BMC Plant Biology

RESEARCH

Open Access

Genome-wide identification of the R2R3-MYB gene family in olive and its association with fatty acid biosynthesis

Hengxing Zhu[1†], Qianli Dai[1†], Feiyi Huang[1], Zhuo Wei[2], Min Lu[1], Chenggong Lei[1], Ximeng Yang[1] and Benwen Chen[3*]

Abstract

MYB transcription factors play an important role in the biosynthesis of fatty acids in plants. Olive (*Olea europaea* L.) is one of the woody plants that has been used the longest for the production of oil. However, their functions in fatty acid metabolism of Olive fruit is not well defined. This study identified 212 *OLR2R3 MYB* genes from the olive genome, which were unevenly distributed across 23 chromosomes. A phylogenetic analysis revealed that they are clustered into six subgroups, and the *OLR2R3 MYB* gene members in different subgroups have similar gene structures and conserved motifs. A collinearity analysis revealed 17 pairs of segmentally duplicated *OLR2R3 MYB* genes. The levels of expression of each *OLR2R3 MYB* gene in the fruit, flower, and bud tissues from various varieties of olive were analyzed. This revealed that 50 genes may be involved in the development of olive fruit. A co-expression network analysis showed that *OLMYB185* is co-expressed with the genes for the biosynthesis of fatty acids. The overexpression of *OLMYB185* in Arabidopsis seeds significantly increased the seed size compared to the wild type. These results further elucidate the role of *OLR2R3 MYB* in the biosynthesis of fatty acids in olive fruit and provide new insights into the function of these genes in olive plants.

Keywords R2R3-MYB, Gene family, Fatty acid, Biosynthesis, Olive

Introduction

MYB is the largest family of transcription factors (TFs) in plants. The MYB proteins contain two domains, including a DNA-binding domain (MYB domain) and a domain responsible for the regulation of protein activity [1]. The MYB domain typically consists of up to four imperfect amino acid sequence repeats (R) with each approximately 52 amino acids long that forms three α-helices [2]. Based on the number of repeats (R) in the MYB domain, the MYB family can be divided into different classes, including 1R-MYB, R2R3-MYB, 3R-MYB, and 4R-MYB. The first functional MYB gene discovered in plants was *ZmMYBC1*, which is involved in the regulation of anthocyanin biosynthesis in *Zea mays* [3]. With the extensive reporting of plant genome information, R2R3-MYB TFs in various plants have been found to be involved in the responses to various physiological environments and regulating their own development [4, 5].

Triacylglycerol (TAG) is the primary storage form of lipids in plants and directly affects the oil content of fruits. TAG is synthesized from fatty acids that are produced in

†Hengxing Zhu and Qianli Dai co-first author.

*Correspondence:
Benwen Chen
455503125@qq.com
[1] Chongqing Key Laboratory of Forest Ecological Restoration and Utilization in the Three Gorges Reservoir Area, Chongqing Academy of Forestry, Chongqing 400036, China
[2] School of Ecology and Environmental Sciences, Yunnan University, Kunming 650091, China
[3] Observation and Research Station of Chongqing Mountain City Forest Ecosystem, Chongqing 400036, China

the plastid and then transported to the endoplasmic reticulum (ER) for completion [6, 7]. In many plants, MYB TFs mediate the biosynthetic pathway for fatty acids. Study of the metabolism in Arabidopsis revealed that *MYB5* increases the accumulation of fatty acids in seeds by upregulating the expression of the negative fatty acid regulator TT8 [8]. Arabidopsis *MYB30* is involved in the regulation of long-chain fatty acid synthesis [9]. Moreover, some MYBs can indirectly affect the content of important TAG components, such as fatty acids. *AtMYB107* promotes the transcription of fatty acyl-CoA reductase (FAR), which facilitates the degradation of fatty acids by this enzyme, while *AtMYB92* promotes the biosynthesis of fatty acids [10]. *CzMYB1*, a conserved MYB transcription factor in green algae, plays an important role in the lipid metabolic pathway that regulates TAG accumulation [8]. The MYB gene family in walnut genome was identified and the candidate genes of *JrMYB116*, *JrMYB118*, *JrMYB119* and *JrMYB121* involved in fatty acid biosynthesis were screened [35]. In addition to providing energy for the growth and development of plants, fatty acids also participate in the process of the maturation of seeds. For example, Arabidopsis WRI1 recruits BLI to the AW boxes of the promoters of genes related to fatty acid biosynthesis to regulate the biosynthesis of fatty acids, and the knockout of BLI results in significantly smaller Arabidopsis seeds compared to the wild type [11].

Olive (*Olea europaea* L.) is a member of the Oleaceae family and an important economic crop for its fruit. The extra virgin olive oil extracted from olive contains approximately 98%-99% fatty acids, particularly monounsaturated fatty acids [12, 13]. Despite the large area of olive cultivated around the world, many varieties are grown, and the yield of oil from the olives varies substantially in different regions. Research on the involvement of MYB TFs in the biosynthesis and metabolism of fatty acids in olive is still limited and merits further study. In this study, we first identified the R2R3 MYB TF family members of olive and analyzed their chromosomal distribution, homology, and evolutionary relationships. We also analyzed the profiles of gene expression in the fruits of different varieties of olives, constructed a co-expression network between the R2R3 MYB TFs and the genes for the biosynthesis of fatty acids, screened for related R2R3 MYB TFs, and confirmed their function in *Arabidopsis*. These results provide insights into the biosynthesis and metabolism of fatty acids in olive.

Methods

Identification of the members and physicochemical properties of MYB Genes

The genomes and gene annotation files of olive, wild olive, sesame (*Sesamum indica* L.), and soybean (*Glycine max* L.) were downloaded from the NCBI database, while the relevant data for *Arabidopsis* were downloaded from TAIR [14]. *Arabidopsis* MYB genes were used as reference genes to preliminarily screen the olive genome using an HMM search to identify possible MYB TFs. The presence of conserved MYB domains was further verified using the Batch CDD-search database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) [15]. Expasy (https://www.expasy.org/) [16] was used to calculate and statistically analyze the biochemical indicators of the candidate olive MYB TF proteins, including their isoelectric point, molecular weight, hydrophilicity average coefficient, and protein instability coefficient.

Phylogenetic Analysis of the R2R3-MYB Gene Family Members

Based on the screened olive MYB genes and *Arabidopsis* MYB genes, a multiple sequence alignment was performed using Clustal X. The screened conserved region was used to construct a phylogenetic tree using MEGA 6.0 with the Neighbor-Joining (NJ) method, and the bootstrap reliability analysis was set to 1,000 iterations.

Gene structure, conserved motif, and domain analysis of the R2R3-MYB gene family members

Gene structure information is extracted from genome annotation files and visualized in TBtools [17]. The conserved motif and domain were analyzed using the MEME online tool and InterPro (https://www.ebi.ac.uk/interpro/) [18], and the number of Motifs was set to 10, Visualization through the gene structure module in TBtools.

Collinearity analysis

McscanX tool of TBtools is used for collinearity analysis, statistic of Collinearity gene pair and tandem gene pair, and the results were visualized using TBtools. To determine the evolutionary pressure of gene pairs in *OLR2R3 MYB* gene family, KaKs_Calculator was used to calculate the ratio of nonsynonymous substitution to synonymous substitution of gene pairs (Ka/Ks).

Analysis of the Expression of the R2R3-MYB Gene Family Members

The transcriptome data for different tissues of olive were downloaded from the NCBI SRA database (https://www.ncbi.nlm.nih.gov/sra) [19]. The t raw data is filtered through Trimmomatic(trimmomatic-0.39). hisat2(version = 2.2.0) was used to index the olive genome, and transcriptome data were compared on the index to obtain expression matrix. DEGseq was used to screen differential genes. The differential expression of OLMYB genes in the transcriptome data was screened

using NovaSeq (Illumina) RNA-seq. Log2 (RPKM + 1) > 5 in any sample transcriptome were screened as differential genes.

WGCNA Screening of the Genes

All the differentially expressed genes (DEGs) were standardized using log2 transformation. Based on these values, a gene co-expression network was constructed using a Weighted Gene Co-Expression Network Analysis (WGCNA) [20]. The adjacency matrix was calculated using the soft threshold, with the soft threshold power selection criterion based on scale-free topology (fit index = 0.9). The adjacency matrix was further transformed into a Topological Overlap Matrix (TOM) to reduce the interference from noise. The gene modules were determined using a fast greedy modular optimization algorithm.

Vector Construction and Transformation of Arabidopsis

The OLMYB185 sequence was cloned into the pCAMBIA1300 plasmid and driven by the CaMV 35S promoter. The constructed vector was transformed into *Agrobacterium tumefaciens* GV3101 and used to transform the wild-type (WT) *Arabidopsis* Col-0 using the floral dip method.

