Food Chemistry 300 (2019) 125246

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Identification of putative genes for polyphenol biosynthesis in olive fruits and leaves using full-length transcriptome sequencing

Rao Guodong[a,*], Zhang Jianguo[a,b,c,*], Liu Xiaoxia[a], Luo Ying[a]

[a] State Key Laboratory of Tree Genetics and Breeding, Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China
[b] Collaborative Innovation Center of Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, China
[c] Key Laboratory of Tree Breeding and Cultivation, State Forestry Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China

ARTICLE INFO

Keywords:
Olea europaea
Polyphenol
Oleuropein
Hydroxytyrosol
Full-length transcripts
Alternative splicing

ABSTRACT

Olive (*Olea europaea*) is a rich source of valuable bioactive polyphenols, which has attracted widespread interest. In this study, we combined targeted metabolome, Pacbio ISOseq transcriptome, and Illumina RNA-seq transcriptome to investigate the association between polyphenols and gene expression in the developing olive fruits and leaves. A total of 12 main polyphenols were measured, and 122 transcripts of 17 gene families, 101 transcripts of 9 gene families, and 106 transcripts of 6 gene families that encode for enzymes involved in flavonoid, oleuropein, and hydroxytyrosol biosynthesis were separately identified. Additionally, 232 alternative splicing events of 18 genes related to polyphenol synthesis were analyzed. This is the first time that the third generations of full-length transcriptome technology were used to study the gene expression pattern of olive fruits and leaves. The results of transcriptome combined with targeted metabolome can help us better understand the polyphenol biosynthesis pathways in the olive.

1. Introduction

Olive (*Olea europaea* L.), a member of the family Oleaceae, is native to the Mediterranean basin and parts of Asia. It was introduced into China in 1956, mainly in the Yunnan, Gansu, Sichuan, and Shaanxi provinces. The tree, famous for its fruit, is of worldwide economic importance for the production of olive oil and table olives. Olive fruit accumulates high oil contents ranging from 28% to 30% of total mesocarp fresh weight. The olive fruits and leaves are rich in secondary metabolites, such as flavonoids, oleuropein, tyrosol, and hydroxytyrosol (HT). As a result, the olive is known to possess various biological properties, such as antioxidant, antiviral, anti-inflammatory, antimicrobial, anticarcinogenic, antihypertensive, antidyslipidemic, cardiotonic, laxative, and antiplatelet (Raederstorff, 2009). In particular, the consumption of olive oil has played an important role in preventing and treating cardiovascular disease and certain cancers (Pauwels, 2011). The majority of studies on olive cultivars in China have focused on fatty acids in the oil (Cheng et al., 2018), while few studies have reported on the active constituents of its fruits and leaves. It is thus necessary to study the olive cultivars that are grown in China in order to guide the cultivation, development, and utilization of olive in this region. Therefore, research was carried out on the main active constituents, including flavonoids, oleuropein, and HT.

Flavonoids constitute one of the largest groups of polyphenols and exist widely in plants. The flavonoids present in olive are abundant in active substances, such as luteolin, apigenin, caffeic acid, ferulic acid, and some others (Brahim, Kelebek, Ammar, Abichou, & Bouaziz, 2017). The first step in the synthesis of flavonoids is to obtain *p*-coumarate-CoA from phenylalanine through continuous catalysis by three enzymes, namely phenylalanine ammonialyase (PAL), cinnamate 4-hydroxylase (C4H), and 4-coumarate CoA ligase (4CL). Synthesis is then initiated with the formation of chalcone by chalcone synthase (CHS) using three molecules of malonyl-CoA and one molecule of *p*-coumarate-CoA. Chalcone isomerase (CHI) catalyzes the synthesis of naringenin, which is crucial in the synthesis of various flavonoids, including isoflavones, flavones, flavanols, and anthocyanins (Zhang et al., 2017).

Oleuropein is the main secoiridoid, representing up to 82% of total biophenols, and is known as the bitter principle of olives (Alagna et al., 2009). The first step in the oleuropein biosynthesis pathway is initiated from geranyl diphosphate and catalyzed by geraniol synthase (GES), following which the product geraniol is converted into 8-hydroxygeraniol by geraniol 8-hydroxylase (G8H). Secoiridoids are derived from iridoids by the opening of the cyclopentane ring, and in the Oleaceae family, the resulting carbonyl group is oxidized and conjugated with a phenolic moiety. In recent years, studies have elucidated

* Corresponding authors at: Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China.
E-mail addresses: rgd@caf.ac.cn (R. Guodong), ralf02@163.com (Z. Jianguo).

https://doi.org/10.1016/j.foodchem.2019.125246
Received 20 March 2019; Received in revised form 22 July 2019; Accepted 23 July 2019
Available online 24 July 2019

the pathway from geranyl diphosphate to secologanin in *Catharanthus roseus* (Miettinen et al., 2014) and *O. europaea* (Mougiou et al., 2018). The pathway involves seven enzymes catalyzing successive reactions from 8-hydroxygeraniol to secologanin: 8-hydroxygeraniol oxidoreductase (8-HGO), iridoid synthase (IS), iridoid oxidase (IO), 7-deoxyloganetic acid-O-glucosyl transferase (7-DLGT), 7-deoxyloganic acid hydroxylase (7-DLH), loganic acid methyltransferase (LAMT), and secologanin synthase (SLS).

HT is another important family of compounds present in olive. HT forms the aglycon part of oleuropein with elenolic acid. The biosynthesis of HT was previously clarified in *O. europaea* using cultured cells (Saimaru & Orihara, 2010). HT is initiated from tyrosine and is preceded by polyphenol oxidase (PPO) and dihydroxyphenylalanine (DOPA) decarboxylase (DDC) through DOPA and dopamine. Dopamine may then be converted into HT via two enzymes, namely primary-amine (copper-containing) oxidase (CuAO) and alcohol dehydrogenase (ALDH). Alternatively, tyrosine could first use tyrosine decarboxylase (TDC), CuAO, and phenyl-acetaldehyde reductase (PAR) to produce tyrosine, the precursor molecule of HT. The exact biosynthetic pathway that leads from tyrosine to HT in olive is yet to be clarified.

To date, the focus of most research on olive polyphenol biosynthesis has been on the use of second-generation sequencing (SGS) platforms (Mougiou et al., 2018; Parra, Paredes, Sanchez-Calle, & Gomez-Jimenez, 2013). Many issues exist with second-generation transcriptional data, such as the short reads and the erroneous splicing tendency, which may result in incompletely assembled transcripts and the loss of some important information. Recently, single-molecule real-time (SMRT) sequencing conducted using the PacBio RS system has provided a third-generation sequencing platform. This technology offers four major advantages: long read lengths, high consensus accuracy, a low degree of bias, and simultaneous capability for epigenetic characterization (Chen et al., 2018). Therefore, the technology is widely used in genome and transcriptome sequencing. Alternative splicing (AS) is an important post-transcriptional regulatory mechanism that significantly enhances transcriptome diversity in multicellular eukaryotes. Full-length transcript sequences allow for the efficient analysis of AS.

In order to identify key genes contributing to the high content of polyphenols and to understand the molecular mechanisms underlying flavonoid, oleuropein, and HT biosynthesis in olive, the present study used both Illumina HiSeq™4000 and SMRT technology to obtain a global overview of transcripts and alternative splicing events during the maturation of olive fruit and leaves. All the genes involved in flavonoid, oleuropein, and hydroxytyrosol biosynthesis were screened, and their expression patterns between different olive fruits and leaves were studied. The AS events in biosynthesis were also analyzed. This is the first study to combine second- and third-generation sequencing to generate the complete and full-length transcriptome of olive. The results are necessary for deducing and assessing the role of the encoded protein and splice variants in gene regulation in olive.

2. Materials and methods

2.1. Plant materials, RNA extraction, and sequencing

Fifteen-year-old olive trees (*O. europaea* L. cv. Leccino) were grown in the research garden of the Research Institute of Forestry, Chinese Academy of Forestry in Gansu, China. The fruit samples were collected at five developmental stages: 50, 80, 110, 140, and 170 days after flowering (F1, F2, F3, F4, and F5). The leaf samples were collected at three different aging degrees: new leaves on the tips of stems of current year branches (NL), old leaves at the base of current year branches (OL), and the leaves on the branches of last year (OY). Three replicates (Three trees) for each sample were collected, flash frozen in liquid nitrogen, and stored at −80 °C until further use.

