



A *De novo* Transcriptomic Approach to Identify Flavonoids and Anthocyanins “Switch-Off” in Olive (*Olea europaea* L.) Drupes at Different Stages of Maturation

Domenico L. Iaria¹, Adriana Chiappetta² and Innocenzo Muzzalupo^{1,3*}

¹ Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, Centro di Ricerca per l'Olivicoltura e l'Industria Olearia, Cosenza, Italy, ² Dipartimento di Biologia, Ecologia e Scienze della Terra, Università della Calabria, Cosenza, Italy, ³ Dipartimento di Farmacia, Scienze della Salute e della Nutrizione, Università della Calabria, Cosenza, Italy

Highlights

- A *de novo* transcriptome reconstruction of olive drupes was performed in two genotypes
- Gene expression was monitored during drupe development in two olive cultivars
- Transcripts involved in flavonoid and anthocyanin pathways were analyzed in Cassanese and Leucocarpa cultivars
- Both cultivar and developmental stage impact gene expression in *Olea europaea* fruits.

During ripening, the fruits of the olive tree (*Olea europaea* L.) undergo a progressive chromatic change characterized by the formation of a red-brown “spot” which gradually extends on the epidermis and in the innermost part of the mesocarp. This event finds an exception in the Leucocarpa cultivar, in which we observe a destabilized equilibrium between the metabolisms of chlorophyll and other pigments, particularly the anthocyanins whose switch-off during maturation promotes the white coloration of fruits. Despite its importance, genomic information on the olive tree is still lacking. Different RNA-seq libraries were generated from drupes of “Leucocarpa” and “Cassanese” olive genotypes, sampled at 100 and 130 days after flowering (DAF), and were used in order to identify transcripts involved in the main phenotypic changes of fruits during maturation and their corresponding expression patterns. A total of 103,359 transcripts were obtained and 3792 and 3064 were differentially expressed in “Leucocarpa” and “Cassanese” genotypes, respectively, during 100–130 DAF transition. Among them flavonoid and anthocyanin related transcripts such as phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H), 4-coumarate-CoA ligase (4CL), chalcone synthase (CHS), chalcone isomerase (CHI), flavanone 3-hydroxylase (F3H), flavonol 3'-hydrogenase (F3'H), flavonol 3'5'-hydrogenase (F3'5'H), flavonol synthase (FLS), dihydroflavonol 4-reductase (DFR), anthocyanidin synthase (ANS), UDP-glucose:anthocyanidin: flavonoid glucosyltransferase (UFGT) were identified. These results contribute to reducing the current gap in information regarding metabolic processes, including those linked to fruit pigmentation in the olive.

Keywords: *Olea europaea*, flavonoid and anthocyanin pathway, RNA-seq, *de novo* assembly, gene expression

OPEN ACCESS

Edited by:

Jörg Linde,
Leibniz-Institute for Natural Product
Research and Infection Biology
-Hans-Knoell-Institute, Germany

Reviewed by:

Thiruvarangan Ramaraj,
National Center for Genome
Resources, USA
Gabriele Bucci,
San Raffaele Scientific Institute, Italy

*Correspondence:

Innocenzo Muzzalupo
innocenzo.muzzalupo@entecra.it

Specialty section:

This article was submitted to
Bioinformatics and Computational
Biology,
a section of the journal
Frontiers in Plant Science

Received: 07 October 2015

Accepted: 21 December 2015

Published: 19 January 2016

Citation:

Iaria DL, Chiappetta A and
Muzzalupo I (2016) A *De novo*
Transcriptomic Approach to Identify
Flavonoids and Anthocyanins
“Switch-Off” in Olive (*Olea europaea*
L.) Drupes at Different Stages of
Maturation. *Front. Plant Sci.* 6:1246.
doi: 10.3389/fpls.2015.01246

INTRODUCTION

The olive tree (*Olea europaea* L. subsp. *europaea* var. *europaea*) is one of the most important and widespread fruit trees in the Mediterranean area. It belongs to the *Oleaceae* family, which includes 600 species within 25 genera. It is widely distributed on all continents, from temperate areas in the north to sub-tropical regions and from low to high altitudes. Native to Mediterranean regions, *Olea europaea* is the only species within the genus *Olea* that produces edible fruits (Green and Wickens, 1989; Wallander and Albert, 2000; Green, 2002; FAOSTAT, 2008¹). The quality of its products, olive oil and table olives, is highly dependent on the agronomic and organoleptic characteristics of its drupes. These characteristics vary in relation to the genetic traits, varieties, the stage of ripeness, as well as in relation to the different susceptibility to environmental growth conditions (Loumou and Giourga, 2003; Conde et al., 2008).

The genuineness of olive oil is important within the "Mediterranean diet." Several research and epidemiological studies link healthy aspects of its components; in particular, olive oil is known to exert protective effects against vascular disease and the onset of cancer (Vauzour et al., 2010). These features are correlated to the high percentage of monounsaturated fats as well as to the high content of antioxidant compounds such as phenols and tocopherols, which, together with other components, characterize the nutraceutical profile of olive products (Pérez-Jiménez et al., 2007; Bruno et al., 2009; Muzzalupo et al., 2011). Phenolic compounds represent a complex mixture in olive derived products responsible for the anti-atherogenic and anti-cancerogenic effects, and for antioxidant properties (Hashim et al., 2008; Llorente-Cortes et al., 2010; Martinelli and Tonutti, 2012). Despite the importance and uniqueness of olive products, the long juvenile developmental phase and its intrinsic self-incompatibility mechanisms slow down current olive breeding programs, which are still very long. Although the current breeding strategies can now benefit from the availability of new polymorphic genetic markers, characterization of the olive germplasm is still far from complete (Baldoni et al., 2009; Muzzalupo, 2012; Muzzalupo et al., 2014).

Therefore, it is of prime importance to focus research programs toward innovative improvement strategies to support conventional programs. In particular, a wider characterization of genes related to plant product quality and to adaptive mechanisms, could provide new information and tools to support both Marker Aided Selection (MAS) strategies and biotechnological approaches. This would aid the development of new growing techniques to increase productivity and quality of this unique species.

Anthocyanins are the most widely distributed group of pigments in plants. They are synthesized via the phenylpropanoid pathway and are mainly responsible for the mauve, red, blue, and purple colors in flowers, fruits, leaves, seeds, and

other organs in most flowering plants. As one of the most ubiquitous class of flavonoids, anthocyanins possess a multitude of biological roles, including protection against solar exposure and ultraviolet radiation, free radical scavenging and anti-oxidative capacity, defense against many different pathogens, and attraction of predators for seed dispersal. Anthocyanins also play a role in consumer preference for flower and fruit quality, potential food health properties, and related horticultural attributes. As a result, classical breeding, as well as transgene technologies, have been used to enhance or create novel colors in ornamental and food crops (Chalker-Scott, 1999; Schaefer et al., 2004; Takahama, 2004; Stommel et al., 2009).

The enzymes involved in the anthocyanin biosynthetic pathway are well characterized. Many of the genes encoding these enzymes have been cloned and share high sequence similarity across species and exhibit tissue- or development-specific expression. Chalcone synthase (CHS) is the first enzymatic step of the biosynthetic pathway (Coe et al., 1981; Dooner, 1983; Koes et al., 1989, 2005). Subsequently chalcone isomerase (CHI) catalyzes the isomerization of chalcone to naringenin (van Tunen et al., 1988, 1989; Grotewold and Peterson, 1994; Griesbach and Beck, 2005). Flavanone 3-hydroxylase (F3H) converts naringenin into dihydrokaempferol, which is converted to anthocyanins by the action of three enzymes. Dihydroflavonol is first converted to a colorless leucoanthocyanidin by dihydroflavonol 4-reductase (DFR). Leucoanthocyanidins are subsequently converted to colored anthocyanidins by anthocyanidin synthase (ANS) finally, the UDP-glucose-flavonoid 3-O-glucosyltransferase (UFGT) creates the anthocyanin-3-glucoside. Within the path, the CHS is the first and key regulatory enzyme of flavonoid biosynthesis and the DFR is the first committed enzyme of anthocyanin biosynthesis in the flavonoid pathway (Holton and Cornish, 1995; Ramsay and Glover, 2005; Martinelli and Tonutti, 2012).

Despite having been recently studied in different olive cultivars (Alagna et al., 2009; Galla et al., 2009; Martinelli and Tonutti, 2012) the molecular mechanisms involved in the regulation of biosynthesis are still unknown.