Measurement of the Size of the Arabidopsis Seeds and Other Indicators

The seed size was determined by measuring the length and width of 50 seeds from the WT Col-0 *Arabidopsis* using ImageJ (NIH, Bethesda, MD, USA). The volume was estimated using the ellipsoid formula (volume = 4/3 × π x length x width x depth).

Statistical Analysis

The statistical significance of various comparisons was analyzed using a one-way analysis of variance (ANOVA) or multiple paired *t*-tests with Origin 7.5 (OriginLab, Northampton, MA, USA).

Results

In this study, we identified 212 OLR2R3 MYB genes from the olive genome and named them *OLMYB1-OLMYB212* based on their chromosomal locations (Supplementary Table 1). A protein analysis revealed that all the OLR2R3 MYBs contain two highly conserved MYB binding domains. The lengths of the OLR2R3 MYB proteins in each gene family vary widely and range from 157 (*OLMYB162*) to 941 (*OLMYB179*) amino acids. Their molecular weights range from 60,198.18 (*OLMYB27*) to 18,001.37 (*OLMYB162*) and their isoelectric points from 10.27 (*OLMYB201*) to 4.78 (*OLMYB57*). Among them, 96 are hydrophobic proteins (PI > 7). All the OLR2R3-MYBs are localized to the nucleus (Fig. 1).

To analyze the developmental relationships and functions of the *OLR2R3 MYB* gene family members in olive, we constructed a phylogenetic tree using the amino acid sequences of the *OLR2R3 MYB* gene family members from olive and *Arabidopsis*. The 212 *OLR2R3 MYB* genes from olive can be divided into six subgroups (O1-O6) (Fig. 2). Combined with the 23 subfamilies in *Arabidopsis*, O1 includes subfamilies S2, S14, S9, S13, and S11; O2 corresponds to subfamilies S12, S21, and S22; O3 includes subfamilies S3, S18, and S24; O4 corresponds to subfamilies S16, S19, and S23; O5 corresponds to subfamilies S4, S6, S7, and S10; and O6 corresponds to subfamilies S20 and S24. No members of the *Arabidopsis*-specific subfamily S9 were identified. O2 was the largest subgroup with 37 family members (Fig. 2). Previous studies have shown that the *OLR2R3 MYB* genes within the same subgroup have similar functions. There is a potential correlation between the subfamily classification and functional similarity of the *OLR2R3 MYB* genes.

We also mapped the chromosomal locations of the 212 *OLR2R3 MYB* genes and found that only 110 of them are located on the chromosomes (Fig. 2). The remaining members could not be successfully mapped owing to the incomplete annotation of the genome. The 110 *OLR2R3 MYB* genes are unevenly distributed across 23 chromosomes. Chromosome 1 contains 10 members; chromosome 22 contains only one member, and chromosome 23 lacks *OLR2R3 MYB* genes.

The analysis of the 212 *OLR2R3 MYB* members included examining conserved motifs, domains, and gene structures. The conserved motif analysis showed that all the *OLR2R3 MYB* genes contain introns, with complex exon and intron distributions. This indicates that the diversity of *OLR2R3 MYB* is related to their mechanisms of formation and evolutionary processes (Fig. 3). Most members within the same subgroup share one or more identical motifs, and closely related members exhibit common motif compositions. This suggests that members of the same subgroup have similar functions. All the family members contain two highly conserved motifs: motif 1 and motif 2. Different groups have distinct conserved motifs, with motif8 only identified in group 2.

We analyzed the collinearity of the *OLR2R3 MYB* gene family and found that there are 7 pairs of tandem duplicated genes, such as *OLMYB63-OLMYB64* and *OLMYB90-OLMYB91*, and 17 pairs of segmentally duplicated genes (Fig. 4). These results indicate that the *OLR2R3 MYB* genes underwent chromosomal segmental duplications and tandem duplications, and segmental duplications occurred at greater distances on the chromosomes. To evaluate the evolutionary relationships of gene families, the selection pressure of *OLR2R3 MYB* gene pairs was examined by the non-synonymous

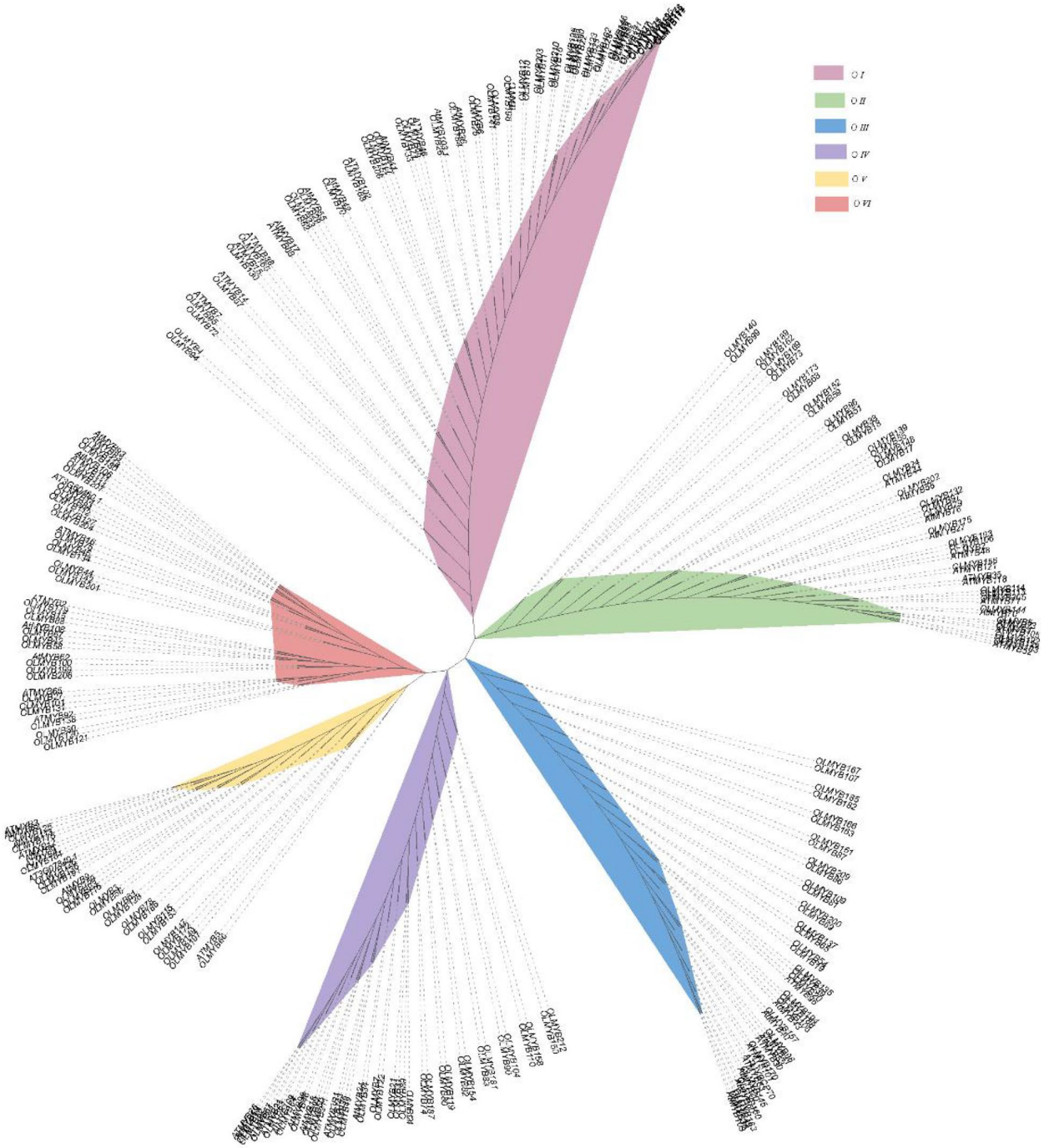


Fig. 1 Phylogenetic tree of olive and *Arabidopsis* R2R3 MYBs. Constructed using the Neighbor-Joining (NJ) method based on the protein sequences, with 1,000 bootstrap replicates

and synonymous substitution ratio (Ka/Ks). The results showed that the Ka/Ks ratio of *OLR2R3 MYB* duplicate gene pairs was less than 1, indicating that *OLR2R3 MYB* gene had undergone purifying selection.

