Available online: www.notulaebotanicae.ro
Print ISSN 0255-965X; Electronic 1842-4309
Not Bot Horti Agrobo, 2019, 47(2):308-319. DOI:10.15835/nbha47211346
Original Article

Avocado Fruit Pulp Transcriptomes in the after-Ripening Process

Liqin LIU[1,a], Bo SHU[1,2,b], Dengwei JUE[1], Yicheng WANG[1], Yongzan WEI[1], Shengyou SHI[1]*

[1]*Chinese Academy of Tropical Agricultural Science, South Subtropical Crops Research Institute, Key Laboratory of Tropical Fruit Biology, Ministry of Agriculture, 524091 Zhanjiang, P. R. China; lolitallq@163.com; bshbest@163.com; juedengwei@126.com; ychw08@163.com; wyz4626@163.com; xiejianghui@21cn.com; ssy7299@163.com (*corresponding author)*
[2]*Yunnan Agricultural University, College of Horticulture and Landscape, 650201 Kunming, P. R. China*
[a,b]*These authors contributed equally to this work*

Abstract

Avocado is an important tropical fruit whose after-ripening process is still poorly understood. The fatty acid, phenolics, flavonoids, and tannins were analyzed in 'Lisa' avocado (*Persea americana* Mill. 'Lisa') fruit pulp during after-ripening. The transcriptome was analyzed to screen for transcripts associated with the aforementioned after-ripening parameters. The results showed that there were no significant differences in the total fatty acid content among the preclimacteric, climacteric, and postclimacteric stages. Nevertheless, the concentrations of C18:3 (α-linolenic acid) were significantly higher in the climacteric and postclimacteric stages than the preclimacteric stage. RNAseq generated 235,082 transcripts and 151,545 unigenes. In addition, 4,324 DEGs were produced among the three stages. KEGG analysis of the DEGs suggested the pathways about "α-linolenic acid metabolism, unsaturated fatty acid biosynthesis", "fatty acid degradation", "linoleic acid metabolism and fatty acid biosynthesis", "linoleic acid metabolism and fatty acid elongation", and "fatty acid elongation" may all contribute to the C18:3 variations in 'Lisa' avocado fruit pulp. Several transcription factors, including the ethylene-related transcription factors, such as NAC, MYB, bHLH, and WRKY, were also identified in the DEGs database. This study generated transcript data and screened the transcription factors involved in the avocado after-ripening process. This information could be used to control after-ripening in avocado and maintain fruit quality during storage.

Keywords: fatty acids; flavonoids; phenolics; RNAseq; tannin; transcription factor
Abbreviations: COG: cluster of orthologous groups of proteins; DEG: differentially expressed gene; FDR: false discovery rate; FPKM: fragments per kilobase of exon per million fragments mapped; GO: gene ontology; HPLC: high-performance liquid chromatography; KEGG: Kyoto Encyclopedia of Genes and Genomes; NR: National Center for Biotechnology Information (NCBI) non-redundant (NR) database.

Introduction

Avocado (*Persea americana* Mill.) is a fruit tree indigenous to tropical- and subtropical regions. Avocado fruit has long been considered a healthful food. Research has shown that it can be useful in the management of hypercholesterolemia and hypertension (Dreher and Davenport, 2013). Consumer demand for avocado has significantly increased in recent years. Avocado trees have been cultivated in the southern Chinese provinces of Guangdong, Guangxi, Hainan, and Yunnan. The area under avocado cultivation has rapidly expanded (Zhang *et al.*, 2015). Postharvest storage is an important aspect of avocado production and has presented some challenges in the production area of China (Zhang *et al.*, 2015). Premature fruit overripening has decreased crop value.

A unique feature of all avocado varieties is that the fruits mature on the tree and ripen after harvest. After-ripening takes 5-7 d at room temperature and the overall fruit quality changes significantly. Several studies indicated that the total fatty acid content remained constant during the after-ripening process (Ozdemir and Topuz, 2004; Villa-Rodriguez *et al.*, 2011). Nevertheless, the amounts of individual fatty acids (oleic, palmitic, palmitoleic, linoleic, and linolenic) in the pulp might differ (Villa-Rodriguez *et al.*, 2011; Dreher and Davenport, 2013). The composition and quantity of total phenolics, flavonoids, tannins and other fruit quality indices may also vary in the after-ripening process (Villa-Rodriguez *et al.*, 2011). It is still unknown

Received: 22 Aug 2018. Received in revised form: 16 Oct 2018. Accepted: 23 Oct 2018. Published online: 29 Oct 2018.

how avocado fruit quality is regulated at the transcription level. Transcription factors participate in the biosynthetic pathways of phytohormones like ethylene and abscisic acid which control after-ripening. Specifically, MADS-box, NAC, SBP/SPL, and AP2/ERF are all involved in the regulation of fruit after-ripening (Shan *et al*., 2012; Gapper *et al*., 2013; Elitzur *et al*., 2016).

The 'Lisa' avocado (*Persea americana* Mill. 'Lisa') cultivar matures in mid-September in Guangdong Zhanjiang. Its production and quality are stable and consistent. This cultivar is well-suited for cultivation in China. In the present study, the quantity of total fatty acids, phenolics, flavonoids and tannins were determined in the after-ripening stage of 'Lisa' avocado fruit. The composition of the fatty acids was analyzed, and the quantity of each component was also measured. Furthermore, the transcriptome was analyzed to screen transcripts involved in fatty acid biosynthesis. The transcription factors were also identified for clarifying the after-ripening process of 'Lisa' avocado. This study generated transcript data and screened for transcription factors involved in after-ripening. This information may help regulate after-ripening in avocado fruit and maintain fruit quality in storage.

Materials and Methods

Materials and experimental design

Nine fruits of similar size, shape, and weight were selected from each of three avocado trees. The trees were 'Lisa' scions grafted onto 'Kampong' rootstocks and had moderately vigorous growth. They were cultivated in the experimental orchard of the South Subtropical Crops Research Institute, Zhanjiang, China. The trees received standard horticultural treatments including weeding, irrigation, and pest control. The fruits were picked when mature fruit fell from the trees. The fruits were washed with sterile water, dried with sterile gauze, and stored at 20 °C. The nine avocado fruits from each tree were divided into three groups, each consisting of three fruits and representing one biological replicate. The samples were obtained at day 0 (0D), day 3 (3D), and day 6 (6D) after harvest (preclimacteric, climacteric, and postclimacteric stages, respectively). Replicates from the same age group were sampled from the same point on all three trees. The degree of fruit softness was used to determine whether the fruits were climacteric.