Total RNA was obtained from the equally mixed samples of three replicates using a Tiangen RNA Prep Pure Plant kit (Tiangen Biotech Co. Ltd., Beijing, China) according to the manufacturer's instructions. The RNA quality was evaluated using a NanoDrop ®ND-1000 spectrophotometer and an Agilent 2100 bioanalyzer (Agilent Technologies, Santa Clara, USA). For Illumina RNA-Seq, RNA (3 μg per sample) was prepared as the input material for construction of sequencing libraries by using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) following the manufacturer's recommendations. Libraries were sequenced using an Illumina HiSeq™4000 platform with high output run mode. SMRT sequencing was conducted using a Pacbio Sequel platform. For SMRT cell library construction, first-strand cDNA was synthesized using a SMARTer PCR cDNA Synthesis Kit (Clontech, Japan), and second-strand cDNA synthesis was performed using DNA Polymerase (NEB, USA). BluePippin size selection (Sage Science, USA) was then used for cDNA fragment screening, followed by cDNA normalization via a Trimmer-2 cDNA Normalization Kit (Evrogen, Russia). Three libraries (1–2, 2–3, and 3–6 KB cDNA fractions) were constructed. Library preparation and sequencing were conducted by NovoGene Science and Technology Corporation (Beijing, China).

2.2. Bioinformatics analysis

Raw data (raw reads) of fastq format were firstly processed through in-house perl scripts. In this step, clean data (clean reads) were obtained by removing reads containing adapter, reads containing ploy-N and low quality reads from raw data. All the downstream analyses were based on the clean data with high quality. Gene structure analysis was performed using the Transcriptome Analysis Pipeline using the Isoform Sequencing (TAPIS) pipeline (Abdelghany et al., 2016). The clean reads were then mapped to the olive reference genome using the Tophat2 tools software program (Cruz et al., 2016; Kim et al., 2013). Plant transcription factors were predicted using iTAK software. A *P*-value < 0.05 and $|\log2 (\text{foldchange})| > 1$ were set as thresholds for significant differential expression. Heatmaps were constructed in the BioMarker cloud platform (http://www.biocloud.net). Differential expression analysis was performed using the DESeq R package with a model based on the negative binomial distribution (Varet, Brillet-Guéguen, Coppée, & Dillies, 2016). Gene Ontology (GO) enrichment analysis of differentially expressed genes was implemented by the GOseq R package. GO terms with corrected P-value less than 0.05 were considered significantly enriched by differential expressed genes (Young, Wakefield, Smyth, & Oshlack, 2010).

The RNA sequence (both ISOSeq and RNA-Seq) raw data was deposited in the NCBI Sequence Read Archive as follows: ISOseq of fruits_Mix (1–2 K): SRR8606699; ISOseq of fruits_Mix (2–3 K): SRR8606700; ISOseq of fruits_Mix (3–6 K): SRR8606701; Fruits_1 (F1) by RNA-seq: SRR8446454; Fruits_2 (F2) by RNA-seq: SRR8446455; Fruits_3 (F3) by RNA-seq: SRR8446452; Fruits_4 (F4) by RNA-seq: SRR8446453; Fruits_5 (F5) by RNA-seq: SRR8446451. Targeted metabolome data have been deposited to the MetaboLights with the accession number of MTBLS814 (https://www.ebi.ac.uk/metabolights/MTBLS814).

2.3. Gene validation and quantitative real-time PCR analysis

Quantitative real-time PCR (qRT-PCR) was used for gene validation. A total of 20 randomly selected genes from the transcriptome database were first analyzed to design qRT-PCR specific primers (Table S1). PCR was conducted using SYBR Green Perfect (Takara, Japan) and StepOnePlusTM System (Applied Biosystems, USA). PCR products were sequenced to verify sequence specificity and correctness, and the dissociation curve was analyzed to verify the specificity of the amplification. A standard curve from purified PCR products was established and used to establish a quantitative correlation between the CT values and the gene copy numbers. Each standard curve contained at least five points, and the quantitative real-time PCR reaction was performed at least three times. The olive actin gene was used as a

reference gene for the validation of transcriptome gene expression.

2.4. Extraction of polyphenols

Each sample (40 mg) was powdered and dried at 60 °C until constant weight. The samples were then repeatedly extracted twice with 3 ml methanol, sonicated in an ultrasonic cleaner at room temperature for 30 min, centrifuged, and the supernatant was then collected. The supernatants from each sample were pooled, concentrated, and dried under vacuum. The powder was dissolved again in 0.4 ml methanol and diluted 200 times. The extract was filtered through a 0.22 μm Millipore membrane. The clear supernatant was then analyzed by liquid chromatography-mass spectrometry (LC–MS). Three replicates were performed for each sample.

2.5. LC–MS conditions and detection procedure

The LC–MS conditions and detection procedure were mainly performed according to previous study (Eleni, Zweigenbaum, & Mitchell, 2015). Chromatographic separations were performed on a Waters ACQUITY UPLC® BEH C_{18} column (2.1 × 100 mm, 1.7 μm). The flow rate was 0.25 ml min^{-1} and the column temperature was maintained at 40 °C. The injection volume of each sample was 5 μl. Eluent A_1 was 0.1% formic acid aqueous solution, eluent A_2 was water, eluent B was methanol. A gradient elution protocol of tyrosol was used as follows: 0–2 min, 90%–10% A_1; 2–2.5 min, 10% A_1; 2.5–3 min, 10%–90% A_1; 3–5 min, 90% A_1. A gradient elution protocol for other polyphenols was used as follows: 0–1 min. 90% A_2; 1–3 min, 90%–10% A_2; 3–5.5 min, 10% A_2; 5.5–6 min, 10%–90% A_2; and 6–9 min, 90% A_2. The MS detection conditions were as follows: electrospray ionization (ESI) source; negative ionization (NI) modes; ionization temperature, 500 °C; ionization voltage, 4.5 kV; collision gas, 6 psi; curtain gas, 30 psi; ion source gas 1, 50 psi; ion source gas 2, 50 psi. Multiple reactions monitoring (MRM) was used for scanning. The ion pairs in this study used for quantitative analysis are listed in Table S2. The linear regression equations, linear ranges, and limits of quantitation are listed in Table S3.

2.6. Statistical analyses

One-way analysis of variance (ANOVA) and bivariate correlation analysis were performed using SPSS 21.0. The differences were considered to be significant at $P < 0.05$.

3. Results and discussion

3.1. Contents of individual polyphenolic compounds

To elucidate the polyphenol contents in different developmental stages of olive, the fruit and leaf extracts were analyzed. Twelve polyphenol compounds were identified and quantified, including five phenolic acids (*p*-coumaric acid, caffeic acid, vanillic acid, ferulic acid, and maslinic acid), two flavonols (rutin and quercetin), two flavones (luteolin and apigenin), one secoiridoid (oleuropein), and two phenols (hydroxytyrosol and tyrosol). In order to assess the data repeatability, we conducted repeated correlation assessments and principal component analysis (Fig. S1). There was a high correlation between the biological repetitions among intra-group samples, which indicated that the data were credible and reproducible.

The contents of the detected polyphenols in the studied stages of the olive fruits and leaves are listed in Table 1. Olive fruit is rich in bioactive pentacyclic triterpenoids, mainly maslinic acid (Fukumitsu et al., 2016). In this study, the most abundant polyphenolic compound found in the olive fruits and leaves was also maslinic acid, with concentrations ranging from 698.67 to 1105.33 ng/mg fresh weight (FW). However, the content of other phenolic acids (*p*-coumaric acid, caffeic acid, vanillic acid, and ferulic acid) was less than 10 ng/mg FW. The highest maslinic acid content among the five developmental stages of olive fruit was found in F3 and indicated a gradual increase followed by a rapid decrease during the fruit-ripening process (Fig. 1A). With increased leaf age from NL to OY, the content of maslinic acid increased gradually (Table 1). Maslinic acid from olive, which is safe and nontoxic for daily consumption, was found to alleviate mild knee joint pain and improve quality of life in the elderly by promoting weight loss (Fukumitsu et al., 2016). In this study, the content of maslinic acid in the leaves was high, particularly in OY, which suggested that olive leaves may have potential in the functional process of making a tea beverage. The sum of the individual phenolic acids accounted for 43.95%–70.30% of the total mass of polyphenolic compounds detected in olive, making phenolic acid the most abundant type of polyphenolic compound.

The total quantity of the two phenolic compounds, hydroxytyrosol and tyrosol, ranked second and third in the olive fruits and leaves, respectively. It is well known that hydroxytyrosol and oleuropein possess powerful antioxidant activity both *in vivo* and *in vitro* (Tripoli et al., 2005). In this study, the content of hydroxytyrosol increased non-significantly at the beginning of ripening (Fig. 1B), which could possibly be attributed to the hydrolysis of components with higher molecular weights in olive fruits (Brahim et al., 2017). Hydroxytyrosol content then decreased dramatically from stage F2 to F5, which is in accordance with observations in other olive cultivars (Brahim et al., 2017).

