig, I. Parmaksiz, E. Dundar, F. Xie, B. Zhang, A. Ipek, S. Uranbey, M. Erayman, E. Ilhan, O. Badad, H. Ghazal, D.A. Lightfoot, P. Kasarla, V. Colantonio, H. Tombuloglu, P. Hernandez, N. Mete, O. Cetin, M. Van Montagu, H. Yang, Q. Gao, G. Dorado, Y. Van De Peer, Genome of wild olive and the evolution of oil biosynthesis, Proc. Natl. Acad. Sci. USA 114 (2017), https://doi.org/10.1073/pnas.1708621114.
[86] G. Xiao, Z.Q. Zhang, C.F. Yin, R.Y. Liu, X.M. Wu, T.L. Tan, S.Y. Chen, C.M. Lu, C. Y. Guan, Characterization of the promoter and 5′-UTR intron of oleic acid desaturase (FAD2) gene in Brassica napus, Gene 545 (2014) 45–55, https://doi.org/10.1016/j.gene.2014.05.008.
[87] A. Salimonti, F. Carbone, E. Romano, M. Pellegrino, C. Benincasa, S. Micali, A. Tondelli, F.L. Conforti, E. Perri, A. Ienco, S. Zelasco, Association study of the 5′UTR intron of the FAD2-2 gene with oleic and linoleic acid content in Olea europaea L, Front. Plant Sci. 11 (2020) 66, https://doi.org/10.3389/fpls.2020.00066.
[88] Á. Soria-García, M.C. Rubio, B. Lagunas, S. López-Gomollón, M.D.L.Á. Luján, R. Díaz-Guerra, R. Picorel, M. Alfonso, Tissue distribution and specific contribution of arabidopsis FAD7 and FAD8 plastid desaturases to the JA- and ABA-mediated cold stress or defense responses, Plant Cell Physiol. 60 (2019) 1025–1040, https://doi.org/10.1093/pcp/pcz017.
[89] J. Shi, Y. Cao, X. Fan, M. Li, Y. Wang, F. Ming, A rice microsomal delta-12 fatty acid desaturase can enhance resistance to cold stress in yeast and Oryza sativa, Mol. Breed. 29 (2012) 743–757, https://doi.org/10.1007/s11032-011-9587-5.
[90] X. Zhao, J. Wei, L. He, Y. Zhang, Y. Zhao, X. Xu, Y. Wei, S. Ge, D. Ding, M. Liu, S. Gao, J. Xu, Identification of fatty acid desaturases in maize and their differential responses to low and high temperature, Genes 10 (2019) 445, https://doi.org/10.3390/genes10060445.
[91] A. Pacchiarelli, M. Lupo, A. Ferrucci, F. Giovanelli, S. Priori, A.L. Pica, C. Silvestri, V. Cristofori, Phenology, yield and nut traits evaluation of twelve European hazelnut cultivars grown in central Italy, Forests 15 (2024) 833, https://doi.org/10.3390/f15050833.