Quality index analysis in avocado pulp

The total phenolic content was analyzed according to Katalinic (2006). Fruit samples were air-dried at 55 °C. One-gram samples were homogenized in a mortar containing 6 mL extraction solution (70% v/v aqueous ethanol). Homogenates were transferred to 10-mL centrifuge tubes and sonicated in an ultrasonic cleaner at 37 °C for 30 min. The centrifuge tubes were then spun at 8,000 × rpm for 10 min at 4 °C. The supernatants were used to determine total phenolics, tannins, and flavonoids. The total phenolics in the extracts were determined according to the Folin–Ciocalteau procedure (Singleton and Rossi, 1965). Briefly, 0.25 mL each of the extract, 200 mg/L gallic acid, and water, 0.5 mL 1:1 diluted Folin–Ciocalteu's phenol reagent, and 4 mL 10% (w/v) Na_2CO_3 were mixed and incubated in darkness for 2 h at 25 °C. The absorbance was then measured at 765 nm. The phenolic content was estimated from a gallic acid standard curve and expressed as mg g^{-1} dry weight. The amount of tannin in the extracts was determined according to Temill *et al.* (1990). F-D solution was prepared by dissolving 50 g sodium tungstate and 10 g phosphomolybdic acid in 375 mL water in a 500-mL conical flask. Twenty-five milliliters 85% H_3PO_4 was then added, and the flask was kept in a boiling water bath for 2 h. The solution was cooled then made up to a constant volume of 500 mL. A 0.4 mL extract aliquot, 1.2 mL F-D solution, 2 mL saturated Na_2CO_3, and 6.4 mL water were mixed and incubated in darkness at 25 °C for 30 min. The absorbance was then measured at 755 nm. The tannin content was estimated from a standard curve and the results were expressed as mg g^{-1} dry weight. The amount of total flavonoids in the extracts was determined using the method of Djeridane *et al*. (2006). In brief, 1 mL extract, 2.7 mL 30% v/v aqueous ethanol, 0.15 mL 0.5 M $NaNO_2$, and 0.15 mL 0.3 M $AlCl_3$ were mixed in a 10-mL centrifuge tube and left to stand for 5 min. Then 1 mL 1 M NaOH was added to the centrifuge tube and the absorbance was measured at 506 nm. The total flavonoids were estimated from a rutin standard curve and expressed as mg g^{-1} dry weight.

Total fatty acids were extracted by the hexane-isopropanol method (Hara *et al*., 1978). The extracted lipids were weighed, resuspended in hexane, converted to fatty acid methyl esters by a base-catalyzed methylation reaction, and analyzed for fatty acid composition using a gas chromatograph coupled with a flame ionization detector (GC-FID; Varian 3800). Fatty acids were quantified against triheptadecanoin added as an internal standard prior to lipid extraction (Kilaru *et al*., 2015).

Total RNA extraction and RNA sequencing

Total RNA was isolated from nine samples of avocado fruit with TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and treated with DNase I to eliminate genomic DNA. RNA quality was determined by 1% agarose gel electrophoresis. A NanoPhotometer spectrophotometer (Implen, Los Angeles, CA, USA) was used to detect degradation and contamination. A Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) measured the RNA concentration. An Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) verified sample integrity.

A total of 1.5 μg RNA per sample was used as input material. Sequencing libraries were generated using a NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, Ipswich, MA, USA) following the manufacturer's recommendations. Index codes were added to the attribute sequences for each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations at elevated temperature in NEBNext first-strand synthesis reaction buffer (5X). First-strand cDNA was synthesized using random hexamer primer and M-MuLV reverse transcriptase (RNase H). Second-strand cDNA was then synthesized with DNA polymerase I and RNase H. The remaining overhangs were converted into blunt ends by

exonuclease/polymerase reactions. After adenylation of the 3' ends of DNA fragments, NEBNext adaptors with hairpin loops were ligated to prepare for hybridization. To select cDNA fragments 150-200 bp long, the library fragments were purified with AMPure XP system (Beckman Coulter, Genomics, Danvers, MA, USA). Then, 3 μL USER enzyme (NEB, Ipswich, MA, USA) were added to the 150-200 bp adaptor-ligated cDNA and maintained at 37°C for 15 min, followed by 5 min at 95°C. PCR was then performed with Phusion® High-Fidelity DNA Polymerase (NEB, Ipswich, MA, USA), universal PCR primers, and index (X) primer. The PCR products were purified using the AMPure XP system and the library quality was assessed with the Agilent Bioanalyzer 2100 system. For quality control, the Agilent 2100 Bioanalyzer and the ABI StepOnePlus Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA, USA) were used to determine sample library quantity and quality. Clustering of the index-coded samples was performed with a cBot Cluster Generation System and a TruSeq PE Cluster Kit v3-cBot-HS (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. After cluster generation, the nine libraries were sequenced on a Hiseq 4000 (Illumina, San Diego, CA, USA) and paired-end reads were generated (Shu *et al*., 2016).

Transcript assembly and functional annotation

Before bioinformatics analysis, the raw sequences were filtered to remove reads containing only adaptor sequences or > 5% unknown nucleotides. Low-quality reads having > 20% bases with a quality value ≤10 were also eliminated. Q20, Q30, GC content, and the sequence duplication level of the clean data were calculated simultaneously. All downstream analyses were based on high-quality clean data. Since the avocado genome had not yet been published, *de novo* assembly was performed with Novogene (Beijing, China) and the short-read assembly program for avocado in the Trinity software package (v. r2014043p1). The min_kmer_cov: 2 and all other parameters were set at their defaults (Grabherr *et al*., 2011).

Functional unigene annotation was first performed using the NR, Swiss-Prot protein (http: //www .expasy.ch/sprot) and KOG/COG (https: //www .biostars.org/p/170077) databases by using DIAMOND v. 0.8.22. The e-values were 1e-5 for NR and Swiss-Prot, and 1e-3 for KOG/COG. BLAST v. 2.2.28+ (e-value 1e-5), KASS v. r140224 (e-value 1e-10), and blast2go v. b2g4 | 2.5 (e-value 1e-6) were used to annotate transcripts against Nt (NCBI [National Center for Biotechnology Information] non-redundant nucleotide sequences), KEGG (http:// www. genome. jp/kegg/), and GO (http:// www .geneontology.org/), respectively. When a unigene did not align with any of the above databases, Hmmscan (HMMER) v. 3 annotated by functional domain prediction (Grabherr *et al*., 2011).