Oleuropein causes a bitter taste, and green olives, which have higher levels of oleuropein, are more bitter (Visioli & Galli, 2001). Oleuropein content also has been shown to be higher in the first stage of fruit maturation, thereafter decreasing as the fruits ripen in our study. The oleuropein content in the leaves was about 6.32 times than that in the fruits on average and was higher in OL than in OY and NL (Fig. 1C, Table 1).

The sum of individual flavones (luteolin and apigenin) in the five fruit stages averaged 1.50% (from 0.26 to 2.74%), which was slightly lower than the proportion of oleuropein (an individual secoiridoid, averaged 2.7%). Furthermore, of the types of polyphenolic compounds, the quantities of flavones were lowest both in the fruits and leaves, with the flavone content in the leaves being about 3 times that of the fruits (Fig. 1D, Table 1). Two flavonols (rutin and quercetin) in the olive fruits were present in proportions ranging from 9.11 to 22.85% (14.41% on average) of the total mass at each stage (Fig. 1E, Table 1). The levels of quercetin in the fruits at all stages were low, ranging from 0.71 to 1.48 ng/mg FW, and the content of quercetin in the leaves was much higher than that in the fruits (4.00–69.76 times). The lowest amounts of quercetin and rutin observed in the fruits were in F4 and F5, respectively. Olive leaves have been determined to possess the highest antioxidant and scavenging power of all the different parts of the olive tree (El & Karakaya, 2009). These properties may be attributed to the higher contents of luteolin, apigenin, quercetin, and oleuropein in the olive leaves. These two flavones both first increased and then decreased during fruit development. The same trend has been observed in other studies (Bouaziz, Jemai, Khabou, & Sayadi, 2010; Brahim et al., 2017). Overall, there were great differences in the sum of the 12 individual polyphenolic compounds among all fruit and leaf samples.

3.2. Identification of transcripts putatively involved in flavonoid biosynthesis in olive

A total of 59,845 transcripts were obtained in all samples tested. FPKM (expected number of Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced) was used to calculate the amount of gene expression in each sample. FPKM intervals analysis showed that FPKM between 0 and 1 exceeded 50% in all samples, followed by FPKM between 3 and 15, accounting for approximately 20% of all transcripts, FPKM was between 1 and 3, between 15 and 60 in about 10% of all transcripts, and FPKM was greater than 60 in about

Table 1
Individual phenolic composition in five developmental stages of olive fruit (ng/mg FW).

	F1	F2	F3	F4	F5	NL	OL	OY
Phenolic acids								
Vanillic acid	NT	NT	8.15 ± 1.38 d	4.65 ± 0.42e	NT	4.79 ± 0.3 f	NT	4.42 ± 0.59 d
Caffeic acid	0.74 ± 0.25 d	0.55 ± 0.27 d	0.4 ± 0.34 d	NT	NT	NT	NT	NT
p-Coumaric acid	NT	NT	2.14 ± 0.59 d	NT	NT	NT	NT	NT
Ferulic acid	0.44 ± 0.11 d	0.62 ± 0.21 d	0.45 ± 0.13 d	0.36 ± 0.07 e	NT	NT	1.14 ± 0.06 e	0.32 ± 0.04 d
Maslinic acid	931.33 ± 102.65 a	1015.33 ± 54.12 a	1105.33 ± 141.23 a	813.33 ± 154.94 a	698.67 ± 28.45 a	708 ± 27.5 a	776 ± 49.76 a	1078.67 ± 249.08 a
Sum of individual phenolic acid	**932.51 (43.95%)**	**1016.50 (58.09%)**	**1116.46 (69.55%)**	**818.34 (65.42%)**	**698.67 (70.30%)**	**712.79 (54.75%)**	**777.14 (50.59%)**	**1083.41 (59.07%)**
Simple phenols								
Hydroxytyrosol	119.2 ± 18.13 c	161.13 ± 43.65 c	102.73 ± 26.87 c	114.4 ± 17.5 bc	95.33 ± 30.35 b	29.27 ± 6.71 ef	155.87 ± 17.39 c	60.27 ± 17.14 d
Tyrosol	476.67 ± 40.51 b	304.4 ± 152.91 b	70.07 ± 12.91 cd	92.33 ± 14.85 cd	72.67 ± 8.29 b	91.07 ± 23.66 c	2.58 ± 0.94 e	25.2 ± 1.56 d
Sum of individual phenols	**595.87 (28.08%)**	**465.53 (26.60%)**	**172.80 (10.77%)**	**206.73 (16.53%)**	**168.00 (16.91%)**	**120.33 (9.24%)**	**158.45 (10.31%)**	**85.47 (4.66%)**
Secoiridoids								
Oleuropein	103 ± 14.67 c	69.73 ± 21.19 d	49.47 ± 9.95 cd	17.45 ± 2.87de	9.28 ± 3.27 cd	252.03 ± 24.25 b	394.12 ± 121.7 b	298.67 ± 70.04 b
Sum of individual secoiridoids	**103.00 (4.85%)**	**69.73 (3.98%)**	**49.47 (3.08%)**	**17.45 (1.39%)**	**9.28 (0.93%)**	**252.00 (19.35%)**	**394.00 (25.65%)**	**298.67 (16.28%)**
Flavones								
Luteolin	5.1 ± 0.38 d	9.09 ± 1.41 d	19.83 ± 4.96 d	32.6 ± 4.13 de	26.73 ± 0.5 c	69.8 ± 20.03 cd	69.67 ± 12.94 de	92.4 ± 22 cd
Apigenin	0.38 ± 0.1 d	0.45 ± 0.01 d	0.76 ± 0.17 d	0.79 ± 0.13 e	0.52 ± 0.04 d	13.49 ± 3.71 f	44.2 ± 3.52 de	22.66 ± 6.45 d
Sum of individual flavones	**5.48 (0.26%)**	**9.54 (0.55%)**	**20.59 (1.28%)**	**33.39 (2.67%)**	**27.25 (2.74%)**	**83.29 (6.40%)**	**113.87 (7.41%)**	**115.06(6.27%)**
Flavonols								
Quercetin	1.48 ± 0.31d	1.29 ± 0.11d	0.83 ± 0.29 d	0.71 ± 0.12e	0.83 ± 0.31d	49.53 ± 5.69 de	5.92 ± 0.62 e	39.67 ± 7.81 d
Rutin	483.33 ± 74.2 b	187.4 ± 57.77 c	245 ± 52.2 b	174.27 ± 22.83 b	89.73 ± 15.23 b	84.07 ± 6.18 c	86.87 ± 12.01 cd	211.93 ± 43.42 bc
Sum of individual flavonols	**484.81 (22.85%)**	**188.69 (10.78%)**	**245.83 (15.32%)**	**174.97 (13.99%)**	**90.57 (9.11%)**	**133.60 (10.26%)**	**92.79 (6.04%)**	**251.60 (13.72%)**
Total polyphenols	**2121.67 (100%)**	**1750.00 (100%)**	**1605.15 (100%)**	**1250.88 (100%)**	**993.77 (100%)**	**1302.02 (100%)**	**1536.24 (100%)**	**1834.20 (100%)**

Values, in ng/mg FW, are expressed as means ± SD (n = 3). Means in the same line followed by different letters (a–g) are significantly different ($p < 0.05$). ND: Not detected.