Tissue- or developmental-specific expression exhibited by anthocyanin structural genes is controlled by a set of regulatory genes. It is known that MYB, bHLH MYC, and WD40 repeat proteins, interacting together to form a regulatory complex that controls anthocyanin structural genes at the transcriptional level (Dixon et al., 2005; Ramsay and Glover, 2005; He et al., 2008; Tian et al., 2008; Alagna et al., 2009; Galla et al., 2009; Stommel et al., 2009; Martinelli and Tonutti, 2012; Ravaglia et al., 2013; Chiappetta et al., 2015).

It has been suggested that a functional MYB-MYC-WD complex directly binds the cis-element of structural gene through MYB transcription factor, while R-like MYC might bind indirectly via a hypothetical R interaction protein (RIP) (Ramsay and Glover, 2005). R-like MYC is centered in the complex that interacts with a MYB factor with WD proteins on its sides. Together, they activate the entire set of anthocyanin biosynthesis genes (Stommel et al., 2009).

¹FAOSTAT 2008 home page Columbia URL: <http://www.columbia.edu/cgi-bin/cul/resolve?ASL9609>

The aim of this work was to define the main transcriptomic profile differences during olive drupe development and to identify the transcripts involved in flavonoid and anthocyanin metabolism.

We have chosen to analyze the transcriptome profile at 100 and 130 days after flowering (DAF), through an Illumina RNA-seq approach, to identify the transcripts along flavonoids and anthocyanins biosynthetic pathways and to monitor their expression levels during ripening. A *de novo* transcriptome reconstruction of olive fruits was performed together with a full expression analysis between samples from “Leucocarpa,” an olive variety characterized by a switch-off in skin color at full ripeness, and “Cassanese,” used as control plant. Significant differences in flavonoid and anthocyanin transcript expression profiles emerged, both during fruit maturation and in relation to genotypes. Consequently, from the wide array of information obtained, our attention was focused on the identified candidate genes set, the expression of which was confirmed by quantitative PCR. In addition, the expression patterns of different MYB, MYC, and WDR transcriptional activators was compared to CHS, DFR, and ANS genes during fruit ripening (Matus et al., 2009; Ravaglia et al., 2013).

MATERIALS AND METHODS

Plant Materials

Olive drupes, of *Olea europaea* L. Leucocarpa and Cassanese cv were used. Drupes were collected from 20-year-old plants, clonally propagated and belonging to the olive germplasm collection of the Agricultural Research Council—Olive Growing and Oil Industry Research Centre, CREA-OLI in Mirto-Crosia (Cosenza, Calabria, Italy). Olive trees were grown using the same field conditions and were located at latitude 39°37′04.57″N, longitude 16°45′42.00″E and altitude 8 m asl).

Fruit sampling was performed as previously described (Matus et al., 2009): for each cultivar, drupes ($n = 30$) were randomly collected at 100 and 130 DAF (**Figure S1**). In order to minimize the effects related to asynchronous fruits maturation within the same tree, drupes with similar pigmentation were picked from all around the external parts of the tree canopy. Concerning drupe pigmentation, the epimesocarp tissues, was totally green in color at 100 DAF whereas at 130 DAF the pulp pigmentation was 50% brown in “Cassanese” and totally unpigmented in “Leucocarpa” drupes (**Figure S1**).

All samples were fixed in liquid nitrogen and stored at -80°C for both RNA-seq and qRT-PCR experiments.

RNA-Seq Library Preparation and Sequencing

In order to obtain a general overview of the transcripts and metabolic pathways involved in fruit maturation and to avoid cross contamination from non-homogeneous tissue separation, sample pooling strategy has been here used (Peng et al., 2003).

Pooling reduces variability by minimizing individual variation and represents an alternative approach to biological replicates in experiments where the interest is not on the individual but rather on characteristics of the population (e.g., common changes in expression patterns; Karp and Lilley, 2007, 2009).

Total RNA was extracted from the epi-mesocarp tissues of drupes collected together, using the RNeasy Plant Mini kit (Qiagen) according to the manufacturer’s instructions. Each RNA sample was subjected to DNase digestion (DNase I, Roche) to remove any DNA contamination and pooled equally, as previously described (Muzzalupo et al., 2012). RNA was quantified by the NanoDrop Spectrophotometer ND-2000 and quality was checked by electrophoresis (28S rRNA/18S rRNA ratios). Samples with a concentration of ≥ 400 ng/ μl , OD260/280 = 1.8~2.2, RNA 28S:18S ≥ 1.0 , and RNA Integrity Number (RIN) ≥ 7.0 were used for cDNA library preparation.

Standard RNA-seq library preparation and sequencing via Illumina HiSeq TM 2000 was carried out by Technology Services of the Institute of Applied Genomics (IGA, Udine, Italy). For each sample a single-end (SE) sequencing cDNA library was constructed with a fragment length range of 50 bp. Each library was created using two replicates, consisting of a separate pool of 30 homogeneous fruits.

RNA-Seq Data Filter and *De novo* Assembly by Trinity

The raw Fastq “reads” (NCBI PDA/SRA accession numbers: SRR1574719, SRR1574772, SRR1573503, SRR1574328, **Table 1**) were analyzed and filtered, respectively with FastQC and Fastx Toolkit softwares to obtain high quality *de novo* transcriptome sequence data. Each sequence set was filtered using the following criteria: (i) reads containing the sequencing adaptor were removed; (ii) reads with unknown nucleotides comprising more than 5% were removed; (iii) low-quality reads with ambiguous sequence “N” were trimmed and discarded.

Since the olive tree does not have a reference genome, the *de novo* assembly of the clean reads into transcripts was performed using the Trinity program (Grabherr et al., 2011; Haas et al., 2013), a useful method for the efficient and robust *de novo* reconstruction of transcriptomes from RNA-seq data (Ward et al., 2012; Gutierrez-Gonzalez et al., 2013; Liang et al., 2013; Liu et al., 2013; Pallavicini et al., 2013; Tulin et al., 2013).

Trinity was run via script using 128 GB of ram, 12 cpu thread and a minimum assembled contig length to report set to 300 bp.

Trinity sequentially combines Inchworm, Chrysalis and Butterfly modules to process large RNA-seq reads data, partitioning the sequence data into many individual de Bruijn graphs, representing transcriptional complexity at a given gene or locus (Grabherr et al., 2011; Haas et al., 2013).

Analysis of Transcript Assembly

For non-model organisms, one metric for evaluating the transcript assembly quality is to examine the number of transcripts that appear to be full-length or nearly full-length

TABLE 1 | Assembled transcripts for each sample.

Sample	Raw reads	Used reads	Assembled transcripts	Contig N50	Mapped reads
Leucocarpa 100 DAF	28,700,100	23,687,921	22,959	754	84.07%
Leucocarpa 130 DAF	28,121,963	23,122,308	26,203	829	84.15%
Cassanese 100 DAF	28,550,901	23,394,526	22,709	767	83.82%
Cassanese 130 DAF	57,106,631	48,153,012	31,485	972	85.49%

if compared to a closely related organism to examine full-length coverage. In this context, a more general analysis was performed aligning the assembled transcripts against all known plant proteins determining the number of unique top matching proteins that are aligned in 70–100% range of its length by full-length transcript analysis (Haas et al., 2013). Therefore, a blastable database has been created to perform a local blastx search where only the single best matching Trinity transcript was outputted for each top matching entry.

To validate our de novo assembly read remapping has been realized using bowtie2 (Langmead and Salzberg, 2012); for each data set a bowtie2 index was created, and then the number of reads that map to our transcriptome have been counted.

Abundance Estimation and Differentially Expressed Trinity Transcripts

For abundance estimation of transcriptome assemblies RSEM software was used (Li and Dewey, 2011). RSEM is a package for estimating gene and isoform expression levels from RNA-seq data. The current version of RSEM, was bundled with the Trinity software package.

Moreover, Trinity currently supports the use of Bioconductor tools (edgeR and DESeq) to compute differential expression analysis in the assembled transcriptome (Anders and Huber, 2010; Robinson et al., 2010; Grabherr et al., 2011; Haas et al., 2013). In order to identify statistically significant differences in transcript expression between samples, the number of reads/transcripts, the depth of sequencing, the transcripts length (longer transcripts generate more fragment reads) and the expression level of the transcripts were considered. Expression values, normalized for each of these factors were measured in FPKM (fragments per feature kilo base per million reads mapped) (Trapnell et al., 2010; Robinson and Oshlack, 2010) and used to make a comparison across multiple samples and replicates. Trinity supports the use of TMM (trimmed mean of *M*-values) normalization (Lekanne Deprez et al., 2002; Dillies et al., 2012), to account for differences in the mass composition of the RNA-seq samples, which does not change the fragment count data, but provides a scaling parameter that yields an effective library size (total map able reads) for each sample. This effective library size is then used in the FPKM calculations.