Differential expression analysis

Reads containing adaptors or > 10% unknown nucleotides, and low-quality reads with > 50% bases with a quality value ≤ 5 were removed to obtain uncontaminated sequences. BAM files were created and run in RSEM v. 1.2.17 to calculate the number of reads mapped on the transcript. The number of mapped and filtered reads for each unigene was calculated to obtain the corresponding FPKM (fragments/kb exon/10^6 fragments mapped) values (Li and Dewey, 2011). DESeq provided statistical routines to determine significant differential expressions in the digital gene expression data using a negative binomial distribution model. The resulting P-values were adjusted using the Benjamini and Hochberg method for controlling the false discovery rate. Genes were considered differentially expressed when they had adjusted P-values < 0.05 according to DESeq. Differential expression analyses of the three replicates were performed using the DEGseq R package. P-values were adjusted using the q-value (Storey 2003). [Q-value < 0.005 | log 2 (foldchange) | > 1] was set as the threshold for significant differential expression. The heat maps of the selected differentially expressed genes (DEGs) from the different treatments were constructed using MeV v. 4.9.0.

GO enrichment analysis

GO enrichment analysis of the DEGs were run using the GOseq R packages based on the Wallenius non-central hyper-geometric distribution which compensates for gene length bias (Young *et al*., 2010).

KEGG pathway enrichment analysis

KEGG is a database used to explain high-level functions in cells, organisms, and ecosystems (Kanehisa *et al*., 2008). It uses molecular-level information from large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies. We used KOBAS software to test the statistical enrichment of DEGs in KEGG pathways Supplementary Fig. 1 (Mao *et al*., 2005). The flow chart for RNA-seq analysis was shown in Supplementary Fig. 1.

Statistical analysis

Experimental data were subjected to ANOVA with SAS v. 8.1 (SAS Institute, Cary, NC, USA). The probabilities of significance were determined for the treatments and the least significant difference (LSD; $p < 0.05$) was used to compare the data.

Results

Quality index analysis in 'Lisa' avocado pulp

The ranges of tannin, total flavonoids, and total phenolics in 'Lisa' avocado fruit pulp were 1.16-1.26 mg/g, 0.31-0.44 mg/g, and 0.47-0.53 mg/g respectively. There were no significant differences in the 0D, 3D, or 6D samples in terms of these indices. Total fatty acid varied in the same way as tannin, total flavonoids, and total phenolics. The total fatty acid content did not significantly change throughout the whole after-ripening process (Fig. 1). The following twelve fatty acids were identified in the 'Lisa' avocado fruit pulp: C14:0 (myristic acid), C16:0 (palmitic acid), C16:1 (9-*cis*-hexadecenoic acid), C17:0 (daturic acid), C17:1 (*cis*-10-heptadecenoic acid), C18:0 (octadecanoic acid), C18:1 (oleic acid), C18:2 (linoleic acid), C18:3 (α-linolenic acid), C20:0 (arachidic acid), C20:1 (*cis*-11-

eicosenoic acid), and C21:0 (heneicosanoic acid). The most abundant fatty acids in the fruit pulp were C16:0 (palmitic acid), C16:1 (9-*cis*-hexadecenoic acid), C18:1 (oleic acid), and C18:2 (linoleic acid). Their concentrations ranged from 0.1-0.5. Nevertheless, none of their concentrations significantly changed over the course of after-ripening. Although the concentration of C18:3 (α-linolenic acid) ranged from 0.004-0.008, it was significantly higher at 3D and 6D than it was at 0D (Fig. 2).

Sequence assembly and annotation

Nine samples were sequenced using a Hiseq genome analyzer (Illumina, San Diego, CA, USA). After quality checking and data cleaning, approximately forty million reads of each sample were obtained. Their average length and GC content were 100 bp and 46%, respectively (Supplementary Table 1). Assembly of all the reads from the nine samples generated 235,082 transcripts and 151,545 unigenes with average lengths of 904 bp and 1,243 bp, respectively (Table 1). These unigenes were annotated by the NR, NT, KO, Swissprot, PFAM, GO, and KOG databases. Of the sequences, 51.5% (78,054) were annotated with reference to NR and the balance (73,491) with reference to the others (Table 1).These transcriptome data were submitted to NCBI sequence read archive No. SRX2944612.

The avocado unigenes most closely matched sequences from lotus (*Nelumbo nucifera*) (35.6%) and grape (*Vitis vinifera*) (11.2%) (Fig. 3a). Of the annotated sequences, 32.8% showed "very strong homology" (*E*-value < 10^{-100}) and 26.2% displayed "strong homology" (10^{-100} < *E*-value < 10^{-45}) to the available sequences (Fig. 3b). Of the unigenes, 37.9% showed "very high similarity" (≥ 80%) and 45.1% presented "high similarity" (80% ≥ similarity ≥ 60%) to the matched sequences (Fig. 3c).

The seven databases were used to classify unigene functions. The avocado unigenes were classified by GO analysis according to "biological process", "cellular component" and "molecular function" (Fig. 4a).