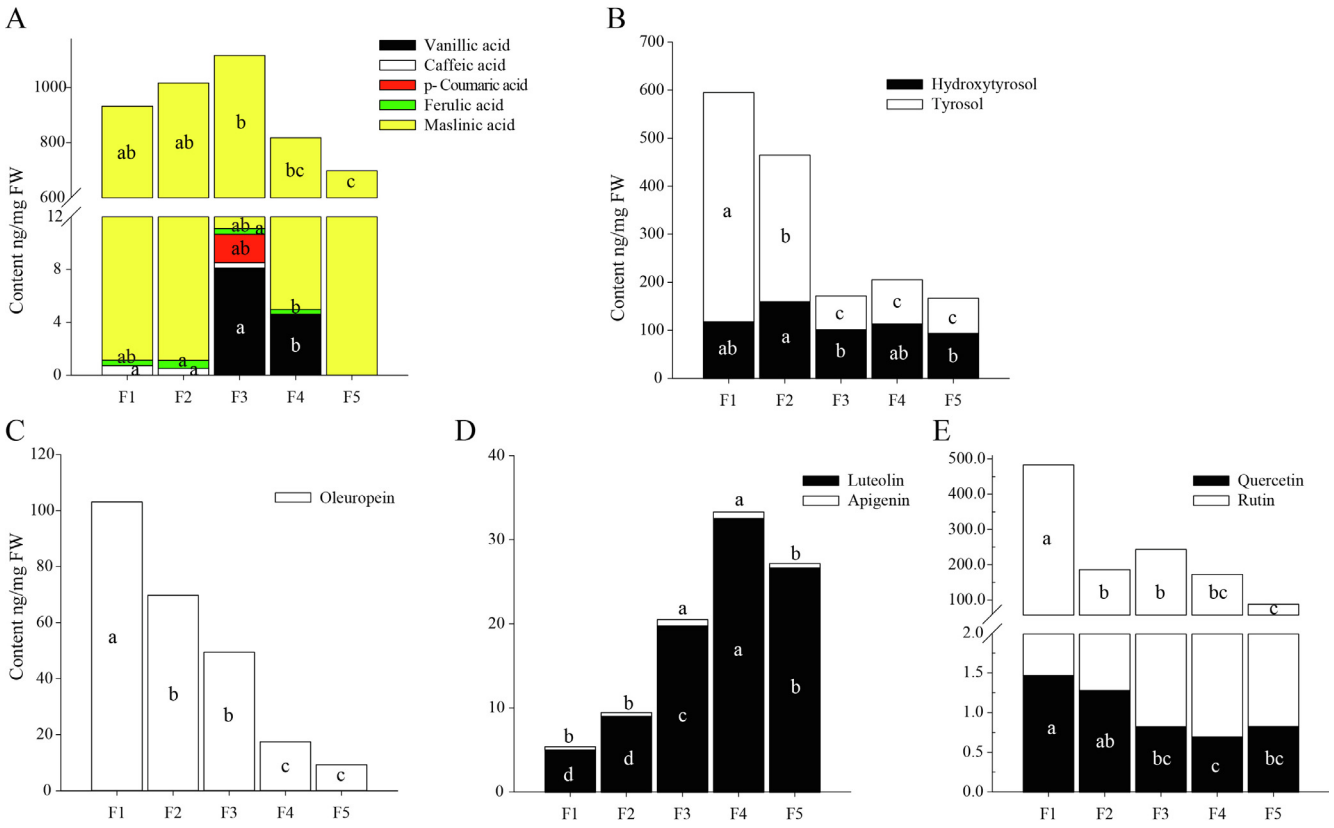

Fig. 1. Variation in the different polyphenolic compounds during the five developmental stages of olive fruit. (A) Five phenolic acid compounds; (B) two phenols; (C) one secoiridoid; (D) two flavones; (E) two flavonols. The different letters (a–g) in the boxes with the same colour indicate statistical significance ($P < 0.05$).

3% of all transcripts. (Table S4). Next we carried out a comparative analysis of the differential genes of fruits and leaves (Fig. S2). The results of the venn diagram showed that F6 and F7 had the most specific expressed genes (2848) in the fruit, while the least difference gene (253) existed between F9 and F10, suggesting that more regulatory genes are needed in the early stage of fruit growth. 69 genes were differentially expressed in all comparison groups, suggesting that these genes play a key regulatory role in fruit development (Fig. S2A). Among the leaves, NL and OL have the most specifically expressed differential genes (2273), while NL and OY have fewer differential genes (188), suggesting that there is a larger biological differences between the new leaves and the old leaves (Fig. S2B). Twenty genes were randomly selected for a real-time qPCR experiment to validate the gene expression level. Results showed that the trend of transcript changes in the transcriptome of these 20 genes is in good agreement with the quantitative results of qPCR. This proves the accuracy of the transcriptome data (Fig. S3)

A total of 991 transcripts were found to be involved in the flavonoid biosynthesis pathway (Fig. 2). Among them, 42 transcripts were not detected in the fruits or leaves of the different developmental stages. A total of 122 transcripts across 17 gene families associated with flavonoid biosynthesis are indicated in the heatmaps in Fig. 2. Our data lists more transcripts in the flavonoid pathway of olive than previous studies (Iaria, Chiappetta, & Muzzalupo, 2015), suggesting that a combination of second and third generation sequencing can increase data acquisition. According to the Venn diagram, 584 transcripts were shared between all five fruit samples tested, and 696 transcripts were expressed in the three leaf samples (Fig. S4A, B). Additionally, 553 transcripts were expressed both in the olive fruits and leaves (Fig. S4C). There were 31 and 143 transcripts that were exclusively expressed in the fruits or leaves, suggesting that they may exhibit tissue-specificity and thus require further study (Fig. S4C). Interestingly, there were more than 100 genes involved in the flavonoid biosynthesis pathway expressed specifically in the leaves than in the fruits, demonstrating that the higher content of flavonoids in the leaves was positively correlated with more genes in the leaves. In Fig. 2A, we found that 61 transcripts were expressed specifically in the F1 stage of the fruits, which was much more than in the other four stages. Green olives have slightly less oil, are bitterer, and can be higher in polyphenols. This finding demonstrated that F1 constitutes the main period of polyphenol synthesis and accumulation. Similarly, there were more exclusively expressed transcripts in NL than in OL and OY, indicating that the new leaves exhibit greater flavonoid accumulation.

Of the transcripts associated with flavonoid synthesis, 387 were differentially expressed in the fruits and leaves. When compared with the F1 stage, the up-regulated transcripts (fold-change > 2) from stage F2 to F5 were 51, 55, 85, and 84, whereas the down-regulated transcripts (fold-change < 0.5) were 114, 129, 164, and 166 (Fig. 2D). Furthermore, 67.6% of the DEGs exhibited down-regulated expression over time, and the numbers of the down-regulated transcripts were greater than the up-regulated transcripts when using F1 as a control (Fig. 2D). In contrast, the transcript numbers did not change sharply in the adjacent groups (i.e., the adjacent groups of F3 vs. F2, F4 vs. F3, and F5 vs. F4) when the prior developing stage served as the control (Fig. 2E). As fruit growth progressed, more down-regulated genes were identified, and the expression of further genes decreased gradually, which accounts for the decreases in the content of some flavonoids, such as rutin (Fig. 2). In the leaf, 130 and 92 transcripts were up- and down-regulated in the group of NL vs. OL, respectively (Fig. 2F). There were fewer up-regulated transcripts in NL vs. OY (36) and OL vs. OY (51), suggesting that young olive leaves are important for secondary metabolite accumulation and biosynthesis.

Notably, the transcripts of the same gene family may exhibit different expression trends. Importantly, we found that the expression patterns of some key genes were consistent with the accumulation patterns of flavonoids at different stages and tissues, including one PAL