We analyzed the promoter binding elements in the upstream 2000 bp promoter regions of the 212 *OLR2R3 MYB* genes (Fig. 5). A total of 11 types of abiotic and biotic stress response elements, 11 types of plant hormone response elements, and 18 types of plant growth and development-related promoter *cis*-elements were found. Among them, the largest number of elements were related to abiotic and biotic stress response, and 744

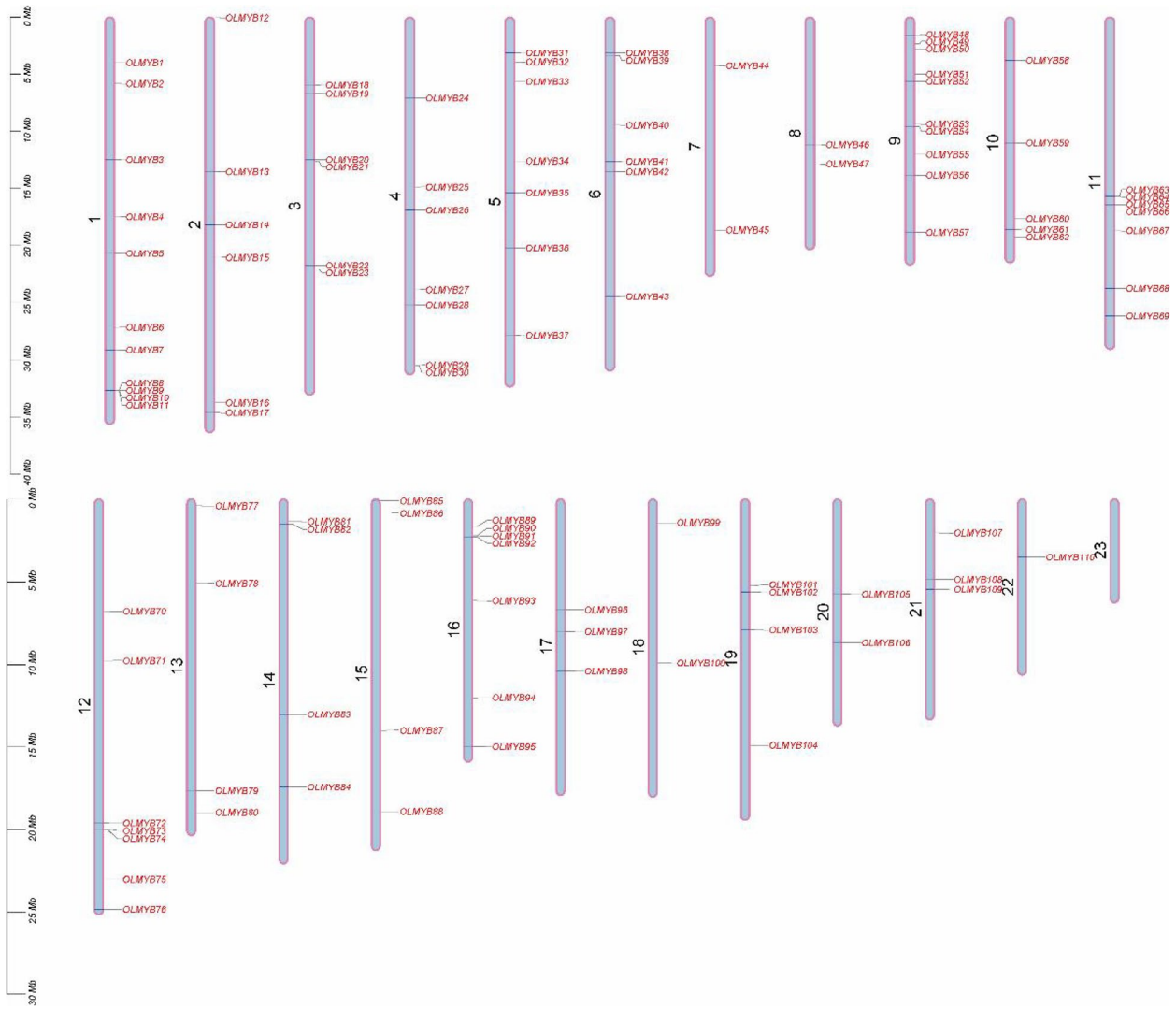


Fig. 2 Distribution of 110 OLMYB genes on 23 olive chromosomes. The left scale represents chromosome length. Scale:Mb

MYC response elements were the most abundant. Other elements are associated with low temperature, drought, and wounding. Of the response elements that are associated with plant growth and development, Box 4 (675) and G box (402) are associated with the light response; the CAT Box (77) is associated with the expression of meristems expression, and the RY-element (14) is associated with seed regulation. These were identified in 13 OLMYB genes. For the plant hormone response elements, ethylene-responsive element (ERE) (346) was the most abundant, while as-1 (21), which is associated with the response of seeds to gibberellin (GA), was the least common. Other response elements associated with gibberellin, salicylic acid, and auxin were also found. In summary, the OLMYB gene family is involved in regulation of the response of olive to various abiotic and biotic stresses, as well as its growth and development.

To further explore the evolutionary relationships of the *OLR2R3 MYB* genes, we examined the homologous evolutionary relationships among olive, wild olive, sesame, soybean, and *Arabidopsis*. The results showed that the olive *OLR2R3 MYB* genes are homologous with the genes from the other four plants, with 162 homologous gene pairs with soybean, 119 with sesame, 101 with wild olive, and 44 with *Arabidopsis* (Fig. 6). The results indicate that olive is more closely related to soybean, while wild olive (*Olea europaea* var europaea) has fewer homologous gene pairs compared to soybean and sesame. Interestingly, the homologous genes of these species with olive vary, which indicates that the MYB genes are not

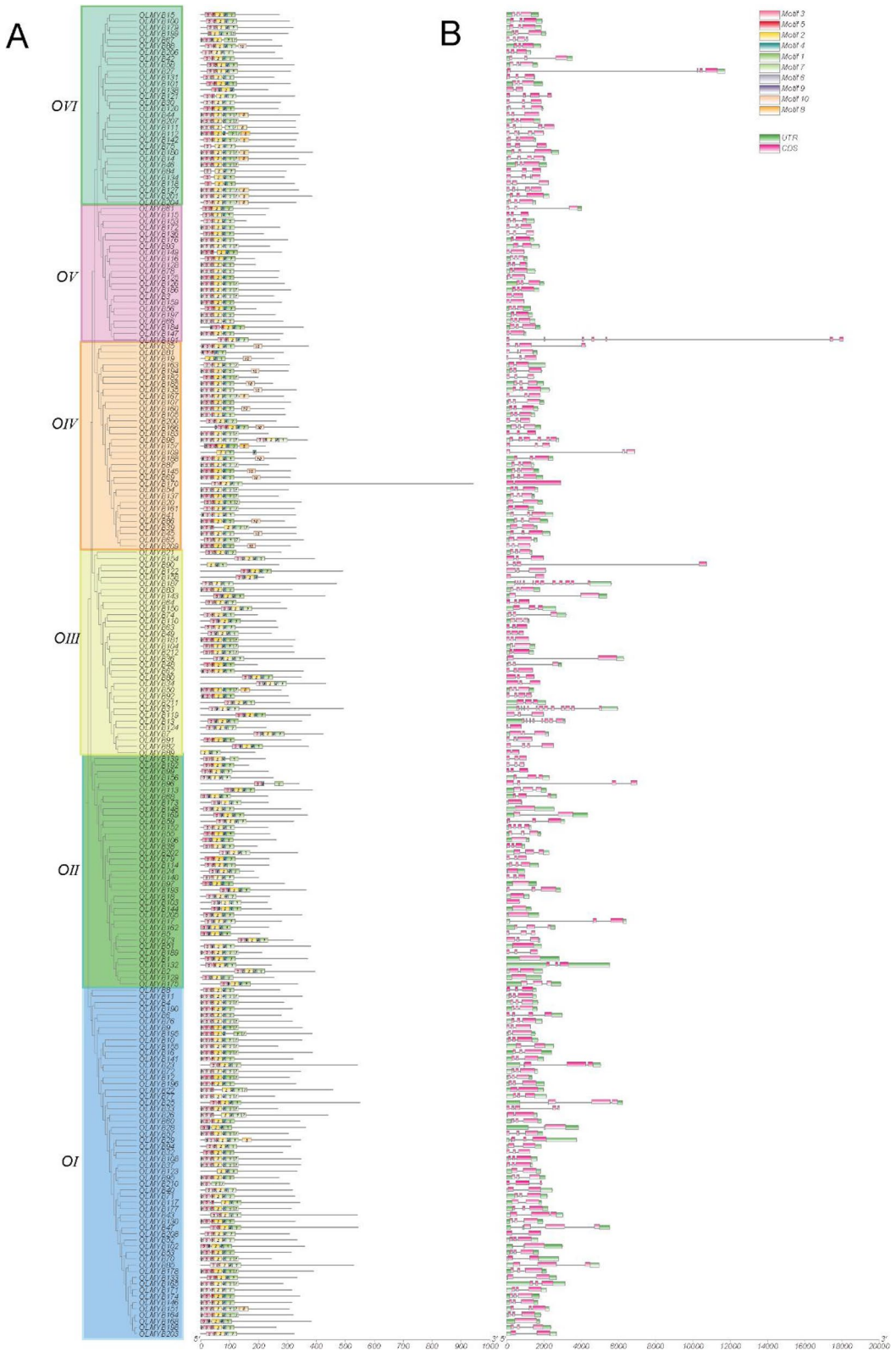


Fig. 3 Conserved motifs and gene structure of 212 OLMYB genes. **A** Conserved motifs in the *OLR2R3 MYB* protein sequences. Different colored boxes with numbers represent different conserved motifs. **B** Exon/intron/UTR regions of the *OLR2R3 MYB* genes. Exons are shown as pink boxes; introns as black lines, and UTRs as green boxes. UTR, untranslated region