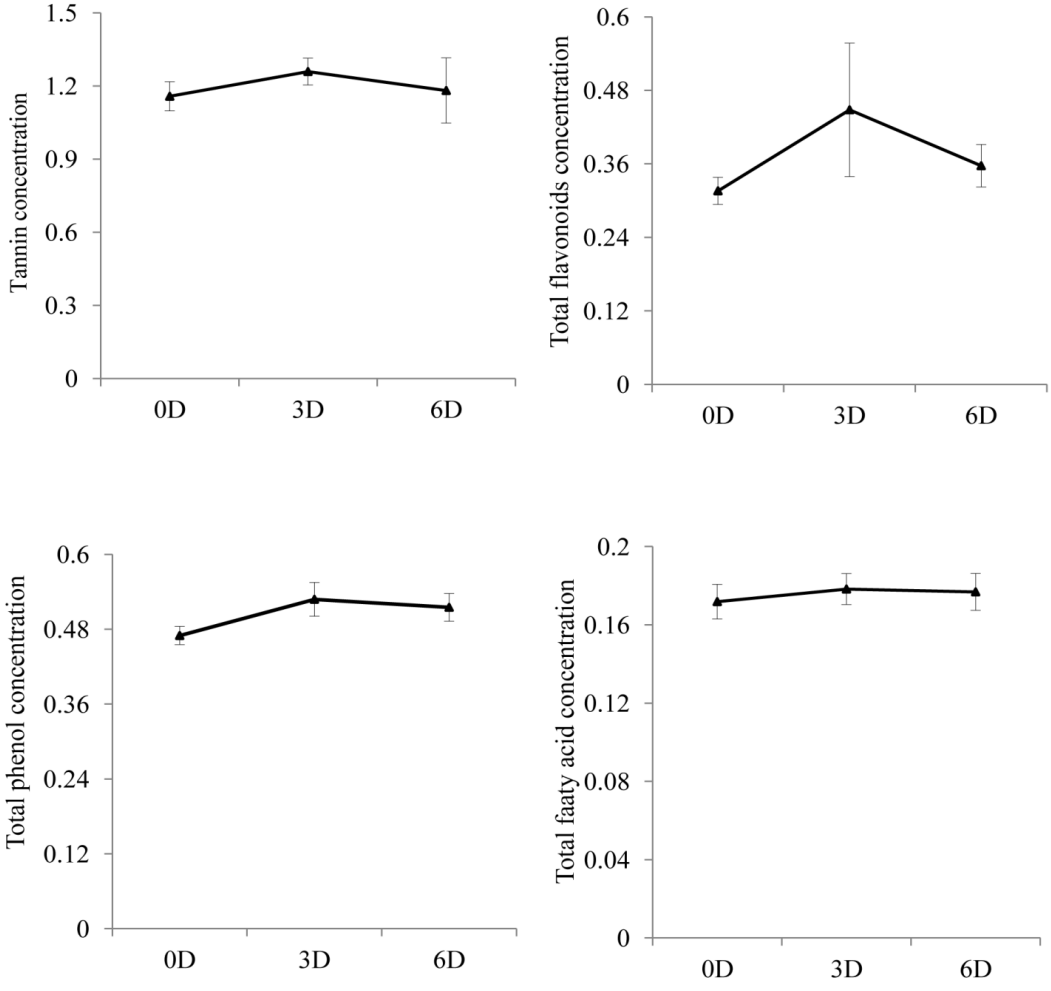


Fig. 1. Variations in tannin (mg g^{-1} dry weight), total flavonoids (mg g^{-1} dry weight), total phenolics (mg g^{-1} dry weight), and total fatty acids (%) in 'Lisa' avocado fruit pulp during after-ripening.*: data (mean ± SE, n = 3) are significantly different (P < 0.05)

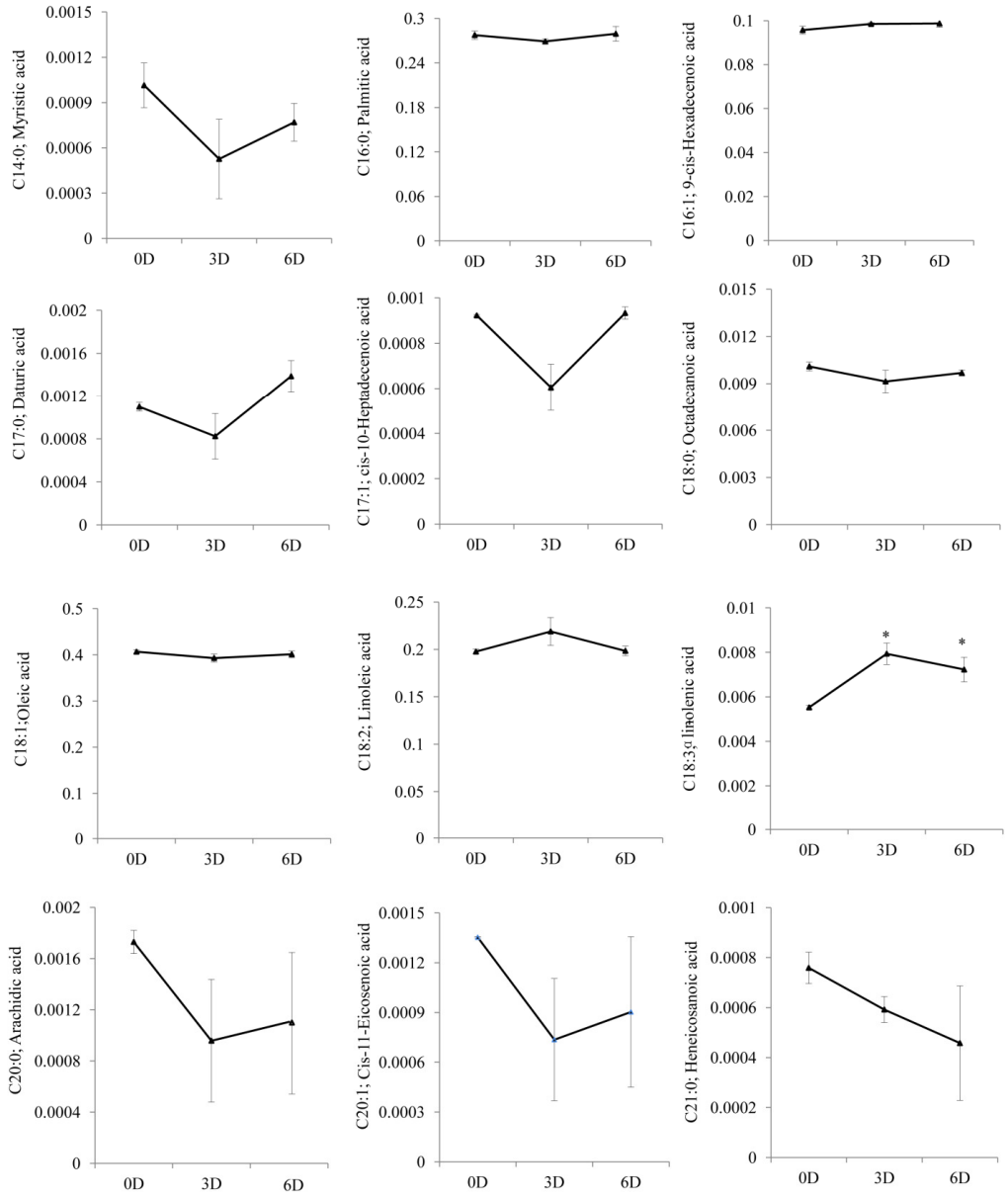


Fig. 2. Variations of fatty acid content (%) in 'Lisa' avocado fruit pulp during after-ripening. *: data (mean ± SE, n = 3) are significantly different ($P < 0.05$)

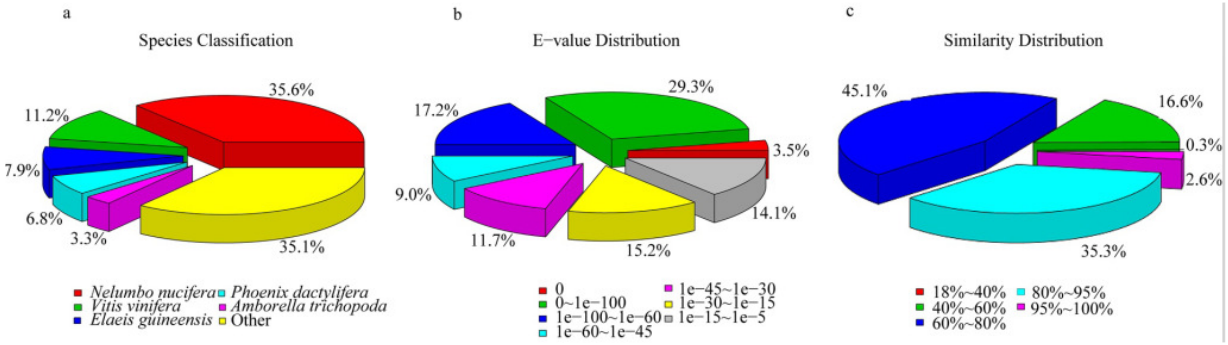


Fig. 3. Outcome of homology search of unigenes against the NR, NT, KO, Swissprot, PFAM, GO, and KOG databases. (a) Species distribution of the top BLAST hits for all homologous sequences. (b) E-value distribution of the top BLAST hits for each unique sequence. (c) Similarity distribution of the BLAST hits for each unique sequence