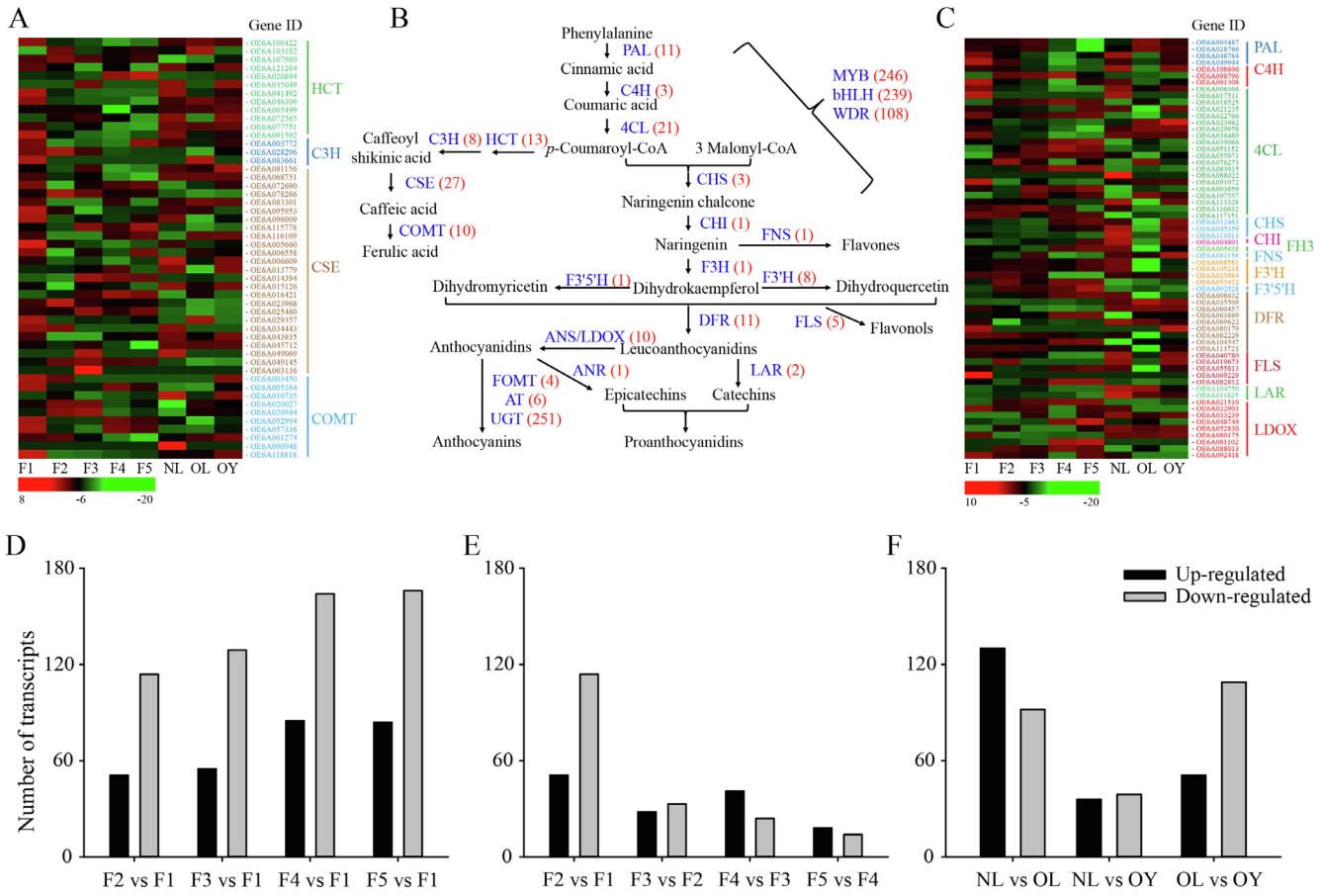

Fig. 2. Biosynthesis of flavonoids in olive and DEGs involved in flavonoid biosynthesis at different developing stages of olive fruit and leaves. (A) Transcript abundance profiles of enzyme genes (HCT, C3H, CSE, COMT) involved in flavonoid synthesis in the different tissues. (B) Simplified diagram of the flavonoid biosynthetic pathways. Numbers in brackets represent gene copy numbers. (C) Transcript abundance profiles of enzyme genes (PAL, C4H, 4CL, CHS, CHI, F3H, FNS, F3′H, F3′5′H, DFR, FLS, LAR, LDOX) involved in flavonoid synthesis in the different tissues. The description of gene abbreviations is provided in Supplementary Table S1. PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate CoA ligase; HCT, p-hydroxycinnamoyl CoA quinate shikimate p-hydroxycinnamoyl transferase; C3H, p-coumarate 3-hydroxylase; CSE, caffeoyl shikimate esterase; COMT, caffeic acid O-methyl transferase; CHS, chalcone synthase; CHI, chalcone isomerase; UGT, UDP- glucosyltransferases; F3H, flavanone 3-hydroxylase; F3′H, flavanone 3′-hydroxylase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; ANR, anthocyanidin reductase. (D) The number of up- and down-regulated unigenes using F1 as a control. (E) The number of up- and down-regulated unigenes between two adjacent developmental stages in olive fruit. (F) The number of up- and down-regulated unigenes between three different developing stages of olive leaf.

(OE6A003487), one C4H (OE6A098796), three 4CLs (OE6A018525, OE6A051152, OE6A116632), one DFR (OE6A113723), one LAR (OE6A011825), one HCT (OE6A072565), one C3H (OE6A028296), eight CSEs (OE6A028296, OE6A013779, OE6A014394, OE6A016421, OE6A023968, OE6A049069, OE6A049145, OE6A063136), and two CSEs (OE6A020027, OE6A020944). They had similar expression patterns with fruit ripening, and their expression curves were bell curves. As fruit development progressed, their expression levels increased gradually, peaking in F3/F4 and then decreasing in F4 and F5. The trend was consistent with flavone accumulation (Fig. 1) and pigment biosynthesis during fruit growth in previous studies in many plants (Deshmukh et al., 2018; Jiang, Song, He, Wang, & Liu, 2018). The coordinated relationship between these transcripts and the flavone level implies that these genes have significant functions in flavonoid biosynthesis in olive.

CHS is the first rate-limiting enzyme in the flavonoid biosynthesis pathway, and a higher level of this enzyme is directly correlated with a higher level of flavonoids (Zuk et al., 2016). We identified three CHS in this study: OE6A012483, OE6A045359, and OE6A113013. The expression level of transcript OE6A113013 was much lower than OE6A012483 and OE6A045359 in the olive fruits. Notably, OE6A045359, which encodes CHS, was most highly (and quite specifically) expressed in NL (Fig. 2C), implying that this transcript played a pivotal role in flavonoid metabolism in the new leaves. DFR provides one entry step in the biosynthesis of anthocyanins and may utilize any one or three of the dihydroflavonols, resulting in the formation of corresponding leucoanthocyanidins (Wu et al., 2016). Of the 11 transcripts that encode DFR, OE6A035509 and OE6A060457 exhibited the same expression trend and both were more highly expressed in the leaves than in the fruits, which were consistent with the accumulation of luteolin, apigenin, and quercetin. Further studies should focus on these two transcripts of DFR, which may constitute genes that control the key nodes of flavonoid metabolism.

The expression of structural genes closely associated with flavonoid biosynthesis are regulated by three transcription factor (TF) families: MYB, bHLH (basic helix-loop-helix, B), and WDR (WD-repeat, W). We detected 204 and 144 TF-regulated DEGs in the fruits and leaves, respectively, during the different developmental stages. Comparative analysis indicated that 89 TF-regulated DEGs existed both in the fruits and leaves, whereas 115 DEGs were expressed only in the fruit (fruit TFs), and 55 DEGs were expressed exclusively in the leaves (leaf TFs) (Table S5). These unique TFs may be involved in the regulation of flavonoid synthesis and metabolism in olive fruits and leaves. Our discovery of a large number of fruit- and leaf-specific TFs implies that they participate in the regulation of physiological processes exclusive to certain tissues. Additionally, the large number of genes encoding

putative fruit TFs involved in a flavonoid biosynthesis suggests that a key role is played by transcriptional regulation during fruit maturation in olive. MYB (OE6A048576, OE6A056119) and bHLH (OE6A032268, OE6A036706) were highly expressed in the fruits compared to the leaves, whereas MYB (OE6A077298, Novel02042) and WDR (OE6A000876) exhibited opposite expression. The expression analysis showed that the TF family exhibits functional differentiation and, thus, further work is necessary to ascertain the role and regulatory mechanism of these TFs.

3.3. Identification of transcripts putatively involved in oleuropein biosynthesis in olive

Oleuropein is a secoiridoid. The biosynthetic steps leading to secoiridoid formation are still unclear because the genes are synthesized only in a restricted number of species, such as *Catharanthus roseus* (Murata & De Luca, 2005), and as a consequence, the enzymes involved in these pathways remain uncertain. However, the pathway from geranyl diphosphate to secologanin has been elucidated (Mougiou et al., 2018), but the follow-up reactions are unclear. Based on this known pathway, we detected some of the required enzymatic functions by observing proposed biosynthetic steps, and a total of 101 transcripts of the nine gene families involved in oleuropein biosynthesis were identified (Fig. 3). Two transcripts of GES, four of 7-DLGT, one of SLS, three of LAMT, and one of 7-DLH were not detected in the olive fruit and leaf databases. Excluding these, there were 90 remaining transcripts involved in oleuropein biosynthesis, among which 7-DLH included the most transcripts (30). The number of differentially expressed transcripts for oleuropein biosynthesis was 62. Our discovery in olive will facilitate the elucidation of oleuropein pathways in other plants that possess oleuropein, such as *Syringa dilatata* and *Osmanthus asiaticus* (Oh et al., 2003; Sugiyama, Machida, Matsuda, & Kikuchi, 1993).

The transcripts of one GES, four G8Hs, three 8-HGOs, one IS, one 7-DLGT, three SLSs, one LAMT, three 7-DLHs, and three IOs (underlined in Fig. 3) declined in expression, which was in accordance with the pattern of oleuropein levels during fruit development (Table 1). Alagna et al. (2009) analyzed two cultivars with different oleuropein contents using comparative transcriptomics, but were unable to elucidate the molecular basis of oleuropein accumulation. It may be addressed by functional annotation of the enzymes involved in oleuropein biosynthesis (Alagna et al., 2009). Hence, the above transcripts may have important roles in oleuropein biosynthesis and could potentially be used as targets for the identification of molecular markers in olive breeding programs aimed at increasing oleuropein production. In addition, some genes related to oleuropein biosynthesis were more highly expressed in the olive leaves than the fruits, such as OE6A119166 (GES), OE6A001750 (G8H), OE6A051723 (IS), and OE6A090361 (LAMT). SLS is a P450 enzyme that catalyzes an unusual ring-opening reaction of loganin in the biosynthesis of secologanin (Bernonville et al., 2015). Notably, the transcript OE6A039927 encoding SLS exhibited the highest FPKM values in the leaves and during the first developmental stage of the fruit. A previous study showed that in *C. roseus*, SLS could be detected preferentially in leaf epidermal cells, and leaf epidermal cells are biosynthetically competent at producing secologanin precursors. Thus the transcript OE6A039927 of SLS may be important for oleuropein biosynthesis. More experimental evidence for this phenomenon is required (Murata and De Luca, 2005). Among the 31 transcripts of DLH, we found that some showed a low expression level at the first sampling and increased later, including OE6A028619, OE6A038177, and OE6A073937. The reason for this may be that they are putatively involved in the biosynthesis of other secondary compounds, such as terpene indole alkaloids, rather than playing a major role in oleuropein synthesis (Miettinen et al., 2014).