Quantitative PCR

Gene expression analysis was performed by quantitative real-time PCR on a 7500 fast real time PCR system (Applied

Biosystems) with SYBR[®] Select Master Mix. The oligonucleotide primer sets (Table 1) used for qRT-PCR analysis were designed using Primer3 (<http://primer3.ut.ee/>).

Each primer pair (Supplementary data, Table S1) generated a single specific amplicon on the 3'-end of target sequence. PCR products were about 150–200 bp long and primer pair average efficiency ranged between 0.95 and 1.0. The housekeeping olive ELONGATION FACTOR 1 (EF1) gene (CAQ17046.1) was used to normalize the expression levels (Galla et al., 2009; Trapnell et al., 2010). Amplification reactions were prepared in a final volume of 20 µl according to the manufacturer's instructions.

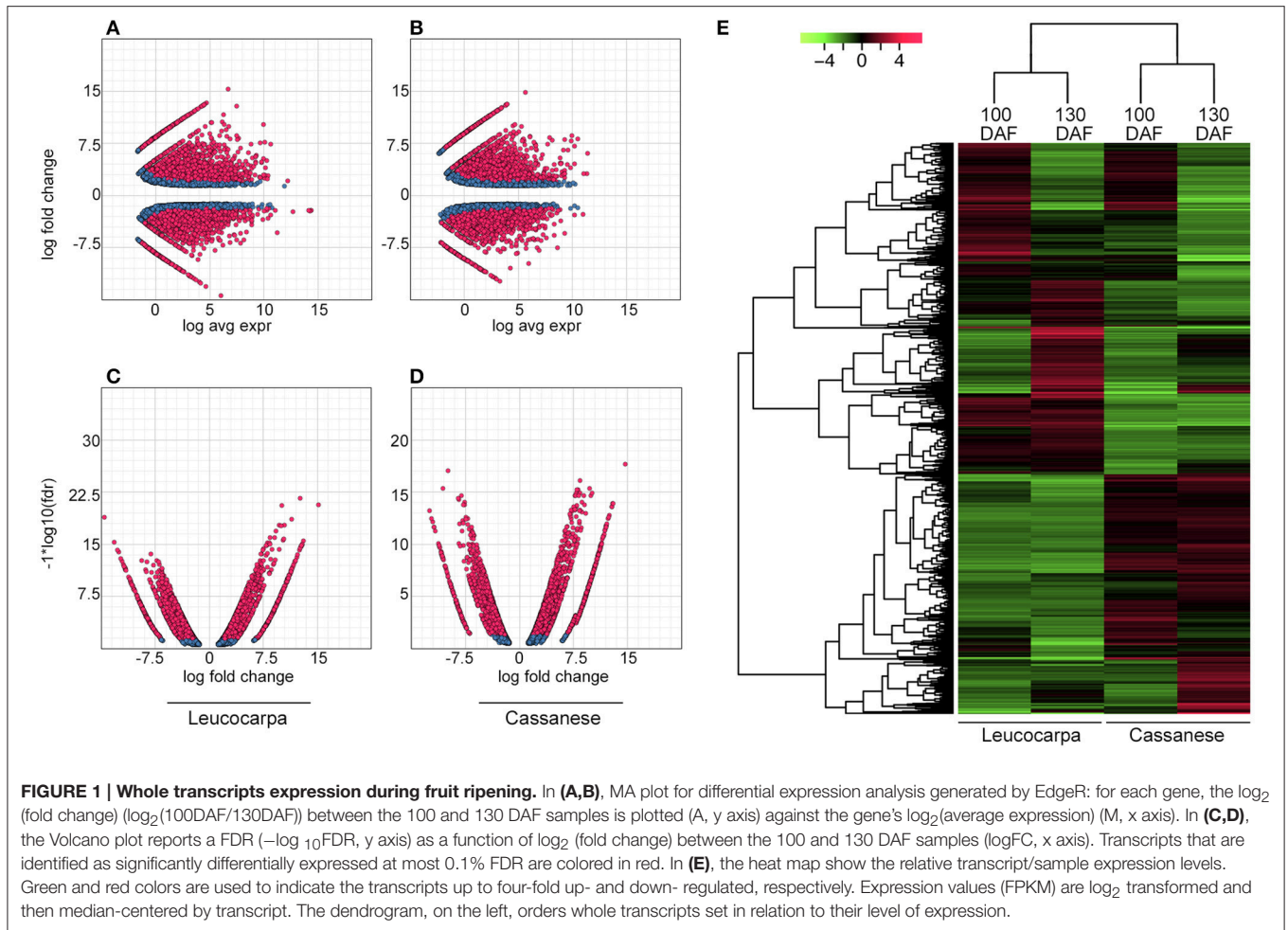
All reactions were run in triplicate in 96-well reaction plates, and negative controls were set. The cycling parameters were as follows: one cycle at 95°C for 3 min to activate the Taq enzyme, followed by 40 cycles of denaturation at 95°C for 10 s and annealing-extension at 58°C for 30 s. To confirm the occurrence of a unique PCR product, the “melting curve” (Lekanne Deprez et al., 2002) was evaluated by an increase of 0.5°C every 10 s within a 60–95°C range and a unique “melting peak” in every reaction was observed. The comparisons of cycle threshold (CT) values were obtained analysing data with the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). The means of gene expression levels were calculated from two biological repeats, obtained from two independent experiments.

Blast2GO

To assign gene ontology (GO) terms in our DE data sets, we used BLASTx 2.2.26+, BLOSUM62 similarity matrix, and Blast2GO database version August 2011 programs (Conesa et al., 2005; Morgulis, 2008). The definition of each GO term was determined by the GO Consortium: <http://www.geneontology.org> and can be found using the EMBL European Bioinformatics Institute QuickGO: <http://www.ebi.ac.uk/QuickGO> or the Gene Ontology Normal Usage Tracking System, GONUTS: http://gowiki.tamu.edu/wiki/index.php/Main_Page.

Pathway assignments were determined following the Kyoto Encyclopedia of Genes and Genomes pathway database (Kanehisa et al., 2008) using BLASTX with an *E*-value threshold of 1.0E-5.

MapMan (<http://mapman.gabipd.org/>) analysis was done using our DE transcripts rearranged as input experimental dataset. Using the Mercator web application we can assign MapMan “Bins” to DNA sequences (Thimm et al., 2004; Lohse et al., 2014). The output was used as a mapping file for data visualization in MapMan. The Mercator tool generates functional predictions by searching a variety of reference



databases (BLAST-based, RPSBLAST based and InterProScan) and subsequently evaluating and compiling the search results for each input gene to propose a functional Bin.

RESULTS

RNA-Seq Library Sequencing and De novo Transcriptome Assembly by Trinity

Starting from four RNA-seq libraries, corresponding to two fruit developmental stages (100 and 130 DAF) of *Olea europaea* “Leucocarpa” and “Cassanese,” 147,789,544 raw reads were generated from 50 bp insert library. A total of 142,479,595 high-quality SE reads were identified and used for *Olea europaea* transcriptome assembly, through the Trinity software. Using the 25-mer in Trinity and a minimum assembled contig length set to 300 bp, we found 103,359 transcripts. The total used reads, the total assembled transcripts, N50 statistics for each sample and remapping results are indicated in Table 1.

A total of 93,623 likely coding sequences were extracted with the Transdecoder utility, to identify the longest ORF (Open Reading Frame) from the transcript assembly, reporting that ORF scored according to the Markov model. In all, 9597 of the

TABLE 2 | Number of differential expressed transcripts during 100–130 DAF transition for each cultivar.

Sample	Total transcripts	DE transcripts
Leucocarpa 100–130 DAF	49,162	3792
Cassanese 100–130 DAF	54,194	3064

olive transcripts had a BLAST hit with an *E*-value of less than $1e-20$, and 19,708 of the extracted reference coding sequences are considered to be approximately “full length,” with the Trinity contigs aligning the matching UniProt reference transcript’s length by more than 70%.