Table 1. Summary of read numbers based on the RNA-Seq data from the pulp of 'Lisa' avocado during after-ripening process

	Transcripts	Unigenes
Total number	235,082	151,545
N50	1,616	1,840
Median Length	482	863
Average length	904	1243
Total length	212,398,204	188,392,442
	Number of Genes	Percentage (%)
Annotated in NR	78,054	51.5
Annotated in NT	49,620	32.74
Annotated in KO	32,865	21.68
Annotated in SwissProt	58,733	38.75
Annotated in PFAM	58,858	38.83
Annotated in GO	59,388	39.18
Annotated in KOG	22,749	15.01
Annotated in all Databases	12,551	8.28
Annotated in at least one Database	88,275	58.25
Total Unigenes	151,545	100

The avocado pulp contained numerous unigenes annotated as "metabolic process" and "cellular process" under the "biological process" category; "binding" and "catalytic activities" under the "molecular function" category; and "cell" and "cell part" under the "cellular component" category (Fig. 4a). KOG analysis classified avocado unigenes into twenty-five categories. "Posttranslational modification, protein turnover, and chaperones", "translation, ribosomal structure, and biogenesis" and "general function prediction only" clustered most of the unigenes in KOG (Fig. 4b). Nineteen categories were classified by KEGG analysis. "Carbohydrate metabolism", "translation" and "folding, sorting and degradation" were the categories with the highest number of unigenes (Fig. 4c).

Differential gene expression analysis

The DEGs in 'Lisa' avocado fruit during after-ripening were assessed by pairwise comparisons of three time points (0D, 3D, and 6D) with the expression fold (log 2 ratio ≥ 1) and FDR ≤ 10^{-3} as the thresholds. The number of DEGs increased with girdling duration.

Thirty-five DEGs (12 upregulated and 23 downregulated) were detected between 3D and 0D. Moreover, 207 DEGs (147 upregulated and 60 downregulated) between 6D and 0D, and 4,222 DEGs (2,283 upregulated and 1,939 downregulated) between 6D and 3D were identified (Fig. 5a).

The DEGs were classified into "biological process" and "molecular function" by GO analysis. The highest annotated DEG numbers were obtained in "metabolic process" and "single-organism metabolic process" under the "biological process" category, and in the "hydrolase activity, hydrolyzing O-glycosyl compounds" and "hydrolase activity, acting on glycosyl bonds" under the "molecular function" category (Fig. 5b).

KEGG pathway enrichment analysis of differentially expressed genes

The major biological processes and their unigenes were screened during 'Lisa' avocado fruit after-ripening. Unlike the DEGs at 0D, most of those at 3D were mapped to "plant–pathogen interaction", "amino sugar and nucleotide sugar metabolism", "mismatch repair" and "base excision repair" for avocado (Figs. 6a and 6b; Supplementary Table 2). In contrast to the DEGs at 0D, those at 6D were mapped to "RNA polymerase", "RNA transport and RNA degradation" and "α-linolenic acid metabolism" (Figs. 6c and 6d; Supplementary Table 2). Compared with the DEGs at 3D, those at 6D mapped to "α-linolenic acid metabolism", "biosynthesis of unsaturated fatty acids", "fatty acid degradation", "linoleic acid metabolism fatty acid biosynthesis", "fatty acid elongation" and "linoleic acid metabolism and fatty acid elongation" (Figs. 6e and 6f; Supplementary Table 2). The number of DEGs at 6D versus that at 3D was highest in the pairwise comparison of the three time points.

Clustering results of time-course data from RNA-Seq by STEM analysis

DEGs with similar expression patterns were clustered into eight distinct subclusters. DEGs grouped in the same subcluster may be functionally correlated. Subclusters I, II, III, IV, and V showed similar expression models: they upregulated at 3D and downregulated at 6D. Nevertheless, the degrees of up- and downregulation differed among them. The DEGs in subcluster I showed -1 to 1 of log 2 at 3D and -2 to 0 of log 2 at 6D. Subcluster II showed 0 to 2 of log 2 at 3D and -3 to -1 of log 2 at 6D. Subcluster III showed 0 to 6 of log 2 at 3D and 0 to -2 of log 2 at 6D. Subcluster IV showed 0 to 4 of log 2 at 3D and -2 to -4 of log 2 at 6D. Subcluster V showed 0 to 5 of log 2 at 3D and -5 to -10 of log 2 at 6D. Unlike subclusters I, II, III, IV, and V, subclusters VI, VII, and VIII were downregulated at 3D and upregulated at 6D. Most DEGs in subcluster VI showed -4 to 1 of log 2 at 3D and 0 to 2 of log 2 at 6D.

The DEGs in subcluster VII showed -4 to 1 of log 2 at 3D and 2 to 8 of log 2 at 6D. The DEGs in subcluster VIII showed -6 to 0 of log 2 at 3D and 0 to 4 of log 2 at 6D (Fig. 7; Supplementary Table 3).