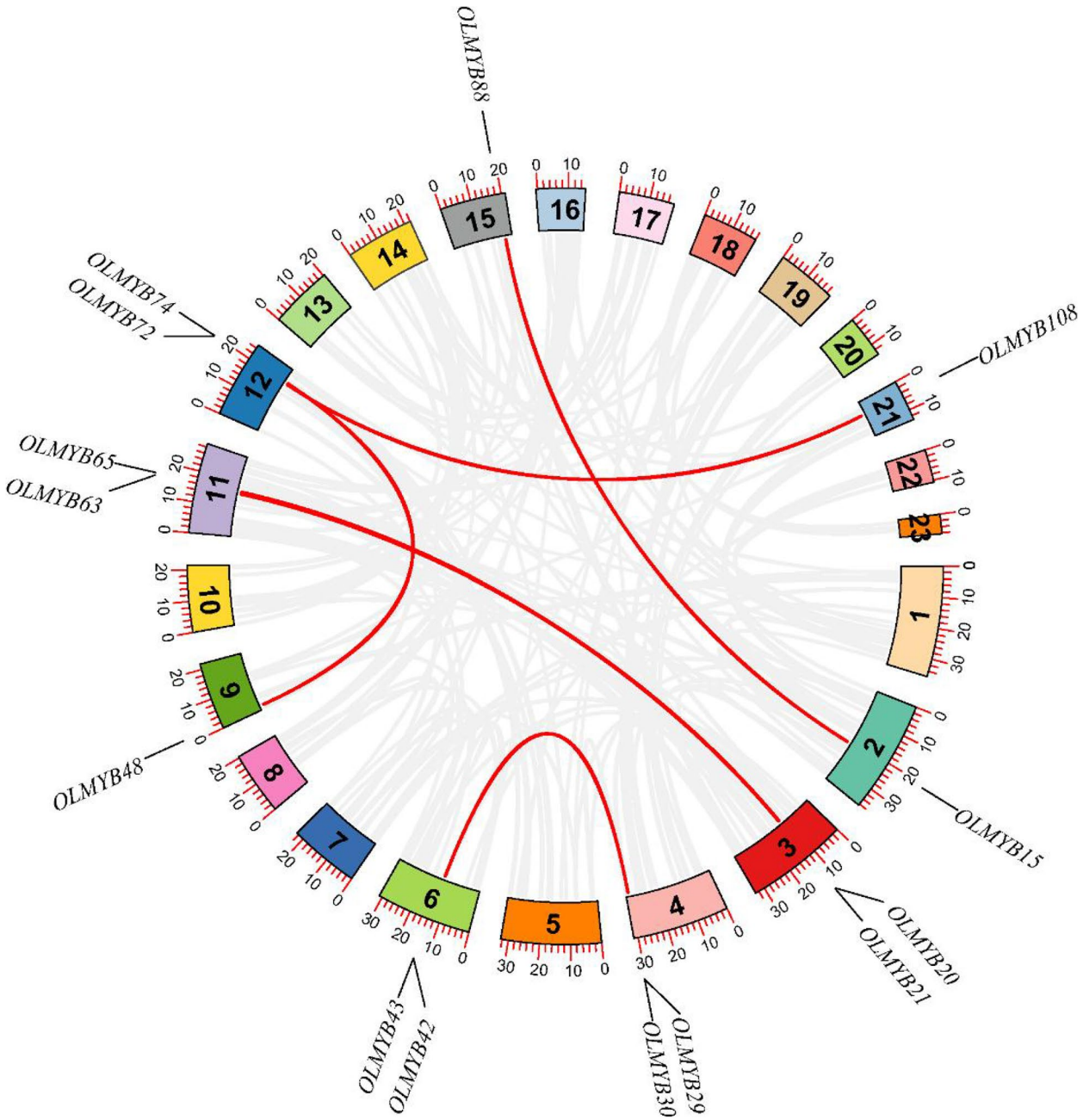
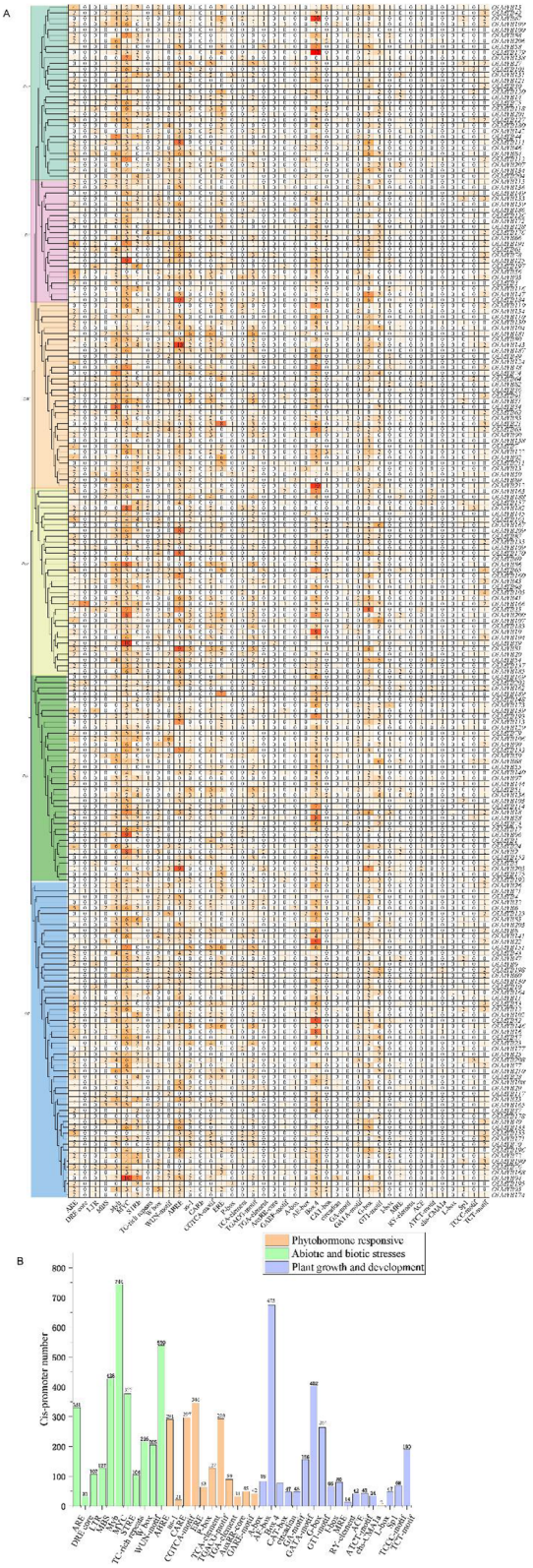


Fig. 4 Intraspecies homology of the OLR2R3 MYBs. Red lines, duplicated gene pairs of *OLR2R3 MYB*. Gray lines, whole genome duplicated gene pairs

conserved during evolution. This may be related to the functional divergence of these genes.

To understand the role of the *OLR2R3 MYB* genes in the growth and development of olive, we obtained transcriptome data on the differentiation of flower buds and fruit from three different cultivars of olive from the NCBI database. Based on the log2 (RPKM + 1) transformed values, we constructed heatmaps of the levels of expression of the *OLR2R3 MYB* genes in the three different olive. Orange represents high levels of expression, while blue represents low levels of expression. The results show that different *OLR2R3 MYB* members have varying levels of expression; *OLMYB185* and *OLMYB149* were highly expressed in the three olive cultivars but not in the flower buds (Fig. 7A). A PPI interaction analysis revealed the protein functional relationships of the homologous genes



◀ **Fig. 5** *Cis*-elements in the *OLMYB* promoters. **A** Different colors, different numbers of *cis*-elements. **B** Statistical chart of the different types of *cis*-elements in *OLR2R3 MYB*: orange, hormone response; green, biotic or abiotic stress response; purple, plant growth and development response

of *OLMYB185* in *Arabidopsis* and demonstrated that it is related to the fatty acid biosynthetic pathway, including *CER2* (ECERIFERUM 2), *FAR3* (fatty acyl-CoA reductase 3), and *KCR1* (3-oxoacyl-CoA reductase 1) (Fig. 7B).