Differential Expression Analysis

To estimate the differential gene expression between fruits of both considered cultivars at each developmental stage, a single assembly, based on combining all reads across all samples as inputs was generated. A single assembly was chosen to avoid difficulty in comparing the results across the different samples, due to differences in assembled transcript lengths and contiguity. Then, reads were aligned separately back to the single assembly, in order to identify the number of DE transcripts with a False

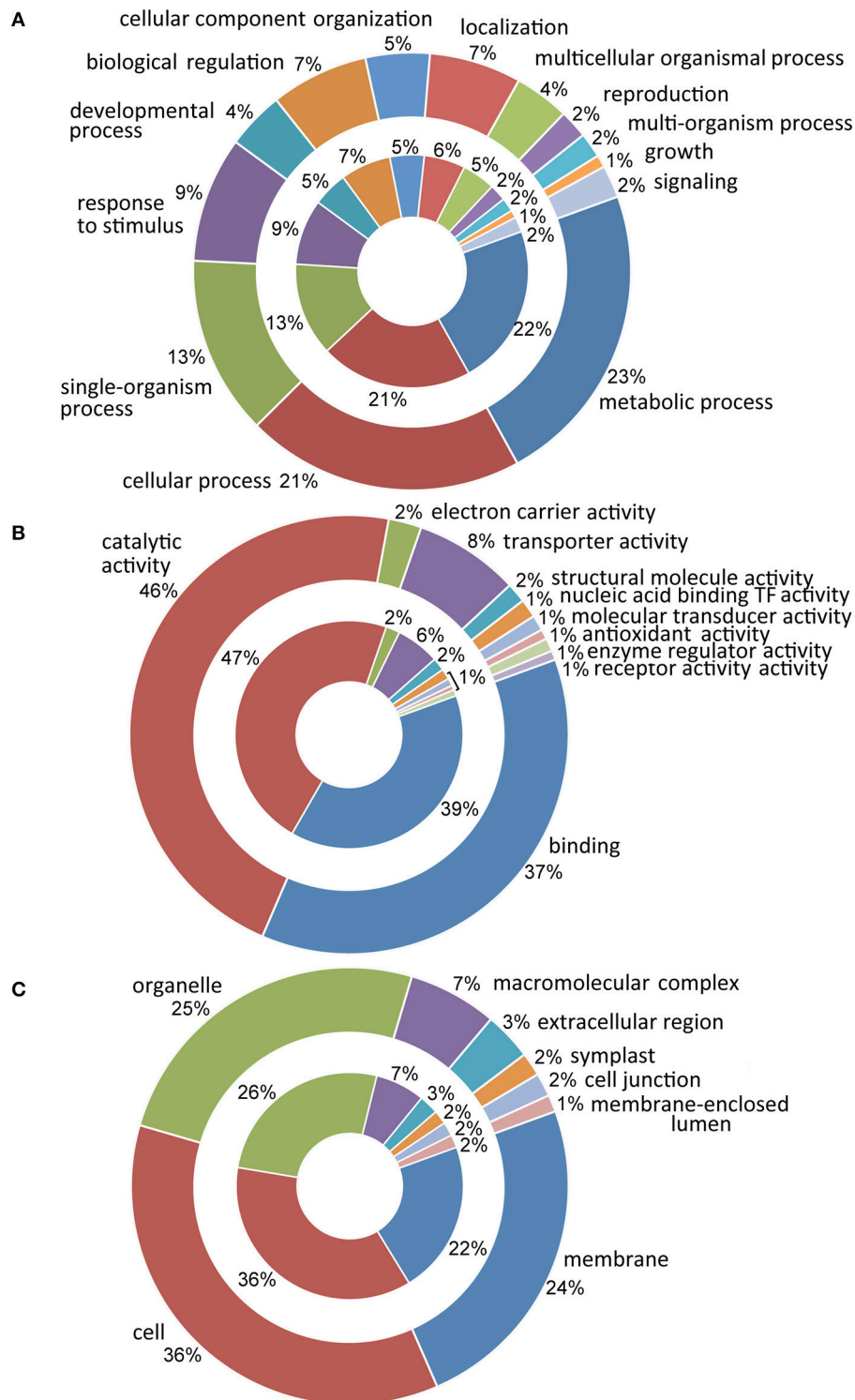
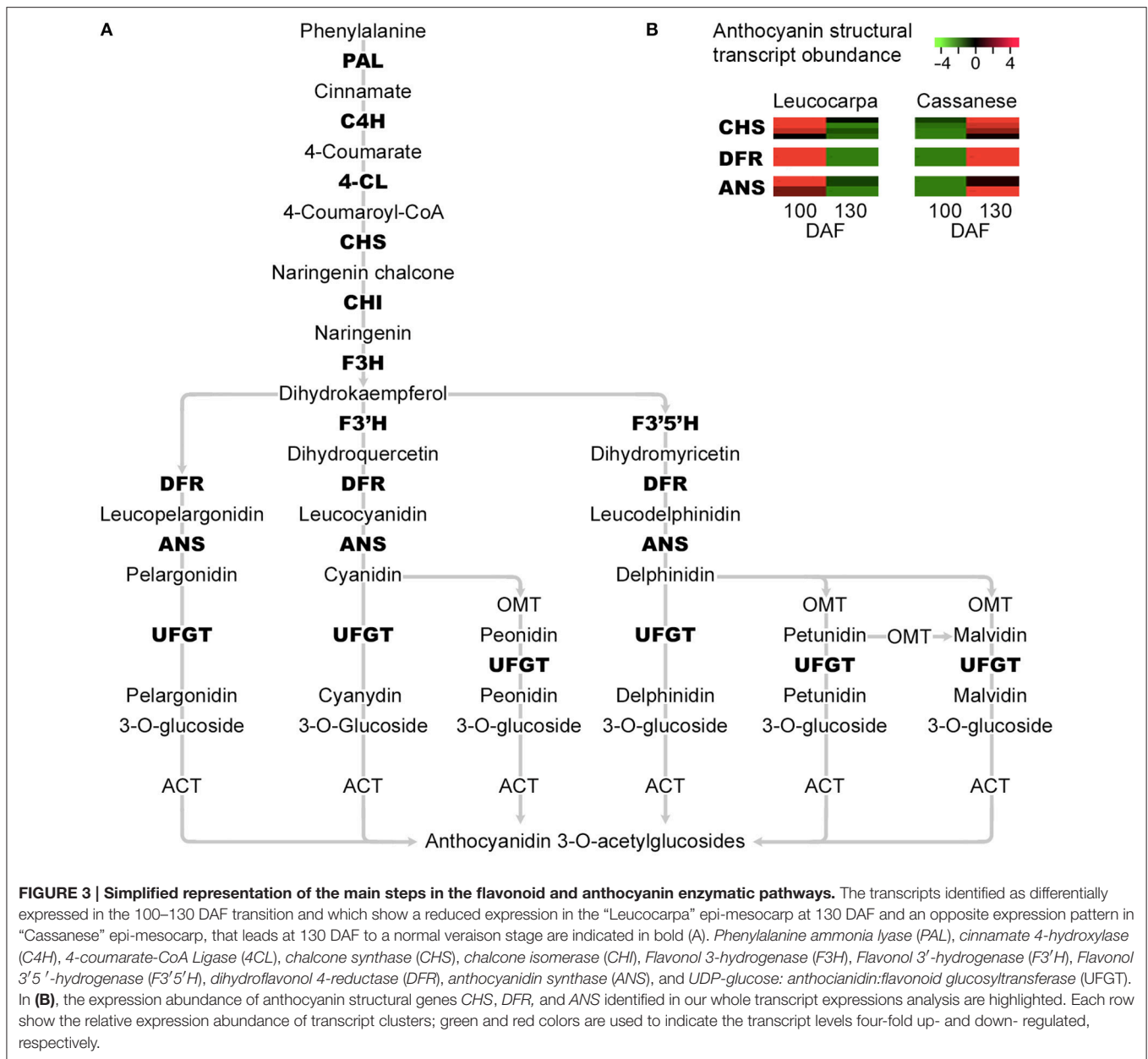


FIGURE 2 | Distribution of ontological categories (level 2 GO terms) in *Leucocarpa* (inner chart) and *Cassanese cvs* (outer chart) DE transcripts according to: biological process (A), molecular function (B) and cellular component (C). In A metabolic process and cellular process are the most represented groupings; the divisions relating to catalytic activity and binding are strongly represented in (B), while in C cell, organelle and membrane categories are represented. At the side of ontological categories, the percentage of the transcripts within each class is reported.