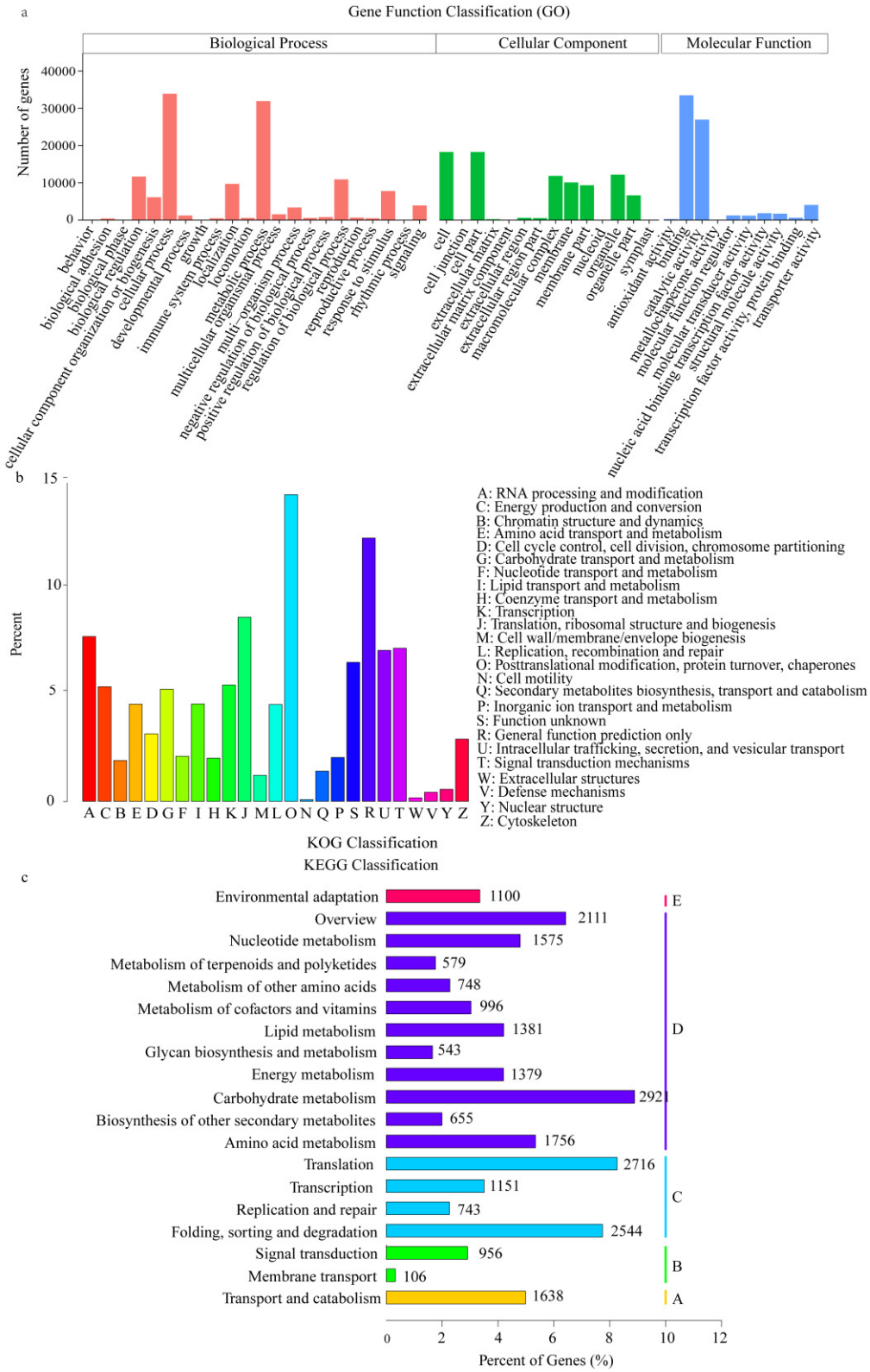


Fig. 4. Histogram of GO (a), KOG (b), and KEGG (c) classifications for transcripts from 'Lisa' avocado fruit pulp. (a) The unigenes corresponded to three main categories: "biological process," "cellular component," and "molecular function." The y-axes indicate the percentage and number of annotated unigenes. (b) The x-axis indicates the KOG functional category, and the y-axes indicate the percentage of annotated unigenes. (c) The x-axis indicates the percentage and number of annotated unigenes, and the y-axes indicate the KEGG functional categories. The number of unigenes in each category is shown on the bar

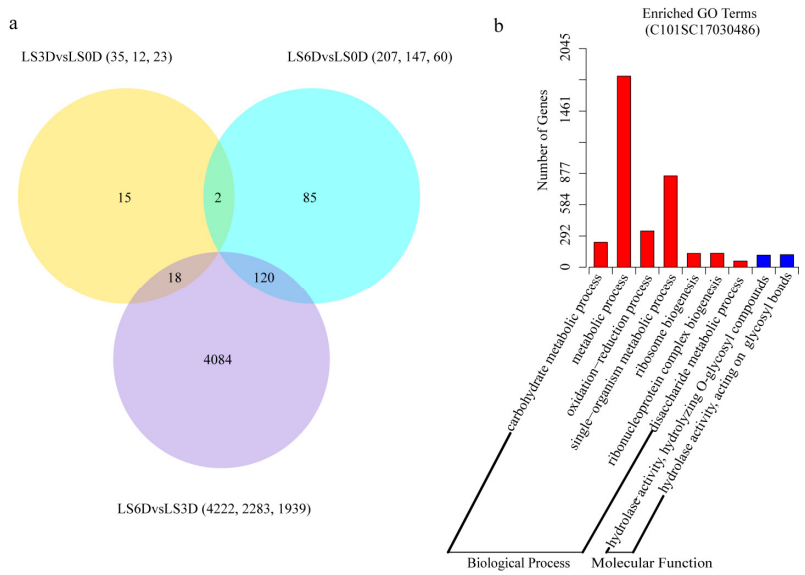


Fig. 5. (a) The quantity of DEGs in 'Lisa' avocado fruit pulp during after-ripening (b) histogram of GO for the DEGs in 'Lisa' avocado fruit pulp