WGCNA

We performed a WGCNA analysis on the transcriptome data and identified 18 modules, with the brown module positively correlated (0.98) with the content of olive oil. *OLMYB185* was also included in the brown module. Calculation of the correlation between *OLMYB185* and the other genes related to fatty acids in the module ($R^2 > 0.85$, $P < 0.05$) revealed that the patterns of expression of *OLMYB185* in the fruit positively correlated with *FAR1* and *ACAT2* (acetyl-CoA acetyltransferase 2) in the fatty acid biosynthetic pathway. This suggests that *OLMYB185* may be involved in the regulation of the biosynthesis of fatty acids in the fruit (Fig. 8). The Top 20 KEGG enrichment of the brown module gene showed that the pathway related to fatty acid biosynthesis was enriched (Fatty acid degradation, Fatty acid Synthesis, Synthesis of unsaturated fatty acids).

Transformation of Arabidopsis

In phylogenetic tree clustering results, we found that *OLMYB185* and *AtMYB94* belong to the O4 subgroup, and *AtMYB94* plays a crucial role in the accumulation of fatty acid biosynthesis. In addition, *OLMYB185* has six ABRE promoter *cis*-elements associated with the accumulation of fatty acid biosynthesis. Therefore, we used Agrobacterium-mediated transfer of *OLMYB185* into *Arabidopsis* to obtain transgenic *Arabidopsis* plants overexpressed with *OLMYB185*, and compared the seed size with that of the wild type of *Arabidopsis* (WT Col-0).

The overexpression of *OLMYB185* in *Arabidopsis* resulted in eight lines. Measurements of the various indicators of the S3 generation *Arabidopsis* seeds showed that the seed surface area and size of the *OLMYB185* overexpression lines were significantly higher than those of the WT Col-0 (Fig. 9).

Discussion

Olive is a type of fruit-bearing tree that yields economically valuable oil. The whole genome sequencing and annotation of regular olive and its wild counterpart have been completed, which enabled the widespread

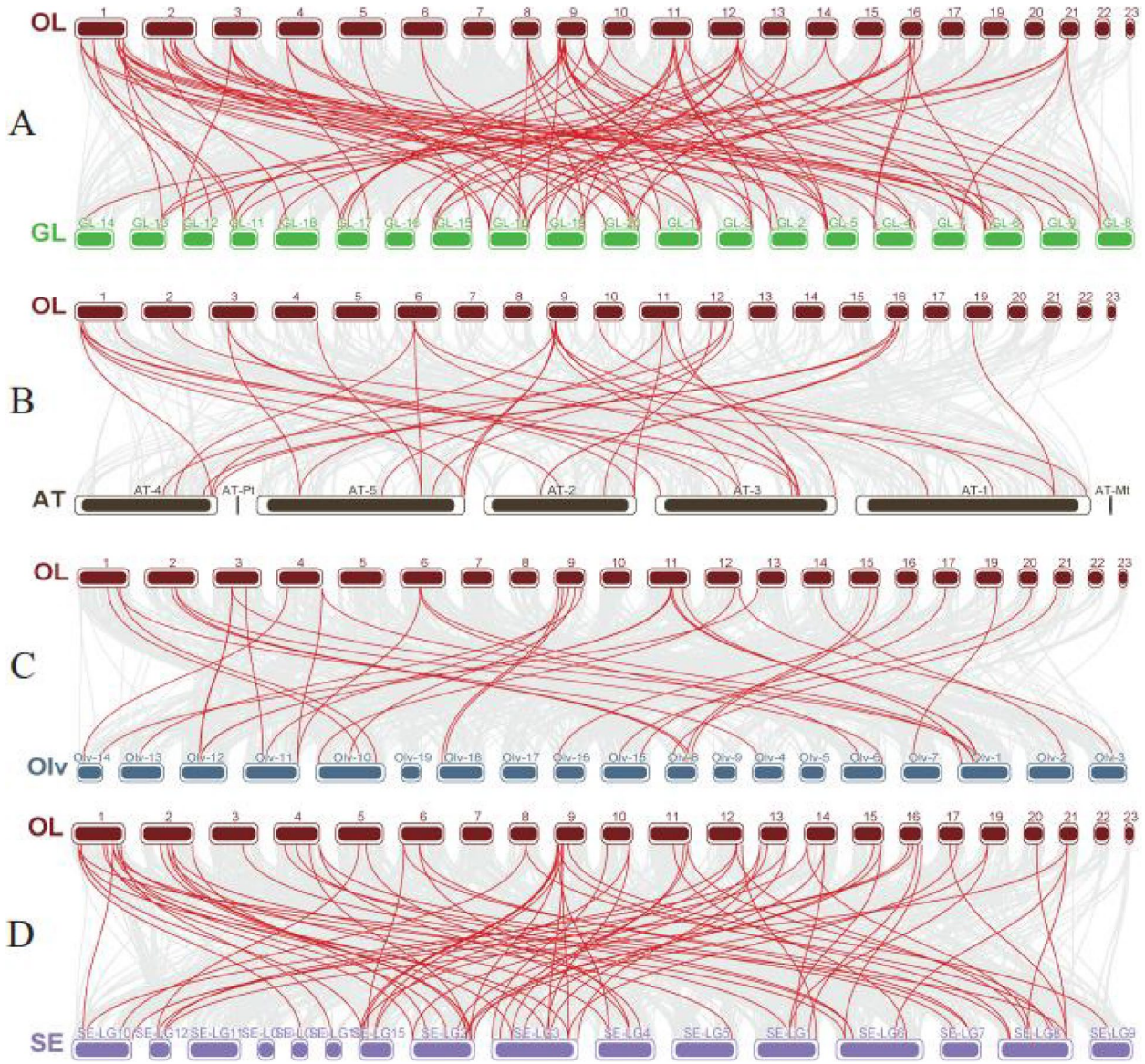


Fig. 6 Homology analysis of the *OLMYB* genes. **A** Collinearity of Olive and soybean MYB genes, (**B**) Collinearity of olive and *Arabidopsis* MYB genes, (**C**) Collinearity of olive and wild olive MYB genes, (**D**) Collinearity of olive and sesame MYB genes. Gray lines, syntenic genes between different species; red lines, homologous gene pairs of MYBs

identification of gene families. The olive genome contains 103 *bZIP* genes, 12 *HSP90* genes, and 40 *FAD* genes [21–23]. This study identified 212 *OLR2R3 MYB* genes in the olive genome, which was higher than the 138 gene sequences found in *Arabidopsis*. This difference in quantity may be related to the whole genome duplication (WGD) event that occurred during the evolutionary process of plants [24].

Predictions of protein chromosome locations and subcellular localization indicate that 110 *OLR2R3 MYB* are unevenly distributed across different olive chromosomes, and chromosome 1 has the most. This pattern of distribution of the MYB genes is similar to the results recorded for other species in previous studies [25].

The phylogenetic analysis constructed using the *Arabidopsis* R2R3 MYB proteins showed that the *OLR2R3 MYB* genes can be divided into six subgroups, which corresponds to 20 *Arabidopsis* subfamilies [26]. However, three *Arabidopsis* subfamily classifications are missing, which suggests that the *OLR2R3 MYB* genes underwent functional diversification during evolution to adapt to environmental selection pressures. Moreover, the