Discovery Rate (FDR) value of at most 0.001 and at least four-fold difference in expression values according to the Trinity protocol.

For this purpose, it was possible to identify the DE transcripts sets of each cultivar, during the 100–130 DAF transition from Trinity scripts that leverage the R software. In this context, 3792 and 3064 DE transcripts (of 49,162 and 54,194 total transcripts, respectively) were identified in “Leucocarpa” and “Cassanese.” The fold change and the statistical significance values between different developmental stage and cultivar were also estimated.

Trinity facilitates analysis of RNA-seq data, including scripts for extracting transcripts that are above some statistical significance (FDR threshold) and fold-change in expression. To examine expression across multiple samples, the FPKM expression values across samples have been normalized, which will

account for differences in RNA composition, afterwards TMM normalization generate a matrix of normalized FPKM values across all samples.

These adjusted library sizes are used to recompute the FPKM expression values. Although the raw fragment counts are used for differential expression analysis, the normalized FPKM values are used below in examining profiles of expression across different samples, each DE set of transcripts was displayed as MA plots (where M = log ratios and A = mean values) (Figures 1A,B), volcano plots (Figures 1C,D) and clustered heat maps (Figure 1E). A correlation matrix (Figure S2) for the different developmental stages across cultivars, reveals that samples are more highly correlated within cultivar than between cultivar.

Functional Annotation of Differentially Expressed Transcript Sets

The *in silico* analysis of the entire sets of DE transcripts, performed by querying databases of genes and proteins (NCBI, ExPASy, InterProScan) and the functional annotation software Blast2GO, have allowed for each sequence to be traced back to the gene family and to be annotated according to the terms of the three main Gene Ontology vocabularies (Figure 2). Since analyses were conducted on the same organ and developmental stages, in both analyzed cultivars a fairly overlapped distribution of GO terms was observed during the developmental transition. In particular, the most represented ontological categories were membrane (GO:0016020), cell (GO:0005623) and organelle (GO:0043226). Molecular functional categories were strongly represented by terms related to catalytic activity (GO:0003824) with 47 and 46% in *Leucocarpa* and *Cassanese* cvs, respectively, followed by binding (GO:0005488) and transporter activity (GO:0005215). Finally, more than 10 categories were identified at the biological process level with metabolic and cellular processes (GO:0008152, GO:0009987), among the groups most represented, highlighting the intense and complex metabolic and regulatory activities during fruit maturation.

In order to trace back to the pathways, such as flavonoids and anthocyanin, (map 00941 and 00942, Figures S3A,B, respectively), which were more closely involved in the transition

between 100 and 130 DAF, the whole DE transcripts set was examined through the Kyoto Encyclopedia of Genes and Genomes (KEGG). Functional analysis was implemented in Mapman, to focus gene expression changes via Image Annotator. All obtained results are consistent with a down regulation of flavonoid and anthocyanins metabolism in *Leucocarpa* cv, while an opposite trend was observed in “*Cassanese*” (Figure S4).

Gene Expression during Olive Fruits Ripening

We performed a quantitative RNA-seq analysis in a cultivar of *Olea europaea* species, whose fruits are characterized by a switch-off in skin color at full ripeness, to identify the genes involved in flavonoid and anthocyanin biosynthesis.

The transcripts set in flavonoid and anthocyanin pathways were identified in our Illumina datasets. It includes 11 transcripts: phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H), 4-coumarate-CoA Ligase (4CL), chalcone synthase (CHS), chalcone isomerase (CHI), flavanone 3-hydroxylase (F3H), flavonol 3'-hydroxylase (F3'H), flavonol 3'5'-hydroxylase (F3'5'H), flavonol synthase (FLS), dihydroflavonol 4-reductase (DFR), anthocyanidin synthase (ANS), UDP-glucose: anthocyanidin:flavonoid glucosyltransferase (UFGT) (Supplementary data, Table S1). Moreover, it was possible to identify different member of MYB, MYC and WD transcription factors related to the regulatory complex that controls anthocyanin structural genes at the transcriptional level (Takahama, 2004).

Interestingly, the quantitative gene expression analysis does not seem to show significant differences during olive fruit development in *Leucocarpa* and *Cassanese* cvs (Table 2). Indeed, focusing attention on the paths that control the biosynthesis of pigments and the natural reduction of photosynthetic pigments during the veraison stage (Pua and Davey, 2010), the “*Leucocarpa*” was characterized by a broad down-regulation of CHS, DFR, and ANS transcripts (Figure 3), during the 100–130 DAF transition compared to *Cassanese* cv.

The estimated fold change of the selected genes was also confirmed by quantitative PCR experiments (Figure 4). In particular, the expression of transcripts putatively involved in the selected pathway were more highly expressed in “*Cassanese*” genotype than in “*Leucocarpa*.”

This genome-wide overview on flavonoid and anthocyanin genes also allowed us to select different members of MYB, MYC, and WD transcription factors (TF), within the differentially expressed gene set, linkable to anthocyanins regulatory circuit (Dixon et al., 2005; He et al., 2008; Tian et al., 2008; Stommel et al., 2009; Jaakola, 2013). The abundance estimation analysis made it possible to compare the identified TFs in all analyzed samples. In the “*Cassanese*” plant, despite a slight decline, the amount of transcripts during 100–130 DAF transition was consistent with the increased anthocyanin structural gene expressions and metabolite accumulation during growth of fruits; whereas in the *Leucocarpa* cv the identified TFs are primarily characterized by lower expression levels and a general reduction in expression abundance during ripening transition. The differences were most evident when the comparison was carried out at the same stage

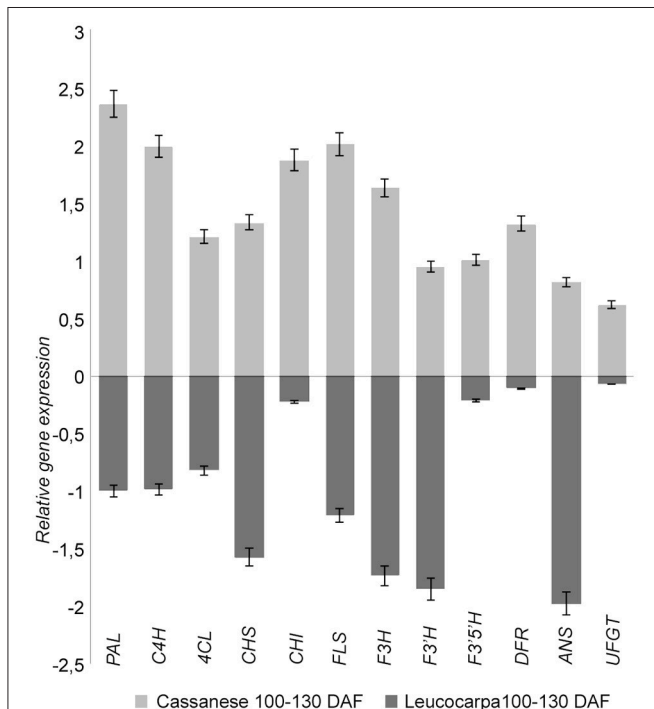
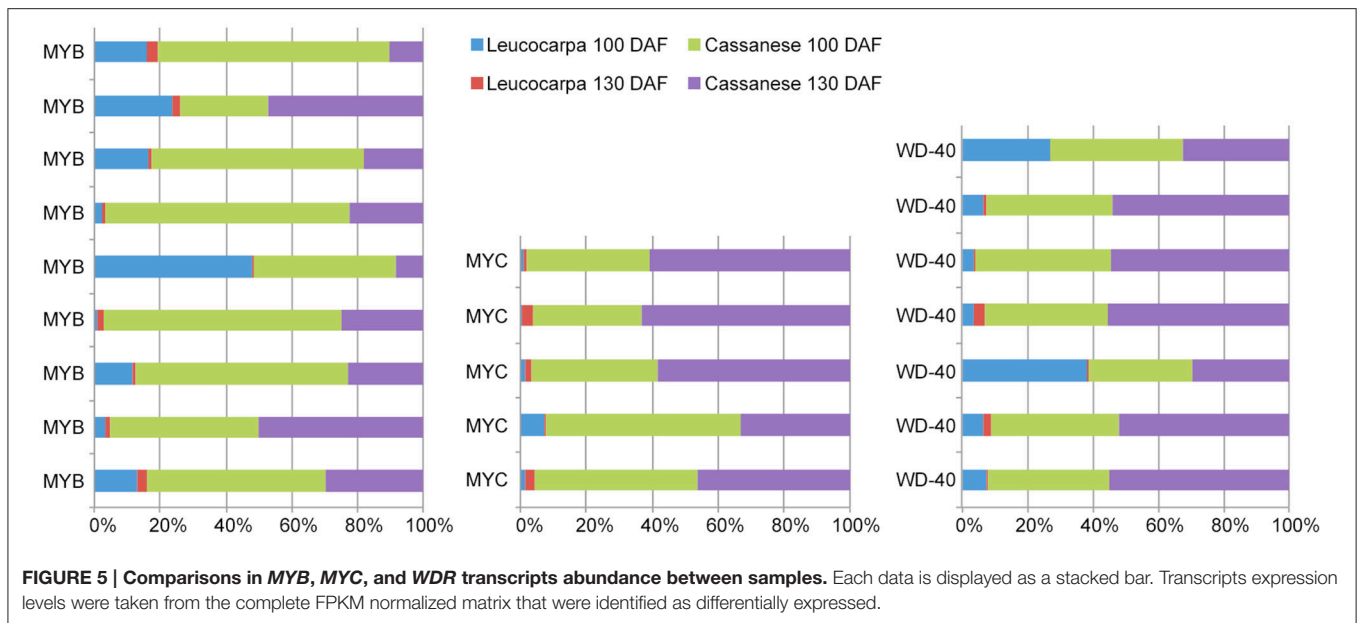