3.4. Identification of transcripts putatively involved in HT biosynthesis in olive

A schematic representation of the putative olive HT and tyrosine

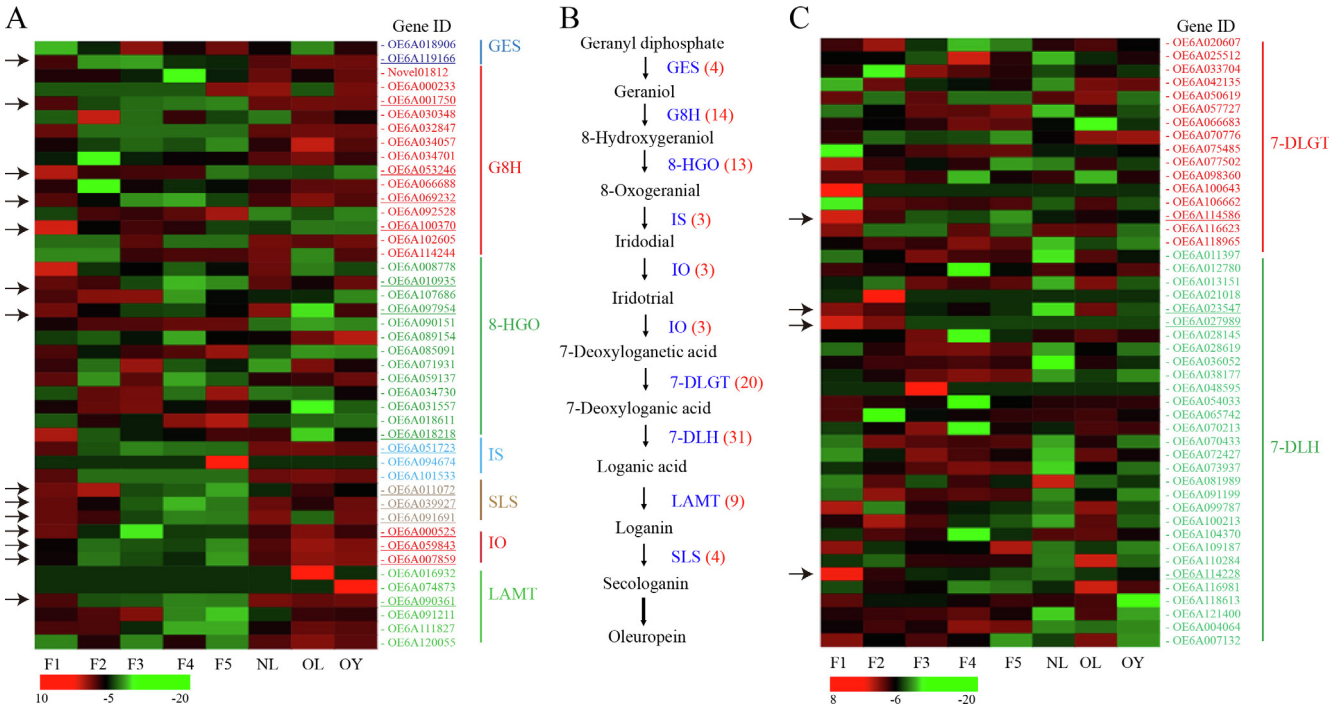

Fig. 3. Biosynthesis of oleuropein in olive. Simplified diagram showing the oleuropein biosynthetic pathways. Numbers in brackets represent gene copy numbers. Underlines and arrows represent the transcripts with notable expression patterns. The heatmaps indicate the transcript abundance profiles of enzyme genes involved in oleuropein synthesis in the different tissues. GES, geraniol synthase; G8H, geraniol 8-hydroxylase; 8-HGO, 8-hydroxygeraniol oxidoreductase, IS, iridoid synthase, IO, iridoid oxidase, 7-DLGT, 7-deoxyloganetic acid-O-glucosyl transferase, 7-DLH, 7-deoxyloganic acid hydroxylase, LAMT, loganic acid methyltransferase; SLS, secologanin synthase.

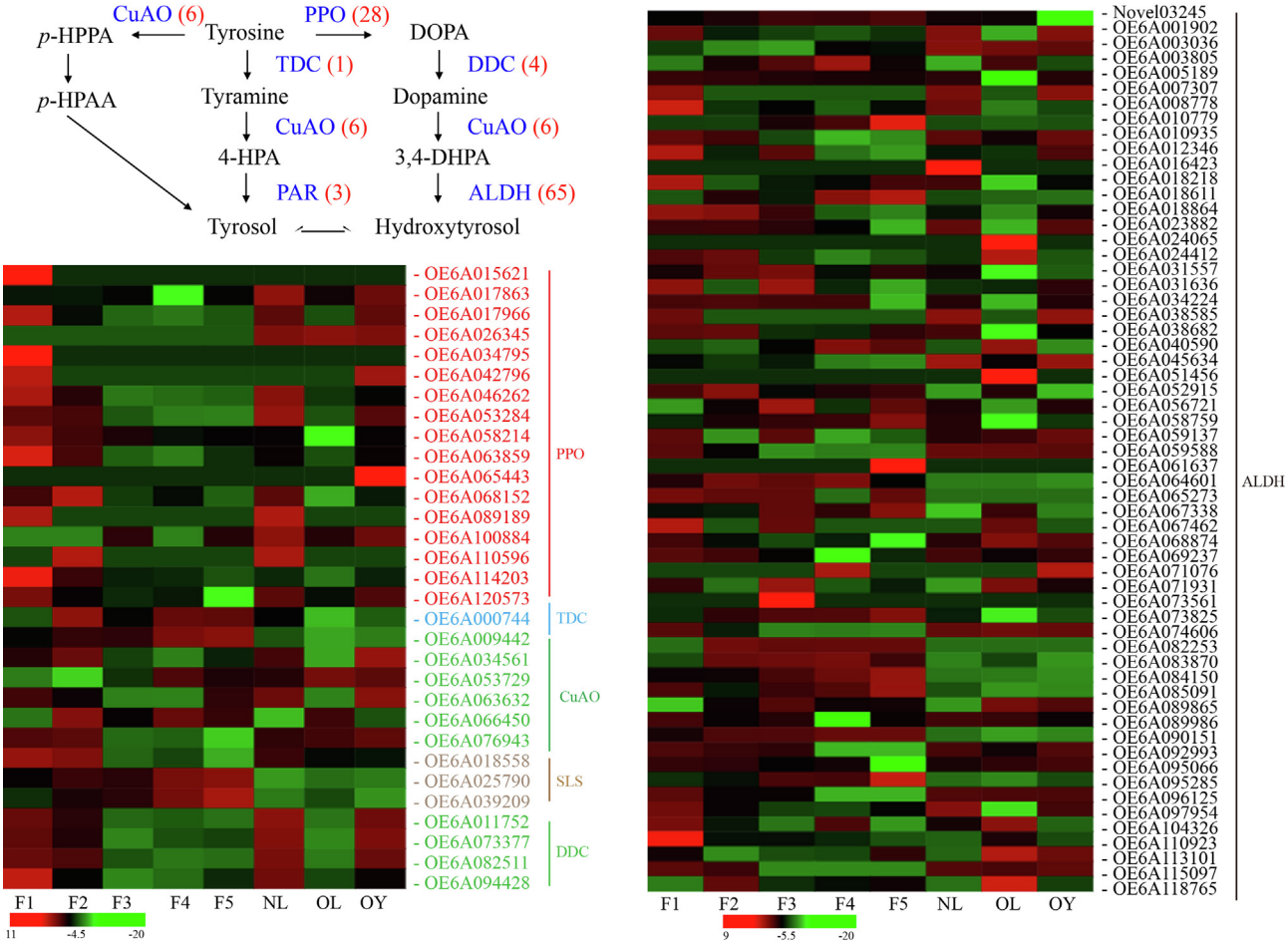

Fig. 4. Biosynthesis of HT in olive. Simplified diagram showing the HT biosynthetic pathways. Numbers in brackets represent gene copy numbers. The heatmaps show the transcript abundance profiles of enzyme genes involved in hydroxytyrosol synthesis in different tissues. PPO, polyphenol oxidase; DDC, DOPA decarboxylase; CuAO, primary-amine (copper-containing) oxidase; ALDH, alcohol dehydrogenase; TDC, tyrosine decarboxylase, PAR, phenyl-acetaldehyde reductase.

biosynthetic pathway is proposed in Fig. 4. One hundred and six transcripts were found to be involved in this pathway. After removing the unexpressed genes, there were 90 transcripts remaining. Fifty-one transcripts were differentially expressed during fruit development, while 35 transcripts were differentially expressed among the different leaf samples. A total of 60 DEGs were discovered in the fruits and leaves.