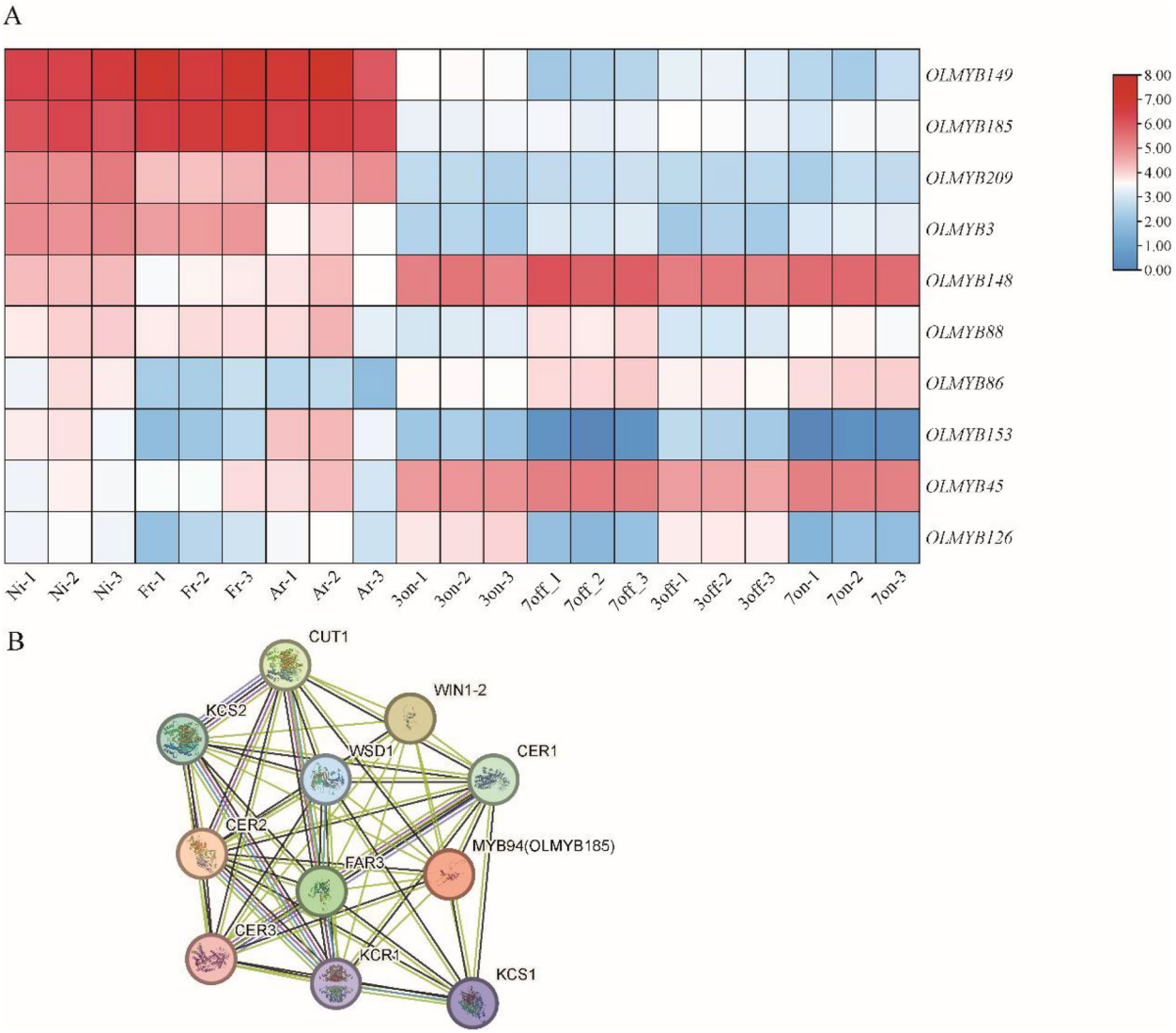


Fig. 7 Patterns of expression of the *OLR2R3 MYB* during the differentiation of flower buds and in the fruits of different olive cultivars. **A** Expression profile using Fragments Per Kilobase of transcript per Million mapped reads (FPKM) data. **B** Protein interaction network of *OLMYB185* predicted using STRING

Arabidopsis-specific S9 subgroup was not found in olive, which indicates that there were differences between their ancestors and the potential disappearance of species-specific MYBs in either olive or *Arabidopsis*. Members within the same subgroup tend to have conserved functions. The results of an analysis of gene structure demonstrate that, similar to the studies of gene structure in other species, all the *OLR2R3 MYB* genes possess conserved MYB gene sequences, while different subgroups have distinct gene sequences [27].

Events, such as tandem repeats and segmental duplications, which are forms of gene duplication, promote the functional diversification of plant genes and the formation of new genes [28]. The olive MYB genes exhibit signs of 17 segmental duplication events and seven tandem repeat gene pairs. The Ka/Ks values of these *OLR2R3 MYB* gene pairs were all less than 1, indicating that the *OLR2R3 MYB* gene underwent purification selection during evolution. This suggests that tandem repeats and segmental duplications may enable olive to adapt to its environment, while the segmental duplications primarily influenced the expansion of the olive MYB gene family. Genes can be duplicated to form multiple new genes among homologous (closely related) species. Olive has a higher number of homologous genes with soybean and sesame compared to *Arabidopsis* [29].

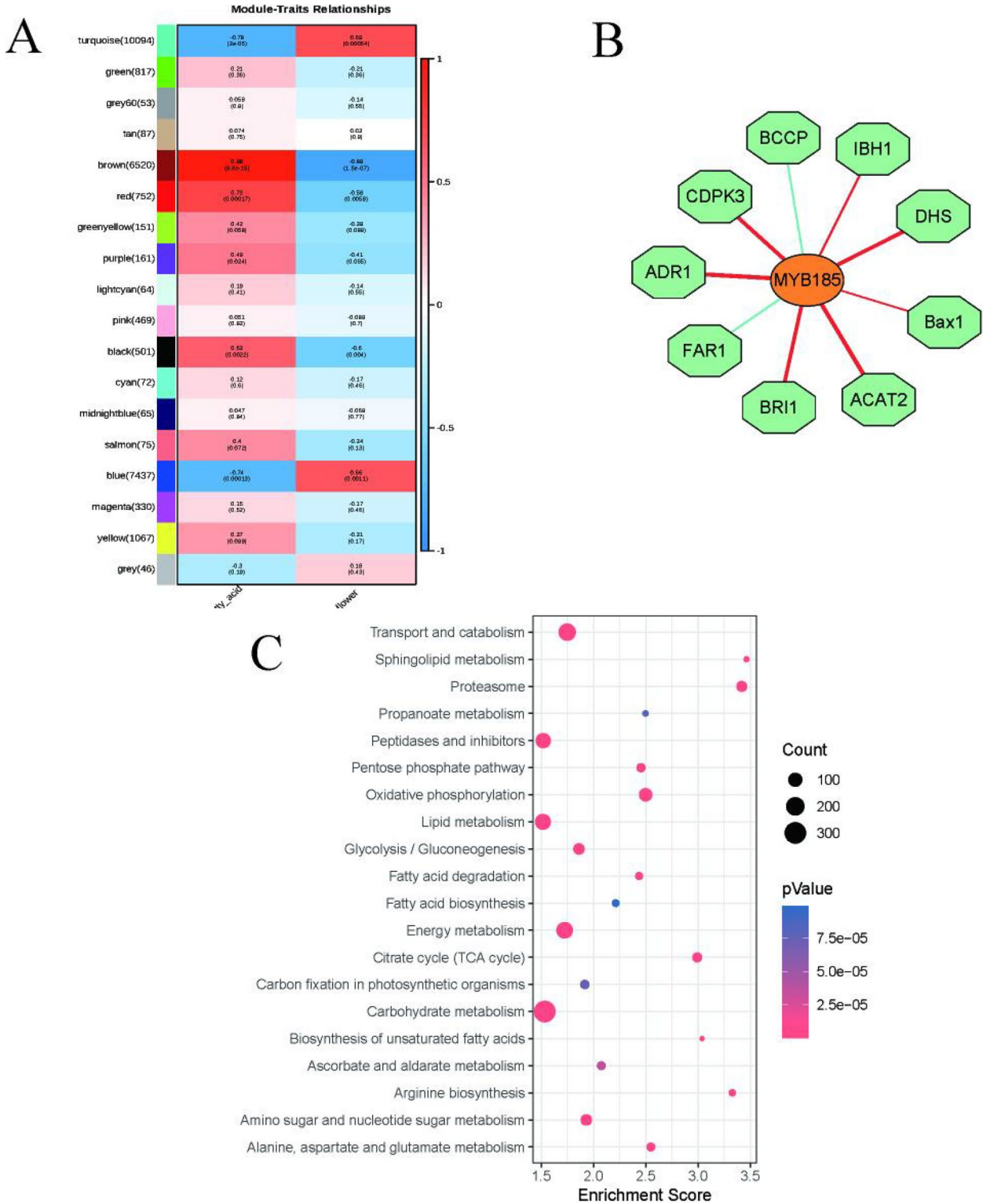


Fig. 8 A WGCNA analysis of the different transcriptomes. **A** Heatmap of module-trait relationships. Red, positive correlation; blue, negative correlation. **B** Co-expression network of the *OLR2R3 MYB*. Yellow boxes, hub genes; green boxes, *OLR2R3 MYB* genes. The thickness of the lines represents the weight of co-expression. Thicker lines, higher weights; thinner lines, lower weights. **C** KEGG enrichment of the brown module genes. KEGG, Kyoto Encyclopedia of Genes and Genomes; WGCNA, Weighted Gene Co-Expression Network Analysis