FIGURE 4 | Relative transcripts expression during fruit ripening in *Leucocarpa* (dark gray) and *Cassanese* (light gray) cvs. The qRT-PCR results (log fold change) are presented as a proportion of the highest value after normalization with the EF1 house-keeping gene; for each cv 100 DAF samples are used as calibrator. The means \pm s.e. of two independent biological replicates are reported.



(100 or 130 DAF) of maturation. 9 MYB, 5 MYC, and 7 WD TF undergo a decrease in expression during transition, in contrast to Cassanese cv where they appear to participate in the activation pathway (Figure 5).

DISCUSSIONS

In the present work we used the Illumina RNA-seq technology to identify the transcripts along flavonoids and anthocyanins biosynthetic pathways and to monitor their expression levels during ripening, by comparing two olive cultivars characterized by different phenological behavior at ripening in terms of anthocyanin accumulation and general pigmentation. We also used a *de novo* transcriptome assembly strategies performed in many plants, including rice, maize, sesame, bamboo, poplar, sweet potato, *Eucalyptus* tree, chickpea, and orchid (Mizrachi et al., 2010; Wang et al., 2010; Fu et al., 2011; Garg et al., 2011; Wei et al., 2011; Zhang et al., 2011a).

The characterization of the genetic entity of olive cultivars has benefited from new molecular biology and high-throughput sequencing methods (Alagna et al., 2009; Galla et al., 2009; Bazakos et al., 2012; Muñoz-Mérida et al., 2013). Through the analysis of massive data it is possible to identify/investigate the genetic pathways that underlie specific, or more general, agronomic traits in the physiological performance of the plants belonging to the *Olea europaea* species.

Between different high-throughput methods, the Illumina sequencing is the best next generation technology, both less costly and more efficient, for transcriptome analysis, if compared with 454 platform, in particular when used in non-model organisms, where genomic sequences are unknown.

Even though this technology has been previously restricted to the re-sequencing of organisms with available reference genomes (Nagalakshmi et al., 2008), its recent improvement has enabled the development of *de novo* strategies for robust transcriptome

reconstruction for non-model plants from short reads and their assembly into unigenes.

Through this approach we identified anthocyanin genes, including PAL, C4H, 4CL, CHS, CHI, F3H, F3'H, F3'5'H, FLS, DFR, ANS, from two olive cultivars.

In addition, different transcription factor members with similarity to MYB, MYC, and WD40 family and involved in anthocyanin biosynthesis were also found. Furthermore, the transcripts abundance of identified genes was correlated to the accumulation rate of anthocyanin metabolites.

The anthocyanin biosynthesis pathway has been extensively studied in several plant species, such as petunia, pears, goji berry, bilberry and black raspberry (Jaakola et al., 2002; Zeng et al., 2014). During the ripening progression, many species including the olive tree accumulate anthocyanin in their fruits (Jaakola et al., 2002; Sweetman et al., 2009; Zhang et al., 2011b). In this context, anthocyanins are considered potent marker to monitor ripening stages and organoleptic quality of fruits.

In apple, the regulatory circuit in anthocyanin biosynthesis is tuned by the MYB-MYC-WD40 protein complexes (Ramsay and Glover, 2005; Schaart et al., 2012). Moreover the R2R3-MYB and bHLH TFs are able to activate structural genes, including CHS, DFR and ANS, and ultimately promote anthocyanin accumulation in fruits (Chagné et al., 2013; Umemura et al., 2013; Zeng et al., 2014). In our case the transcripts abundance of MYB, MYC, and WD40-*type* TFs was higher in Cassanese cultivar than in Leucocarpa and was also directly related to anthocyanin accumulation.

In conclusion, the comparative approach performed provide an invaluable resource to identify genes involved in fruit maturation and to define the metabolic pathway and tissue specific functional genomics in non-model plant species. The characterization of transcripts from flavonoid and anthocyanin biosynthetic pathways and the analysis of their expression level in olive fruits is an important goal to understand the veraison

event of fruits and to increase the knowledge on these antioxidant molecules, important for human health.

AUTHOR CONTRIBUTIONS

DI performed research and discussed results. AC designed research analyzed data and discussed result. IM designed research analyzed data and discussed results. All authors contributed to improving the papers and approved the final manuscript.

ACKNOWLEDGMENTS

The authors are very grateful to Dr. Sabrina Micali (CREA-FRU of Roma, Italy) for excellent technical and scientific assistance. This research was supported by the “Certificazione della composizione varietale, dell’origine geografica e dell’assenza di prodotti di sintesi negli oli extravergini di oliva—CERTOLIO 2012-2014” projects and by grant from University of Calabria (MIUR ex 60) to Prof. AC.

REFERENCES

- Alagna, F., D’Agostino, N., Torchia, L., Servili, M., Rao, R., Pietrella, M., et al. (2009). Comparative 454 pyrosequencing of transcripts from olive genotypes during fruit development. *BMC Genomics* 10:399. doi: 10.1186/1471-2164-10-399
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11:R106. doi: 10.1186/gb-2010-11-10-r106
- Baldoni, L., Cultrera, N. G., Mariotti, R., Ricciolini, C., Arcioni, S., Vendramin, G. G., et al. (2009). A consensus list of microsatellite markers for olive genotyping. *Mol. Breeding* 24, 213–231. doi: 10.1007/s11032-009-9285-8
- Bazakos, C., Manioudaki, M. E., Therios, I., Voyiatzis, D., Kafetzopoulos, D., Awada, T., et al. (2012). Comparative transcriptome analysis of two olive cultivars in response to NaCl-Stress. *PLoS ONE* 7:e42931. doi: 10.1371/journal.pone.0042931
- Bruno, L., Chiappetta, A., Muzzalupo, I., Gagliardi, C., Iaria, D., Bruno, A., et al. (2009). Role of geranylgeranyl reductase gene in organ development and stress response in olive (*Olea europaea*). *plants. Funct. Plant Biol.* 36, 370–381. doi: 10.1071/FP08219
- Chagné, D., Lin-Wang, K., Espley, R. V., Volz, R. K., How, N. M., Rouse, S., et al. (2013). An ancient duplication of apple MYB transcription factors is responsible for novel red fruit-flesh phenotypes. *Plant Physiol.* 161, 225–239. doi: 10.1104/pp.112.206771
- Chalker-Scott, L. (1999). Environmental significance of anthocyanins in plant stress responses. *Photochem. Photobiol.* 70, 1–9. doi: 10.1111/j.1751-1097.1999.tb01944.x
- Chiappetta, A., Muto, A., Bruno, L., Woloszynska, M., Van Lijsebettens, M., and Bitonti, M. B. (2015). A dehydrin gene isolated from feral olive enhances drought tolerance in *Arabidopsis* transgenic plants. *Front. Plant Sci.* 6:392. doi: 10.3389/fpls.2015.00392
- Coe, E. H., McCormick, S., and Modena, S. A. (1981). White pollen in maize. *J. Hered.* 72, 318–320.
- Conde, C., Delrot, S., and Gerós, H. (2008). Physiological, biochemical and molecular changes occurring during olive development and ripening. *J. Plant Physiol.* 165, 1545–1562. doi: 10.1016/j.jplph.2008.04.018
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610
- Dillies, M. A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., et al. (2012). A comprehensive evaluation of normalization methods for

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.01246>

Figure S1 | “Cassanese” (A–A’) and “Leucocarpa” (B–B’) fruits, in the top and bottom respectively, sampled during 100 (A,B) and 130 (A’,A”,B’,B”) DAF transition.