In heterologous systems, tyrosol is converted to HT by tyrosinase (TYR) (Satoh, Tajima, Munekata, Keasling, & Lee, 2012). However, we did not find any enzyme that catalyzed this step in the present study, and thus the exact biosynthetic pathway that leads from tyrosine to HT in olive is yet to be clarified. ALDH is a small polygenic family that has played an important role in the development of population biology and evolutionary genetics (Thompson, Fernandes, de Souza, de Freitas, & Salzano, 2010). A total of 65 and 59 ALDHs were identified and expressed in the olive samples and possessed the largest number of transcripts among the genes involved in HT biosynthesis. Therefore, the ALDH data in this study provides essential background information for future specific structure-function and evolutionary investigations. The findings discussed above suggested that the main pathway of HT synthesis in olive is via PPO, DDC, CuAO, and ALDH (Fig. 4).

The transcript OE6A025790, which encodes PAR, was highly expressed in the olive fruits. The FPKM values all exceeded 1000 during the five stages of fruit development. We found that the expression values increased as the fruit grew, with expression peaking during the last developmental stage (3720). PAR catalyzes the production of tyrosol, the precursor molecule of HT, and thus future research might focus on this transcript. The expression profile of all four transcripts of the gene *DDC* and some transcripts (17) of the gene *ALDH* were correlated with the HT content and decreased during fruit development. CuAO oxidizes various substrates containing biologically active amine groups and produces the corresponding amino aldehydes, H_2O_2 and NH_3 (Sebela, Frebort, Lemr, Brauner, & Pec, 2000). PPO catalyzes the oxidation reaction in the presence of oxygen, and it might participate in different steps of phenolics metabolism (Ortega-Garcia, Blanco, Peinado, & Peragon, 2008). OE6A068152 and OE6A034561 mRNA levels, putatively coding for two key enzymes PPO and CuAO, were analyzed, and their expression peaked at F2.

3.5. *Identification of alternative spliced transcripts involved in the polyphenol biosynthetic pathway in olive*

Full-length transcript sequences allow for the effective analysis of alternative splicing (AS). Of all the genes related to flavonoid metabolism, oleuropein biosynthesis, and HT biosynthesis, we found 232 alternative transcripts of 18 genes. The AS events include seven known types, namely intron retention (IR), exon skipping (ES), alternative 5′ splice site donor (A5), alternative 3′ splice site acceptor (A3), alternative first exon (AF), alternative last exon (AL), and mutually exclusive exon (MX). Compared with humans, the most common type of AS event in plants appears to be IR, and the abundance of IR is conserved in diverse plants (Xu et al., 2016). In our study, the most predominant and rarest types of AS events in all samples were IR and MX, respectively (Fig. S5). The identification of these alternative splicing events is

beneficial for our understanding of the molecular mechanisms of polyphenol biosynthesis in olive.

Eight candidate genes and three transcription factors related to flavonoid metabolism with AS events were detected. These genes included 4CL (1), CHS (1), F3′H (1), FOMT (1), GT (23), PAL (1), HCT (2), CSE (10), MYB (37), bHLH (24), and WDR (85) (Supplementary Table S6). 4CL (OE6A113329), CHS (OE6A045359), and F3′H (OE6A068581) were found to express alternatively spliced isoforms, and the alternatively spliced junctions were characterized as IR. 4CL expressed three isoforms, whereas CHS, FOMT, PAL and F3′H each expressed two isoforms. The three isoforms of 4CL have five, six, and seven exons, respectively. As suggested in metazoans (Kalsotra & Cooper, 2011) and *Brassica rapa* (Tong et al., 2013), the AS events in regulation-related TFs, MYB, bHLH, and WDR may serve as important and prevalent mechanisms for these TF-encoding genes to form protein complexes that function in transcription in olive. Three genes related to oleuropein biosynthesis exhibited AS events: GES, 8-HGO, and 7-DLH. Among them, GES (OE6A119166) and 7-DLH (OE6A073937 and OE6A121400) had two isoforms, whereas 8-HGO (OE6A031557) and 7-DLH (OE6A013151) had three isoforms, and 7-DLH (OE6A038177) had four isoforms (Supplementary Table S5). In terms of HT biosynthesis, the genes TDC, CuAO, PAR, and ALDH exhibited AS events. TDC (OE6A000744) expressed two isoforms, and these two isoforms have 12 and 13 exons, respectively. CuAO is critical in HT biosynthesis. In the six CuAOs found in this study, two had AS events: OE6A034561 and OE6A076943. They had three and two isoforms, respectively, all of which are potentially interesting. It has been shown that the abundance and activity of splicing factors affects the AS events of target genes (Sun & Xiao, 2015). Zhu et al. (2018) proposed the hypothesis that gene function not only depends on conventional full-length transcripts, but also on the extraordinary AS transcripts. Additionally, AS may act as a major mediator at the post-transcriptional level during plant development (Zhu et al., 2018). Therefore, the newly identified AS events of some genes involved in polyphenol synthesis in this study should be further studied to test this hypothesis.

4. Conclusion

Polyphenols at five different developmental stages of olive fruits and three different olive-leaf ages were tested in this study. The most abundant polyphenolic compound found in the olive fruits was also maslinic acid, with concentrations reaching to 1000 ng/mg fresh weight (FW). Three main polyphenols (namely flavonoid, oleuropein, and hydroxytyrosol) biosynthesis pathway genes were analyzed for the first time by using a combined SGS and SMRT sequencing approach. Our data list more transcripts in these three biosynthesis pathway of olive than previous studies, and indicate the combination of SGS and SMRT sequencing can increase data acquisition. A total of 122 transcripts of 17 gene families, 101 transcripts of the 9 gene families, and 106 transcripts of 6 gene families that encode for enzymes involved in flavonoid, oleuropein, and HT biosynthesis were separately identified. 232 alternative splicing events of 18 genes related to polyphenol synthesis were also identified in this study. This is the first time that the third generation full-length transcriptional sequencing technology has been used to study polyphenol synthesis and related gene expression patterns in olive, which provides a valuable resource for future research on gene discovery, molecular breeding, and metabolic engineering in olive.

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.

Acknowledgments

This work was supported by The National Key R&D Program of China (2018YFD1000603-3) and Fundamental Research Funds for the Central Non-profit Research Institution of Chinese Academy of Forestry (CAFYBB2018QB001). We thank Suzhou Bionovogene (http://www.bionovogene.com) for its help in data analysis. We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2019.125246.

References

Abdelghany, S. E., Hamilton, M., Jacobi, J. L., Ngam, P., Devitt, N., Schilkey, F., & Reddy, A. S. N. (2016). A survey of the sorghum transcriptome using single-molecule long reads. *Nature Communications, 7*, 11706.

Alagna, F., D'Agostino, N., Torchia, L., Servili, M., Rao, R., Pietrella, M., ... Perrotta, G. (2009). Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics, 10*(1), 399.

Bernonville, T. D. D., Foureau, E., Parage, C., Lanoue, A., Clastre, M., Londono, M. A., ... Besseau, S. (2015). Characterization of a second secologanin synthase isoform producing both secologanin and secoxyloganin allows enhanced de novo assembly of a *Catharanthus roseus* transcriptome. *BMC Genomics, 16*(1), 619.

Bouaziz, M., Jemai, H., Khabou, W., & Sayadi, S. (2010). Oil content, phenolic profiling and antioxidant potential of Tunisian olive drupes. *Journal of the Science of Food and Agriculture, 90*(10), 1750–1758.

Brahim, S. B., Kelebek, H., Ammar, S., Abichou, M., & Bouaziz, M. (2017). LC-MS phenolic profiling combined with multivariate analysis as an approach for the characterization of extra virgin olive oils of four rare Tunisian cultivars during ripening. *Food Chemistry, 229*, 9–19.

Chen, J., Tang, X., Ren, C., Wei, B., Wu, Y., Wu, Q., & Pei, J. (2018). Full-length transcriptome sequences and the identification of putative genes for flavonoid biosynthesis in safflower. *BMC Genomics, 19*(1), 548.

Cheng, C., Wang, D., Xia, H., Wang, F., Yang, X., Pan, D., ... Yang, Y. (2018). A comparative study of the effects of palm olein, cocoa butter and extra virgin olive oil on lipid profile, including low-density lipoprotein subfractions in young healthy Chinese people. *International Journal of Food Science and Nutrition,* 1–12.

Cruz, F., Julca, I., Gomez-Garrido, J., Loska, D., Marcet-Houben, M., Cano, E., & Gabaldon, T. (2016). Genome sequence of the olive tree, Olea europaea. *Gigascience, 5*, 29.