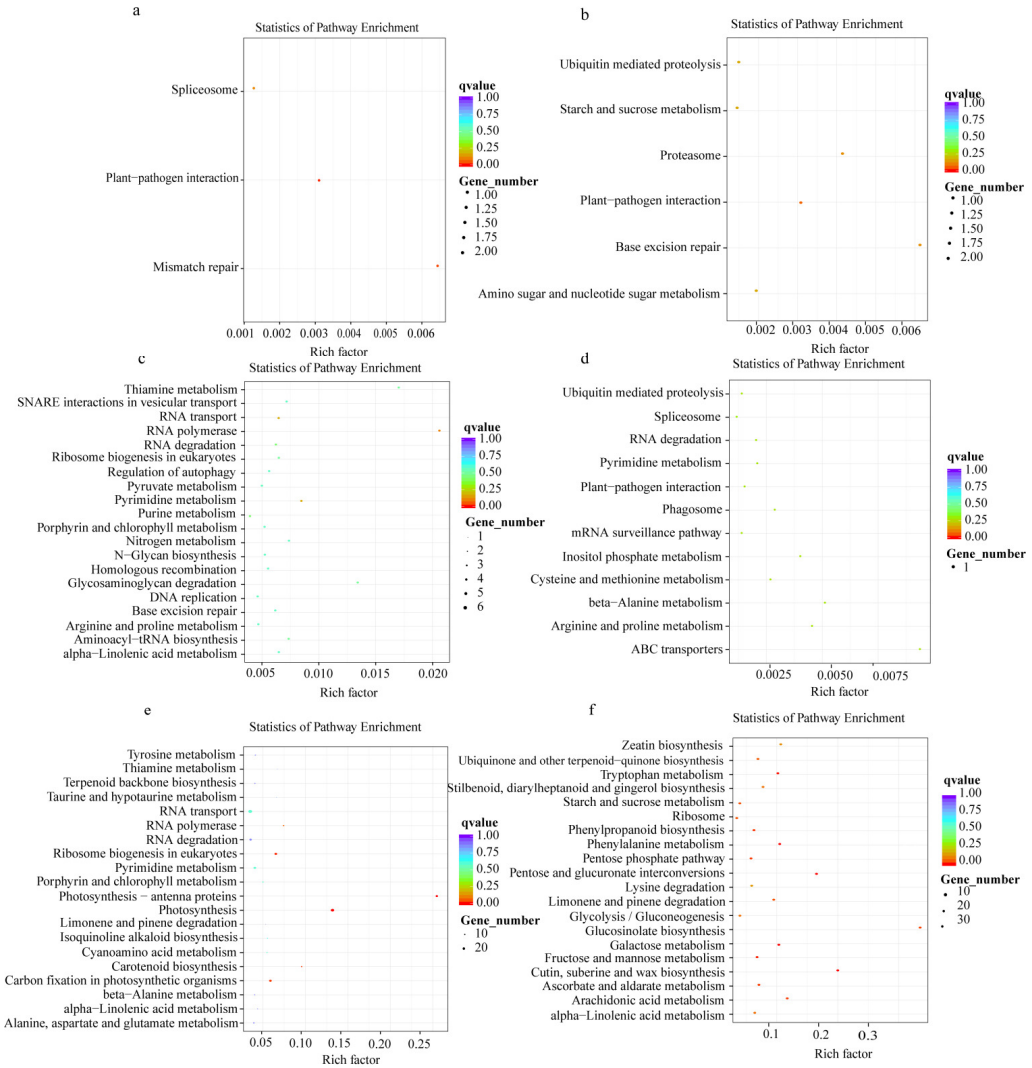


Fig. 6. KEGG pathway enrichment analysis of differentially expressed genes. (a) KEGG analysis of upregulated DEGs at 3D versus those at 0D. (b) KEGG analysis of downregulated DEGs at 3D versus those at 0D. (c) KEGG analysis of upregulated DEGs at 6D versus those at 0D. (d) KEGG analysis of downregulated DEGs at 6D versus those at 0D. (e) The KEGG analysis of upregulated DEGs at 6D versus those at 3D. (f) KEGG analysis of downregulated DEGs at 6D versus those at 3D. The unigenes belonging to each corresponding KEGG pathway is shown in Supplementary Table 2

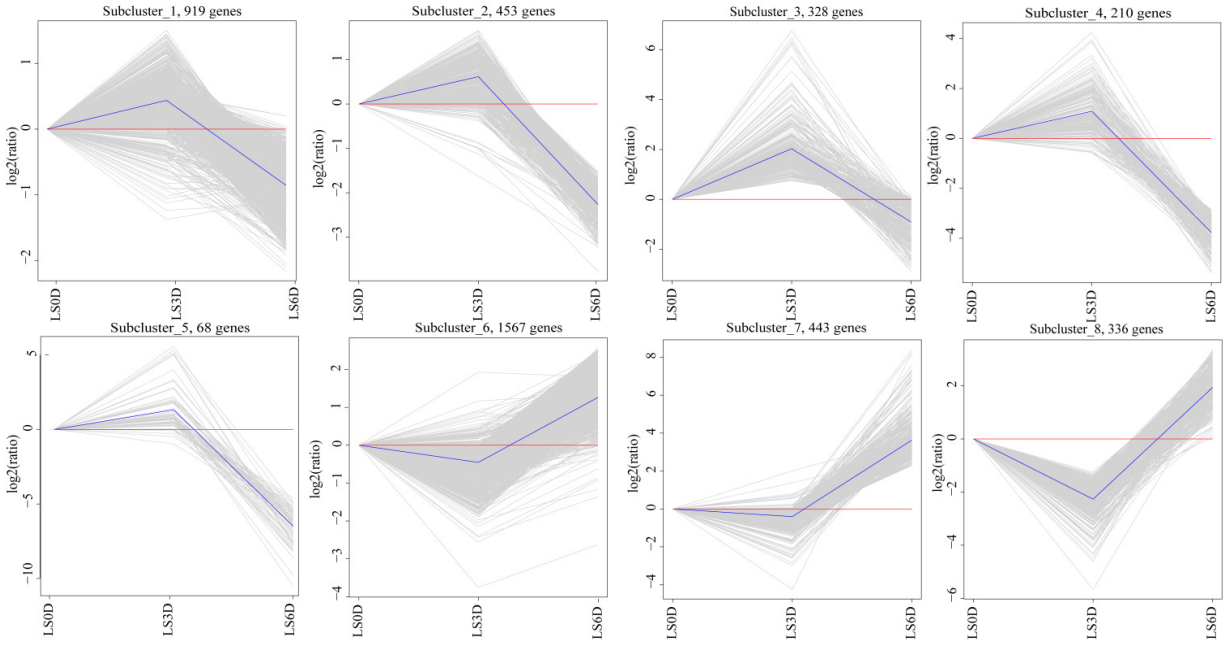


Fig. 7. Clustering of time-course data from RNAseq by short time-series expression miner analysis. Each box corresponds to one of the model temporal expression profiles. The unigenes belonging to each corresponding subcluster is shown in Supplementary Table 3. 0D, 3D, and 6D are the numbers of days into the 'Lisa' avocado fruit pulp after-ripening process. The y-axes indicate the fpkm values of each DEGs

Gene Transcription Factor

Transcription factor DEGs in 'Lisa' avocado fruit after-ripening were selected to construct a heat map which divided the avocado pulp transcription factor DEGs into two subclusters. The unigenes in subcluster I were upregulated at 6D. Those in subcluster II were induced at 3D but downregulated at 6D (Fig. 5A). Eleven transcription factors associated with ethylene production were identified in this experiment. Five of them were in subcluster I and six in subcluster II. Two NAC transcription factors were also identified: one in subcluster I, and the other in subcluster II. Other transcription factors like MYB, bHLH, and WRKY were also identified (Fig. 8).