Figure S2 | Correlation matrix for sample across cultivar: samples are more highly correlated within cultivar than between cultivar.

Figure S3 | KEGG pathway for flavonoid and anthocyanins biosynthesis, map: 00941 (A) and 00942 (B). The isoforms of differentially expressed transcripts controlling flavonoid as well as anthocyanin biosynthesis were mapped.

Figure S4 | MapMan visualization of changes in expression levels of genes associated with secondary metabolism. Green denotes down-regulation and red up-regulation. Changes in (A,B) gene expression after 130 DAF compared to 100 DAF as calibrator, in Cassanese and Leucocarpa cvs respectively.

Table S1 | Sequences list and qRT-PCR primers set to validate selected targets.

- Illumina high-throughput RNA sequencing data analysis. *Brief. Bioinformatics* 14, 671–683. doi: 10.1093/bib/bbs046
- Dixon, R. A., Xie, D. Y., and Sharma, S. B. (2005). Proanthocyanidins - a final frontier in flavonoid research? *New Phytol.* 165, 9–28. doi: 10.1111/j.1469-8137.2004.01217.x
- Dooner, H. K. (1983). Coordinate genetic regulation of flavonoid biosynthetic enzymes in maize. *Mol. Gen. Genet.* 189, 136–141. doi: 10.1007/BF00326066
- Fu, C. H., Chen, Y. W., Hsiao, Y. Y., Pan, Z. J., Liu, Z. J., Huang, Y. M., et al. (2011). OrchidBase: a collection of sequences of the transcriptome derived from orchids. *Plant Cell Physiol.* 52, 238–243. doi: 10.1093/pcp/pcq201
- Galla, G., Barcaccia, G., Ramina, A., Collani, S., Alagna, F., Baldoni, L., et al. (2009). Computational annotation of genes differentially expressed along olive fruit development. *BMC Plant Biol.* 9:128. doi: 10.1186/1471-2229-9-128
- Garg, R., Patel, R. K., Tyagi, A. K., and Jain, M. (2011). *De novo* assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Res.* 18, 53–63. doi: 10.1093/dnares/dsq028
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883
- Green, P. S. (2002). A revision of *Olea* L. (Oleaceae). *Kew Bull.* 140, 57–91. doi: 10.2307/4110824
- Green, P. S., and Wickens, G. E. (1989). *The Olea Europaea Complex*. Edinburgh: Univ Press.
- Griesbach, R. J., and Beck, R. M. (2005). Sequence analysis of the chalcone synthase gene in four *Petunia* taxa. *J. Am. Soc. Hort. Sci.* 130, 360–365.
- Grotewold, E., and Peterson, T. (1994). Isolation and characterization of a maize gene encoding chalcone flavonone isomerase. *Mol. Gen. Genet.* 242, 1–8.
- Gutierrez-Gonzalez, J. J., Tu, Z. J., and Garvin, D. F. (2013). Analysis and annotation of the hexaploid oat seed transcriptome. *BMC Genomics* 14:471. doi: 10.1186/1471-2164-14-471
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084
- Hashim, Y. Z., Rowland, I. R., McGlynn, H., Servili, M., Selvaggini, R., Taticchi, A., et al. (2008). Inhibitory effects of olive oil phenolics on invasion in human colon adenocarcinoma cells *in vitro*. *Int. J. Cancer* 122, 495–500. doi: 10.1002/ijc.23148
- He, F., Pan, Q-H, Shi, Y., and Duan, C. Q. (2008). Biosynthesis and genetic regulation of proanthocyanidins in plants. *Molecules* 13, 2674–2703. doi: 10.3390/molecules13102674