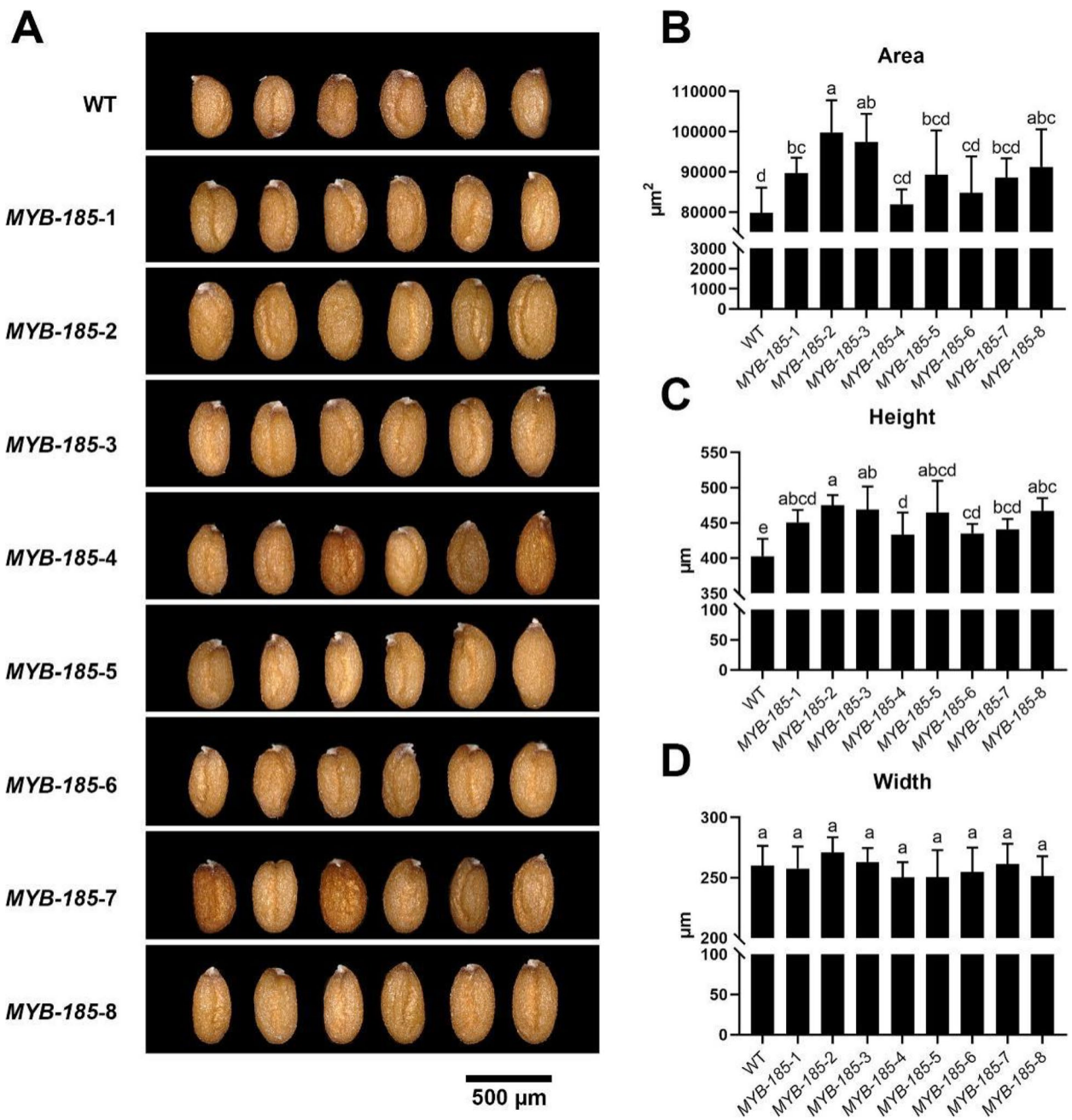


Fig. 9 *OLMYB185* overexpression in fruits. **a** Fruit size of different *OLMYB185* overexpression *Arabidopsis* lines. **b** Surface area of the fruits from different lines of *OLMYB185* overexpressed in *Arabidopsis*. **c** Height of the fruits from different lines of *OLMYB185* overexpressed in *Arabidopsis*. **d** Width of the fruits from different lines of *OLMYB185* overexpressed in *Arabidopsis*

It is worth noting that olive has fewer homologous genes than wild olive, which suggests that environmental factors played a role in shaping the traits of olive during the process of domestication.

Different *cis*-regulatory elements in the sequences of the promoters may result in different patterns of expression [30]. To explore whether the *OLR2R3 MYB* genes are regulated by *cis*-acting elements to enable the adaptation of olive to its environment, we analyzed the *cis*-acting elements present in the 2 kb upstream promoter sequences of the *OLR2R3 MYB* genes. *OLR2R3 MYB* contains 11 types of abiotic and biotic stress-related, 11 types of plant hormone-related, and 18 types of plant growth and development-related promoter

cis-acting elements. There are abundant EREs, and hormone response elements, such as those for gibberellin and salicylic acid, are also present [31]. This indicates that the *OLR2R3 MYB* gene members participate in the regulatory networks of hormone signal transduction. Additionally, the presence of abiotic and biotic stress response elements, those involved in the development of plant meristems (CAT-box), and seed-specific regulation elements (RY-element) suggests that the *OLR2R3 MYB* gene family may participate in the responses of plants to various abiotic stresses and in the regulation of growth and development [32, 33].

The combination of the phylogenetic tree analysis and the analysis of the *cis*-elements related to growth and development of the plants aids in the selection of candidate genes for the accumulation of oil in the fruit [34]. We analyzed the patterns of expression of the MYB genes in different tissues of the olive plants and the fruit of different varieties. Preliminary transcriptome screening led to the identification of two candidate genes, *OLMYB185* and *OLMYB149*, which may be related to these traits in olive fruit. A further co-expression network analysis revealed that *OLMYB185* is associated with genes related to fatty acid biosynthesis, which suggests that the *OLMYB185* may directly or indirectly participate in this process. The WGCNA analyzed the enrichment results of the brown module gene KEGG, and the results showed an enrichment of the pathway related to the biosynthesis of fatty acids. *OLMYB185* strongly correlated with *ACAT2*, *FAR1* and other genes associated with the biosynthesis of fatty acids. The overexpression of *OLMYB185* in *Arabidopsis thaliana* resulted in larger fruits compared to the WT, which indicated that *OLMYB185* is involved in the biosynthesis and accumulation of fatty acids in the fruit.

Seed size and oil content are important agronomic traits during crop domestication [36]. Fatty acids are synthesized mainly in the plastid of plant cells and combine with glycerol to form triglycerides, which are the main form of plant seed oil [37]. Overexpression of the fatty acid-associated *AtFAX1* gene in Arabidopsis resulted in increased oil content and increased seed size, most likely due to increased oil and protein accumulation [38]. *AP2* family members are involved not only in fatty acid synthesis but also in the regulation of seed size. It has been reported that overexpression of the *AP2* gene increases the size of seed outer coat cells, indicating the role of *AP2* in the regulation of seed size [39]. In this study, the overexpression of the *OLMYB185* gene associated with the fatty acid pathway in Arabidopsis Thaliana showed significant changes in seed size, which may be related to the regulation of fatty acid accumulation by the *OLMYB185* gene.

Conclusions

Members of the R2R3-type transcription factor gene family (*OLR2R3 MYB*) were retrieved based on the whole genome information of olive. A total of 212 OLR2R3-MYB genes were identified, which is a substantial number compared to those of other species. The genes from olive were divided into six subgroups using the *Arabidopsis thaliana* MYB genes as a reference. It is worth noting that the *OLR2R3 MYB* genes had previously been divided into six subgroups. Thus, it is likely that these six subgroups represent a more detailed classification.

An additional analysis was conducted on the gene structure, sequence organization, repeats, homology, and *cis*-regulatory elements of these *OLR2R3 MYB* genes. Transcriptome datasets related to the development of fruit were analyzed using olive as the research subject, and the differentially expressed genes (DEGs) associated with the development of fruit were identified.

The use of a co-expression network analysis and the examination of patterns of expression enabled the screening of *OLMYB185*, which is related to the biosynthesis of fatty acids in fruit, as a candidate gene. The overexpression of *OLMYB185* in the model plant *Arabidopsis thaliana* resulted in larger fruit phenotypes compared to the wild type.

This study contributes to understanding the evolution and functional diversification of MYB genes in the olive genome and provides useful information for the breeding of edible cultivars of olive.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06096-7.

Supplementary Material 1.

Acknowledgements
We thank Dr. Liping Wei and Ke Liu for technical assistance.

Authors' contributions
Hengxing Zhu, Qianli Dai and Zhuo Wei conceived and designed the project. Feiyi Huang, Min Lu and Chen Benwen performed the experiments. Chenggong Lei and Ximeng Yang analyzed data. Zhuo Wei wrote the manuscript. Qainli Dai and Ximeng Yang supervised and revised the manuscript. All authors read and approved the manuscript.

Funding
This study was supported the Key Project of Chongqing Technology Innovation and Application Development Special Project (cstc2020jscx-tpyzxX0016).

Data availability
The genomes and gene annotation files of olive, wild olive, sesame and soybean were downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov/), with the primary accession code GCA_902713445.1, GCA_002742605.1, GCA_000512975.1, GCA_000004515.5. While the relevant data for Arabidopsis were downloaded from TAIR (https://www.Arabidopsis.org/).