Deshmukh, A. B., Datir, S. S., Bhonde, Y., Kelkar, N., Samdani, P., & Tamhane, V. A. (2018). De novo root transcriptome of a medicinally important rare tree Oroxylum indicum for characterization of the flavonoid biosynthesis pathway. *Phytochemistry, 156*, 201–213.

El, S. N., & Karakaya, S. (2009). Olive tree (Olea europaea) leaves: Potential beneficial effects on human health. *Nutrition Reviews, 67*(11), 632–638.

Eleni, M., Zweigenbaum, J. A., & Mitchell, A. E. (2015). Ultrahigh-pressure liquid chromatography triple-quadrupole tandem mass spectrometry quantitation of polyphenols and secoiridoids in california-style black ripe olives and dry salt-cured olives. *Journal of Agricultural & Food Chemistry, 63*(9), 2400–2405.

Fukumitsu, S., Villareal, M. O., Aida, K., Hino, A., Hori, N., Isoda, H., & Naito, Y. (2016). Maslinic acid in olive fruit alleviates mild knee joint pain and improves quality of life by promoting weight loss in the elderly. *Journal of Clinical Biochemistry and Nutrition, 59*(3), 220–225. https://doi.org/10.3164/jcbn.16-40.

Iaria, D. L., Chiappetta, A., & Muzzalupo, I. (2015). A De novo Transcriptomic Approach to Identify Flavonoids and Anthocyanins "Switch-Off" in Olive (Olea europaea L.) Drupes at Different Stages of Maturation. *Frontiers in Plant Science, 6*, 1246. https://doi.org/10.3389/fpls.2015.01246.

Jiang, Y., Song, H. Y., He, J. R., Wang, Q., & Liu, J. (2018). Comparative transcriptome analysis provides global insight into gene expression differences between two orchid cultivars. *PLoS One, 13*(7), e0200155.

Kalsotra, A., & Cooper, T. A. (2011). Functional consequences of developmentally regulated alternative splicing. *Nature Reviews Genetics, 12*(10), 715–729. https://doi.org/10.1038/nrg3052.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*.

Miettinen, K., Dong, L., Navrot, N., Schneider, T., Burlat, V., Pollier, J., & Ilc, T. (2014). Corrigendum: The seco-iridoid pathway from *Catharanthus roseus*. *Nature Communications, 5*(4), 3606–3616.

Mougiou, N., Trikka, F., Trantas, E., Ververidis, F., Makris, A., Argiriou, A., & Vlachonasios, K. E. (2018). Expression of hydroxytyrosol and oleuropein biosynthetic genes are correlated with metabolite accumulation during fruit development in olive, Olea europaea, cv Koroneiki. *Plant Physiology and Biochemistry, 128*, 41–49.

Murata, J., & De Luca, V. (2005). Localization of tabersonine 16-hydroxylase and 16-OH tabersonine-16-O-methyltransferase to leaf epidermal cells defines them as a major

site of precursor biosynthesis in the vindoline pathway in *Catharanthus roseus*. *The Plant Journal, 44*(4), 581–594.
Oh, H., Ko, E. K., Kim, D. H., Jang, K. K., Park, S. E., Lee, H. S., & Kim, Y. C. (2003). Secoiridoid glucosides with free radical scavenging activity from the leaves of *Syringa dilatata*. *Phytotherapy Research, 17*(4), 417–419.
Ortega-Garcia, F., Blanco, S., Peinado, M. A., & Peragon, J. (2008). Polyphenol oxidase and its relationship with oleuropein concentration in fruits and leaves of olive (*Olea europaea*) cv. 'Picual' trees during fruit ripening. *Tree Physiology, 28*(1), 45–54.
Parra, R., Paredes, M. A., Sanchez-Calle, I. M., & Gomez-Jimenez, M. C. (2013). Comparative transcriptional profiling analysis of olive ripe-fruit pericarp and abscission zone tissues shows expression differences and distinct patterns of transcriptional regulation. *BMC Genomics, 14*(1), 866.
Pauwels, E. K. (2011). The protective effect of the Mediterranean diet: Focus on cancer and cardiovascular risk. *Medical Principles and Practice, 20*(2), 103–111. https://doi.org/10.1159/000321197.
Raederstorff, D. (2009). Antioxidant activity of olive polyphenols in humans: A review. *International Journal for Vitamin and Nutrition Research, 79*(3), 152–165. https://doi.org/10.1024/0300-9831.79.3.152.
Saimaru, H., & Orihara, Y. (2010). Biosynthesis of acteoside in cultured cells of *Olea europaea*. *Journal of Natural Medicines, 64*(2), 139–145.
Satoh, Y., Tajima, K., Munekata, M., Keasling, J. D., & Lee, T. S. (2012). Engineering of L-tyrosine oxidation in Escherichia coli and microbial production of hydroxytyrosol. *Metabolic Engineering, 14*(6), 603–610. https://doi.org/10.1016/j.ymben.2012.08.002.
Sebela, M., Frebort, I., Lemr, K., Brauner, F., & Pec, P. (2000). A study on the reactions of plant copper amine oxidase with C3 and C4 aliphatic diamines. *Archives of Biochemistry and Biophysics, 384*(1), 88–99. https://doi.org/10.1006/abbi.2000.2081.
Sugiyama, M., Machida, K., Matsuda, N., & Kikuchi, M. (1993). A secoiridoid glycoside from *Osmanthus asiaticus*. *Phytochemistry, 34*(4), 1169–1170.
Sun, Y., & Xiao, H. (2015). Identification of alternative splicing events by RNA sequencing in early growth tomato fruits. *BMC Genomics, 16*, 948. https://doi.org/10.1186/s12864-015-2128-6.
Thompson, C. E., Fernandes, C. L., de Souza, O. N., de Freitas, L. B., & Salzano, F. M. (2010). Evaluation of the impact of functional diversification on Poaceae, Brassicaceae, Fabaceae, and Pinaceae alcohol dehydrogenase enzymes. *Journal of Molecular Modeling, 16*(5), 919–928. https://doi.org/10.1007/s00894-009-0576-0.
Tong, C., Wang, X., Yu, J., Wu, J., Li, W., Huang, J., & Liu, S. (2013). Comprehensive analysis of RNA-seq data reveals the complexity of the transcriptome in Brassica rapa. *BMC Genomics, 14*, 689. https://doi.org/10.1186/1471-2164-14-689.
Tripoli, E., Giammanco, M., Tabacchi, G., Di Majo, D., Giammanco, S., & La Guardia, M. (2005). The phenolic compounds of olive oil: Structure, biological activity and beneficial effects on human health. *Nutrition Research Reviews, 18*(1), 98–112. https://doi.org/10.1079/NRR200495.
Varet, H., Brillet-Guéguen, L., Coppée, J. Y., & Dillies, M. A. (2016). SARTools: A DESeq2- and EdgeR-based r pipeline for comprehensive differential analysis of RNA-Seq Data. *PLoS One, 11*(6), e0157022.
Visioli, F., & Galli, C. (2001). Antiatherogenic components of olive oil. *Current Atherosclerosis Repotrs, 3*(1), 64–67.
Wu, Q., Wu, J., Li, S. S., Zhang, H. J., Feng, C. Y., Yin, D. D., & Wang, L. S. (2016). Transcriptome sequencing and metabolite analysis for revealing the blue flower formation in waterlily. *BMC Genomics, 17*(1), 897. https://doi.org/10.1186/s12864-016-3226-9.
Xu, Z., Luo, H., Ji, A., Zhang, X., Song, J., & Chen, S. (2016). Global identification of the full-length transcripts and alternative splicing related to phenolic acid biosynthetic genes in Salvia miltiorrhiza. *Frontiers in Plant Science, 7*, 100. https://doi.org/10.3389/fpls.2016.00100.
Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biology, 11*(2) R14 R14.
Zhang, F. S., Wang, Q. Y., Pu, Y. J., Chen, T. Y., Qin, X. M., & Gao, J. (2017). Identification of genes involved in flavonoid biosynthesis in Sophora japonica through transcriptome sequencing. *Chemistry & Biodiversity, 14*(12).
Zhu, J., Wang, X., Xu, Q., Zhao, S., Tai, Y., & Wei, C. (2018). Global dissection of alternative splicing uncovers transcriptional diversity in tissues and associates with the flavonoid pathway in tea plant (*Camellia sinensis*). *BMC Plant Biology, 18*(1), 266. https://doi.org/10.1186/s12870-018-1497-9.
Zuk, M., Dzialo, M., Richter, D., Dyminska, L., Matula, J., Kotecki, A., ... Szopa, J. (2016). Chalcone Synthase (CHS) gene suppression in flax leads to changes in wall synthesis and sensing genes, cell wall chemistry and stem morphology parameters. *Frontiers in Plant Science, 7*, 894. https://doi.org/10.3389/fpls.2016.00894.