- Holton, T. A., and Cornish, E. C. (1995). Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell* 7, 1071–1083. doi: 10.1105/tpc.7.7.1071
- Jaakola, L. (2013). New insights into the regulation of anthocyanin biosynthesis in fruits. *Trends Plant Sci.* 18, 477–483. doi: 10.1016/j.tplants.2013.06.003
- Jaakola, L., Määttä K., Pirttilä AM, Törrönen, R., Kärenlampi, S., and Hohtola, A. (2002). Expression of genes involved in anthocyanin biosynthesis in relation to anthocyanin, proanthocyanidin, and flavonol levels during bilberry fruit development. *Plant Physiol.* 130, 729–739. doi: 10.1104/pp.006957
- Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36, D480–D484. doi: 10.1093/nar/gkm882
- Karp, N. A., and Lilley, K. S. (2007). Design and analysis issues in quantitative proteomics studies. *Proteomics* 7, 42–50. doi: 10.1002/pmic.200700683
- Karp, N. A., and Lilley, K. S. (2009). Investigating sample pooling strategies for DIGE experiments to address biological variability. *Proteomics* 9, 388–397. doi: 10.1002/pmic.200800485
- Koes, R. E., Spelt, C. E., and Mol, J. N. (1989). The chalcone synthase multigene family of *Petunia hybrida* (V30): differential, light regulated expression during flower development and UV light induction. *Plant Mol. Biol.* 12, 213–225. doi: 10.1007/BF00020506
- Koes, R., Verweij, W., and Quattrocchio, F. (2005). Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends Plant Sci.* 10, 236–242. doi: 10.1016/j.tplants.2005.03.002
- Langmead, B., and Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lekanne Deprez, R. H., Fijnvandraat, A. C., Ruijter, J. M., and Moorman, A. F. (2002). Sensitivity and accuracy of quantitative real-time polymerase chain reaction using SYBR green I depends on cDNA synthesis conditions. *Anal. Biochem.* 307, 63–69. doi: 10.1016/S0003-2697(02)00021-0
- Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* 12:323. doi: 10.1186/1471-2105-12-323
- Liang, C., Liu, X., Yiu, S. M., and Lim, B. L. (2013). *De novo* assembly and characterization of *Camelina sativa* transcriptome by paired-end sequencing. *BMC Genomics* 14:146. doi: 10.1186/1471-2164-14-146
- Liu, S., Li, W., Wu, Y., Chen, C., and Lei, J. (2013). *De novo* transcriptome assembly in chili pepper (*Capsicum frutescens*) to identify genes involved in the biosynthesis of capsaicinoids. *PLoS ONE* 8:e48156. doi: 10.1371/journal.pone.0048156
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Llorente-Cortes, V., Estruch, R., Mena, M. P., Ros, E., González, M. A., Fitó M, et al. (2010). Effect of Mediterranean diet on the expression of pro-atherogenic genes in a population at high cardiovascular risk. *Atherosclerosis* 208, 442–450. doi: 10.1016/j.atherosclerosis.2009.08.004
- Lohse, M., Nagel, A., Herter, T., May, P., Schroda, M., Zrenner, R., et al. (2014). Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. *Plant Cell Environ.* 37, 1250–1258. doi: 10.1111/pce.12231
- Loumou, A., and Giourga, C. (2003). Olive groves: the life and identity of the Mediterranean. *Agric. Hum. Values* 20, 87–95. doi: 10.1023/A:1022444005336
- Martinelli, F., and Tonutti, P. (2012). Flavonoid metabolism and gene expression in developing olive (*Olea europaea* L.) fruit. *Plant Biosyst.* 146, 164–170. doi: 10.1080/11263504.2012.681320
- Matus, J. T., Loyola, R., Vega, A., Peña-Neira, A., Bordeu, E., Arce-Johnson, P., et al. (2009). Post-veraison sunlight exposure induces MYB-mediated transcriptional regulation of anthocyanin and flavonol synthesis in berry skins of *Vitis vinifera*. *J. Exp. Bot.* 60, 853–867. doi: 10.1093/jxb/ern336
- Mizrachi, E., Hefer, C. A., Ranik, M., Joubert, F., and Myburg, A. A. (2010). *De novo* assembled expressed gene catalog of a fast-growing *Eucalyptus* tree produced by Illumina mRNA-Seq. *BMC Genomics* 11:681. doi: 10.1186/1471-2164-11-681
- Morgulis, A. (2008). Database indexing for production MegaBLAST searches. *Bioinformatics* 24, 1757–1764. doi: 10.1093/bioinformatics/btn322
- Muñoz-Mérida, A., González-Plaza, J. J., Cañada, A., Blanco, A. M., García-López Mdel, C., Rodríguez, J. M., et al. (2013). *De novo* assembly and functional annotation of the olive (*Olea europaea*) transcriptome. *DNA Res.* 20, 93–108. doi: 10.1093/dnares/dss036
- Muzzalupo, I. (2012). *Olive Germplasm – Italian Catalogue of Olive Varieties*. Rijeka: InTech
- Muzzalupo, I., Macchione, B., Bucci, C., Stefanizzi, F., Perri, E., Chiappetta, A., et al. (2012). LOX gene transcript accumulation in olive (*Olea europaea* L.) fruits at different stages of maturation: relationship between volatile compounds, environmental factors, and technological treatments for oil extraction. *Sci. World J.* 2012, 1–9. doi: 10.1100/2012/532179
- Muzzalupo, I., Stefanizzi, F., Perri, E., and Chiappetta, A. (2011). Transcript levels of CHL P gene, antioxidants and chlorophylls contents in olive (*Olea europaea* L.) pericarps: a comparative study on eleven olive cultivars harvested in two ripening stages. *Plant Foods Hum. Nutr.* 66, 1–10. doi: 10.1007/s11130-011-0208-6
- Muzzalupo, I., Vendramin, G. G., and Chiappetta, A. (2014). Genetic biodiversity of Italian olives (*Olea europaea*) germplasm analyzed by SSR markers. *Sci. World J.* 2014, 1–12. doi: 10.1155/2014/296590
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., et al. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320, 1344–1349. doi: 10.1126/science.1158441
- Pallavicini, A., Canapa, A., Barucca, M., Alf Ldi, J., Biscotti, M. A., Buonocore, F., et al. (2013). Analysis of the transcriptome of the Indonesian coelacanth *Latimeria menadoensis*. *BMC Genomics* 14:538. doi: 10.1186/1471-2164-14-538
- Peng, X., Wood, C. L., Blalock, E. M., Chen, K. C., Landfield, P. W., and Stromberg, A. (2003). Statistical implications of pooling RNA samples for microarray experiments. *BMC Bioinformatics* 4:26. doi: 10.1186/1471-2105-4-26
- Pérez-Jiménez, F., Ruano, J., Perez-Martinez, P., Lopez-Segura, F., and Lopez-Miranda, J. (2007). The influence of olive oil on human health: not a question of fat alone. *Mol. Nutr. Food Res.* 51, 1199–1208. doi: 10.1002/mnfr.200600273
- Pua, E. C., and Davey, M. R. (2010). *Plant Developmental Biology - Biotechnological Perspectives*, Vol. 1. Heidelberg; Dordrecht; London; New York, NY: Springer.
- Ramsay, N. A., and Glover, B. J. (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci.* 10, 63–70. doi: 10.1016/j.tplants.2004.12.011
- Ravaglia, D., Espley, R. V., Henry-Kirk, R. A., Andreotti, C., Ziosi, V., Hellens, R. P., et al. (2013). Transcriptional regulation of flavonoid biosynthesis in nectarine (*Prunus persica*) by a set of R2R3 MYB transcription factors. *BMC Plant Biol.* 13:68. doi: 10.1186/1471-2229-13-68
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Robinson, M. D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11:R25. doi: 10.1186/gb-2010-11-3-r25
- Schaart, J. G., Dubos, C., Romero De La Fuente, I., van Houwelingen, A. M. M. L., de Vos, R. C. H., Jonker, H. H., et al. (2012). Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (*Fragaria x ananassa*) fruits. *New Phytol.* 197, 454–467. doi: 10.1111/nph.12017
- Schaefer, H. M., Schaefer, V., and Levey, D. J. (2004). How plant-animal interactions signal new insights in communication. *Trends Ecol. Evol.* 19, 577–584. doi: 10.1016/j.tree.2004.08.003
- Stommel, J. R., Lightbourn, G. J., Winkel, B. S., and Griesbach, R. J. (2009). Transcription factor families regulate the anthocyanin biosynthetic pathway in *Capsicum annuum*. *J. Am. Soc. Hort. Sci.* 134, 244–251.
- Sweetman, C., Deluc, L. G., Cramer, G. R., Ford, C. M., and Soole, K. L. (2009). Regulation of malate metabolism in grape berry and other developing fruits. *Phytochemistry* 70, 1329–1344. doi: 10.1016/j.phytochem.2009.08.006
- Takahama, U. (2004). Oxidation of vacuolar and apoplastic phenolic substrates by peroxidases: physiological significance of the oxidation reactions. *Phytochem. Rev.* 3, 207–219. doi: 10.1023/B:PHYT.0000047805.08470.e3
- Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., et al. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* 37, 914–939. doi: 10.1111/j.1365-313X.2004.02016.x
- Tian, L., Pang, Y., and Dixon, R. A. (2008). Biosynthesis and genetic engineering of proanthocyanidins and (iso)flavonoids. *Phytochem. Rev.* 7, 445–465. doi: 10.1007/s11101-007-9076-y
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-seq reveals

- unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621
- Tulin, S., Aguiar, D., Istrail, S., and Smith, J. (2013). A quantitative reference transcriptome for *Nematostella vectensis* early embryonic development: a pipeline for de novo assembly in emerging model systems. *Evodevo* 4:16. doi: 10.1186/2041-9139-4-16
- Umemura, H., Otagaki, S., Wada, M., Kondo, S., and Matsumoto, S. (2013). Expression and functional analysis of a novel MYB gene, MdMYB110a_JP, responsible for red flesh, not skin color in apple fruit. *Planta* 238, 65–76. doi: 10.1007/s00425-013-1875-3
- van Tunen, A. J., Hartman, S. A., Mur, L. A., and Mol, J. N. (1989). Regulation of chalcone flavone isomerase (CHI) gene expression in *Petunia hybrida*: the use of alternative promoters in corolla, anthers and pollen. *Plant Mol. Biol.* 12, 539–551. doi: 10.1007/BF00036968
- van Tunen, A. J., Koes, R. E., Spelt, C. E., van der Krol, A. R., Stuitje, A. R., Mol, J. N., et al. (1988). Cloning of the two chalcone flavanone isomerase genes from *Petunia hybrida*: coordinate light-regulated and differential expression of flavonoid genes. *EMBO J.* 7, 1257–1263.
- Vauzour, D., Rodriguez-Mateos, A., Corona, G., Oruna-Concha, M. J., and Spencer, J. P. E. (2010). Polyphenols and human health: prevention of disease and mechanisms of action. *Nutrients* 2, 1106–1131. doi: 10.3390/nu2111106
- Wallander, E., and Albert, V. A. (2000). Phylogeny and classification of Oleaceae based on rps16 and trnL-F sequence data. *Am. J. Bot.* 12, 1827–1841. doi: 10.2307/2656836
- Wang, Z., Fang, B., Chen, J., Zhang, X., Luo, Z., Huang, L., et al. (2010). De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (*Ipomoea batatas*). *BMC Genomics* 11:726. doi: 10.1186/1471-2164-11-726
- Ward, J. A., Ponnala, L., and Weber, C. A. (2012). Strategies for transcriptome analysis in non model plants. *Am. J. Bot.* 99, 267–276. doi: 10.3732/ajb.1100334
- Wei, W., Qi, X., Wang, L., Zhang, Y., Hua, W., Li, D., et al. (2011). Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. *BMC Genomics* 12:451. doi: 10.1186/1471-2164-12-451
- Zeng, S., Wu, M., Zou, C., Liu, X., Shen, X., Hayward, A., et al. (2014). Comparative analysis of anthocyanin biosynthesis during fruit development in two *Lycium* species. *Physiol. Plant.* 150, 505–516. doi: 10.1111/pp.12131
- Zhang, J., Wang, X., Yu, O., Tang, J., Gu, X., Wan, X., et al. (2011b). Metabolic profiling of strawberry (*Fragaria x ananassa* Duch) during fruit development and maturation. *J. Exp. Bot.* 62, 1103–1118. doi: 10.1093/jxb/erq343
- Zhang, Y. J., Ma, P. F., and Li, D. Z. (2011a). High-throughput sequencing of six bamboo chloroplast genomes: phylogenetic implications for temperate woody bamboos (*Poaceae: Bambusoideae*). *PLoS ONE* 6:e20596. doi: 10.1371/journal.pone.0020596

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Iaria, Chiappetta and Muzzalupo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.