Declarations

Ethics approval and consent to participate
The methods involved in this study were carried out in compliance with local and national regulations.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 13 November 2024 Accepted: 9 January 2025
Published online: 20 May 2025

References

1. Ambawat S, Sharma P, Yadav NR, Yadav RC. MYB transcription factor genes as regulators for plant responses: an overview. Physiol Mol Biol Pla. 2013;19(3):307–21.
2. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. MYB transcription factors in Arabidopsis. Trends Plant Sci. 2010;15(10):573–81.
3. Huang X, Li Y, Chang Z, Yan W, Xu C, Zhang B, He Z, Wang C, Zheng M, Li Z, et al. Regulation by distinct MYB transcription factors defines the roles of OsCYP86A9 in anther development and root suberin deposition. Plant J. 2024;118(6):1972–90.
4. Ren D, Rao Y, Yu H, Xu Q, Cui Y, Xia S, Yu X, Liu H, Hu H, Xue D, et al. MORE FLORET1 Encodes a MYB Transcription Factor That Regulates Spikelet Development in Rice. Plant Physiol. 2020;184(1):251–65.
5. Wan J, Wang R, Zhang P, Sun L, Ju Q, Huang H, Lü S, Tran LS, Xu J. MYB70 modulates seed germination and root system development in Arabidopsis. iScience. 2021;24(11):103228.
6. Kong Q, Yang Y, Guo L, Yuan L, Ma W. Molecular Basis of Plant Oil Biosynthesis: Insights Gained From Studying the WRINKLED1 Transcription Factor. Front Plant Sci. 2020;11:24.
7. Chapman KD, Ohlrogge JB. Compartmentation of triacylglycerol accumulation in plants. J Biol Chem. 2012;287(4):2288–94.
8. Shi M, Yu L, Shi J, Liu J. A conserved MYB transcription factor is involved in regulating lipid metabolic pathways for oil biosynthesis in green algae. New Phytol. 2022;235(2):576–94.
9. He S, Min Y, Liu Z, Zhi F, Ma R, Ge A, Wang S, Zhao Y, Peng D, Zhang D, et al. Antagonistic MADS-box transcription factors SEEDSTICK and SEPALLATA3 form a transcriptional regulatory network that regulates seed oil accumulation. J Integr Plant Biol. 2024;66(1):121–42.
10. Raffaele S, Vailleau F, Léger A, Joubès J, Miersch O, Huard C, Blée E, Mongrand S, Domergue F, Roby D. A MYB transcription factor regulates very-long-chain fatty acid biosynthesis for activation of the hypersensitive cell death response in Arabidopsis. Plant Cell. 2008;20(3):752–67.
11. Wei X, Mao L, Wei X, Xia M, Xu C. MYB41, MYB107, and MYC2 promote ABA-mediated primary fatty alcohol accumulation via activation of AchnFAR in wound suberization in kiwifruit. Hortic Res-England. 2020;7(1):86.
12. To A, Joubès J, Thueux J, Kazaz S, Lepiniec L, Baud S. AtMYB92 enhances fatty acid synthesis and suberin deposition in leaves of Nicotiana benthamiana. Plant J. 2020;103(2):660–76.
13. Boskou D. Olive oil: chemistry and technology: AOCS Publishing; 2006:35.
14. Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, et al. The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. Nucleic Acids Res. 2003;31(1):224–8.
15. Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M et al. CDD: specific functional annotation with the Conserved Domain Database. Nucleic Acids Res. 2009;37(Database issue):D205-D10.
16. Duvaud S, Gabella C, Lisacek F, Stockinger H, Ioannidis V, Durinx C. Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users. Nucleic Acids Res. 2021;49(W1):W216–27.
17. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol Plant. 2020;13(8):1194–202.
18. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bradley P, Bork P, Bucher P, Cerutti L et al. InterPro, progress and status in 2005. Nucleic Acids Res. 2005, 33(Database issue):D201-D05.
19. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, et al. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2022;50(D1):D20–6.
20. Wang R, Shu P, Zhang C, Zhang J, Chen Y, Zhang Y, Du K, Xie Y, Li M, Ma T, et al. Integrative analyses of metabolome and genome-wide transcriptome reveal the regulatory network governing flavor formation in kiwifruit (Actinidia chinensis). New Phytol. 2022;233(1):373–89.
21. Rong S, Wu Z, Cheng Z, Zhang S, Liu H, Huang Q. Genome-wide identification, evolutionary patterns, and expression analysis of *bZIP* gene family in olive (Olea europaea L.). Genes. 2020;11(5):510.
22. Bettaieb I, Hamdi J, Bouktila D. Genome-wide analysis of HSP90 gene family in the Mediterranean olive (Olea europaea subsp. europaea) provides insight into structural patterns, evolution and functional diversity. Physiol Mol Biol Pla. 2020;26(11):2301–18.
23. Niu E, Gao S, Hu W, Zhang C, Liu D, Shen G, Zhu S. Genome-wide identification and functional differentiation of fatty acid desaturase genes in Olea europaea L. Plants. 2022;11(11):1415.
24. Barrera-Redondo J, Ibarra-Laclette E, Vázquez-Lobo A, Gutiérrez-Guerrero YT. Sánchez de la Vega G, Piñero D, Montes-Hernández S, Lira-Saade R, Eguiarte LE: The genome of Cucurbita argyrosperma (silver-seed gourd) feveals faster rates of protein-coding gene and long noncoding RNA turnover and neofunctionalization within Cucurbita. Mol Plant. 2019;12(4):506–20.
25. Ding P, Tang P, Li X, Haroon A, Nasreen S, Noor H, Attia KA, Abushady AM, Wang R, Cui K et al. Genome-wide identification, phylogeny and expression analysis of the R2R3-MYB gene family in quinoa (Chenopodium quinoa) under abiotic stress. Funct Plant Biol. 2024;51:23261.
26. Li H, Wen X, Wei M, Huang X, Dai S, Ruan L, Yu Y. Genome-wide identification, characterization, and expression pattern of MYB gene family in Melastoma candidum. Horticulturae. 2023;9(6):708.
27. Rao X, Qian Z, Xie L, Wu H, Luo Q, Zhang Q, He L, Li F. Genome-wide identification and expression pattern of MYB family transcription factors in Erianthus fulvus. Genes. 2023;14(12):2128.
28. Katiyar A, Smita S, Lenka SK, Rajwanshi R, Chinnusamy V, Bansal KC. Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis. BMC Genomics. 2012;13:544.
29. Zhou C, Chen Y, Wu Z, Lu W, Han J, Wu P, Chen Y, Li M, Jiang H, Wu G. Genome-wide analysis of the MYB gene family in physic nut (Jatropha curcas L.). Gene. 2015;572(1):63–71.
30. Ma R, Liu B, Geng X, Ding X, Yan N, Sun X, Wang W, Sun X, Zheng C. Biological Function and Stress Response Mechanism of MYB Transcription Factor Family Genes. J Plant Growth Regul. 2023;42(1):83–95.
31. Jung HW, Lim CW, Hwang BK. Isolation and functional analysis of a pepper lipid transfer protein III (CALTPIII) gene promoter during signaling to pathogen, abiotic and environmental stresses. Plant Sci. 2006;170(2):258–66.
32. Shi J, Yao J, Tong R, Wang S, Li M, Song C, Wan R, Jiao J, Zheng X. Genome-wide identification of laccase gene family from Punica granatum and functional analysis towards potential involvement in lignin biosynthesis. Horticulturae. 2023;9(8):918.
33. Huang Y, Xu Z, Xu L, Wei P, Zhang D, He X, Guo S, Peng C, Ma H. Promoter cloning and expression analysis of transcription factor gene GMMYB92 in soybean (Glycine max. L). Pak J Agr Sci. 2016;53(3):487–94.
34. Beltrán G, del Río C, Sánchez S, Martínez L. Seasonal changes in olive fruit characteristics and oil accumulation during ripening process. J Sci Food Agr. 2004;84(13):1783–90.
35. Dongxue Su, Zheng J, Yi Y, Zhang S, Feng L, Quzhen D, Qiong De, Zhang W, Wang Q, Feng Xu. Genome-Wide Identification of the MYB Gene Family and Screening of Potential Genes Involved in Fatty Acid Biosynthesis in Walnut. Phyton-Int J Experiment Botany. 2024;9:2317–37.
36. Zuo JF, Ikram M, Liu JY, Han CY, Niu Y, Dunwell JM, Zhang YM. Domestication and improvement genes reveal the differences of seed size- and oil-related traits in soybean domestication and improvement. Comput Struct Biotechnol J. 2022;20:2951–64.

37. Parchuri P, Bhandari S, Azeez A, Chen G, Johnson K, Shockey J, Smertenko A, Bates PD. Identification of triacylglycerol remodeling mechanism to synthesize unusual fatty acid containing oils. Nat Commun. 2024;15(1):3547.
38. Tian Y, Lv X, Xie G, Zhang J, Xu Y, Chen F. Seed-specific overexpression of *AtFAX1* increases seed oil content in *Arabidopsis*. Biochem Biophys Res Commun. 2018;500(2):370–5.
39. Ohto MA, Floyd SK, Fischer RL, Goldberg RB, Harada JJ. Effects of *APETALA2* on embryo, endosperm, and seed coat development determine seed size in *Arabidopsis*. Sex Plant Reprod. 2009;22(4):277–89.

Publisher's Note