Discussion

The effectiveness of RNAseq in 'Lisa' avocado fruit after-ripening

RNAseq detects low-expressing reads (Garber *et al*., 2011; Wang *et al*., 2014) and identifies novel transcripts (Handa *et al*., 2015). Nevertheless, avocado transcriptome data is still limited. The two parts known avocado transcriptome accrue disease and develop fruit, respectively. The 454 pyrosequencing technique characterizes the root diseases transcripts in avocado. Mahomed and Van den Berg (2011) obtained 10,000 reads and assembled 371 contigs by 454 pyrosequencing to identify the influence of *Phytophthora* root rot in avocado. Approximately 124 Mb of data and 7,685 contigs were realized by 454 pyrosequencing to identify the response gene for avocado root rot (Reeksting *et al*., 2014). Using 454 pyrosequencing, Djami-Tchatchou *et al*. (2012) found 70.6 Mb of sequence data and annotated 639 unigenes characterizing *Colletotrichum gloeosporioides* infection. The 454 pyrosequencing process characterized avocado development transcripts also. It generated 4,530,278 high-quality reads (931,834 generated by GS-FLX+, and 3,598,444 pairs generated by MiSeq) from various *Persea drymifolia* seeds, fruits, leaves, roots, stems, aerial buds, and flowers. These transcripts were implicated in fatty acid metabolism and fruit after-ripening (Ibarra-Laclette *et al*., 2015). There were 151,788 contigs generated for *Persea americana* mesocarp oil biosynthesis using 454 GS FLX (Kilaru *et al*., 2015). Avocado pulp transcriptome was analyzed using a Hiseq 4000 genome analyzer (Illumina, San Diego, CA, USA). Assembly of all the reads from the nine samples generated 235,082 transcripts and 151,545 unigenes with average lengths of 904 bp and 1,243 bp, respectively (Table 1). Of the sequences, 51.5% (78,054) were annotated to the NR database, and the balance to the others (Table 1).The avocado transcriptome data reported in this study increased the number of transcripts related to avocado after-ripening. The Differential Gene Expression analysis, KEGG Pathway Enrichment analysis and Gene Transcription Factor screening help explain the quality changes during avocado after-ripening.

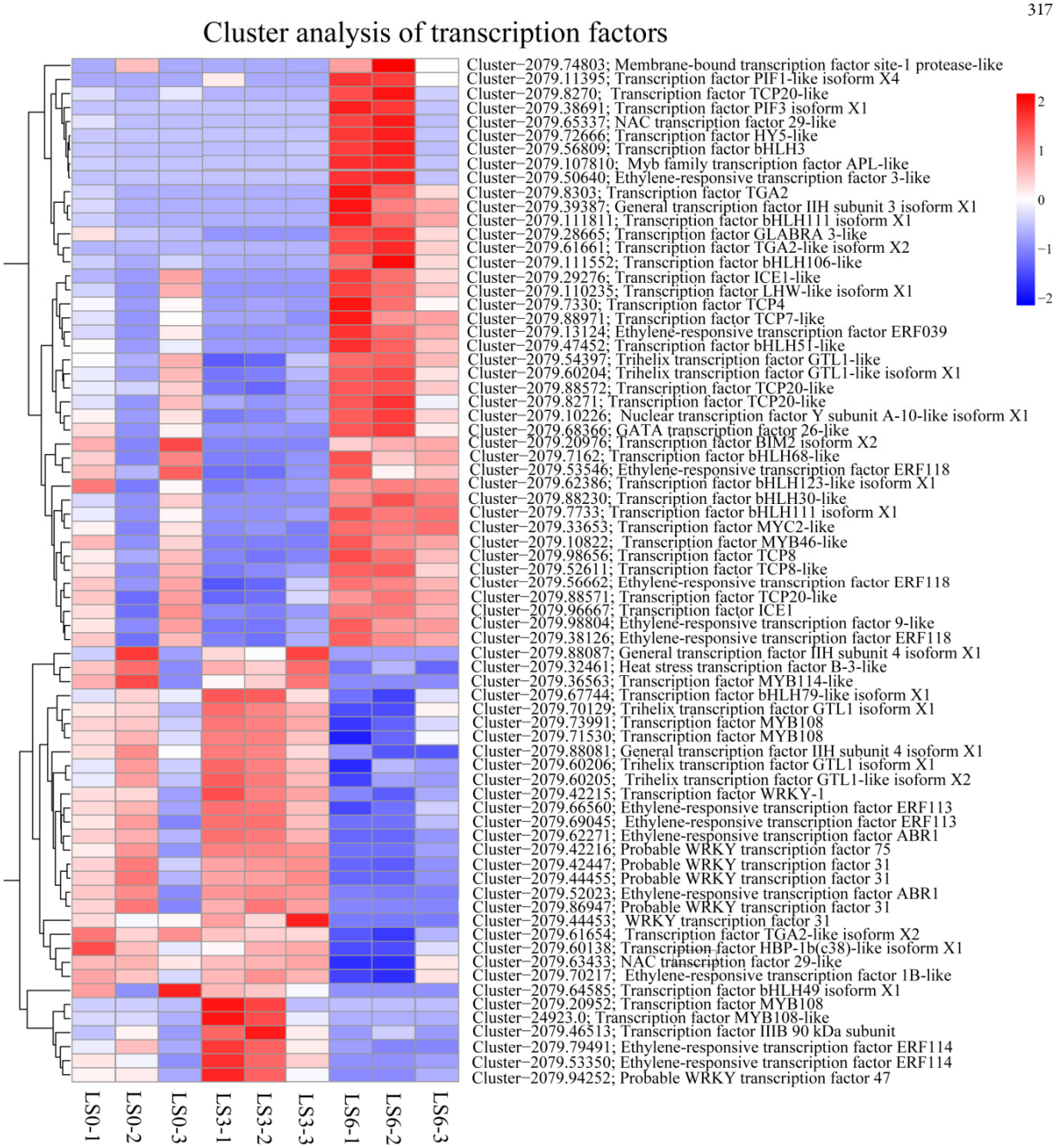


Fig. 8. Heat map of transcription factor DEGs in 'Lisa' avocado fruit pulp after-ripening. Three replicates of 'Lisa' avocado fruit pulp after-ripening at 0D. These include LS0-1, LS0-2, LS0-3, etc. at 3D and 6D. The heat map was constructed based on the fpkm values. All transcription factors cluster into two different subgroups.

Fatty acid variation and the potential biosynthesis pathway

Avocado is valued as a food crop because of its rich flavor, high oil quality and health benefits attributed to its fatty acid content. In the fruit pulp, the fatty acid content steadily increases during fruit development and reaches a plateau a few days after harvest (Werman and Neeman, 1987; Gaydou *et al*., 1987; Kilaru *et al*., 2015; Ibarra-Laclette *et al*., 2015). The fatty acid composition does not significantly changes during the after-ripening period. It consists mainly of oleic (C18:1), palmitic (C16:0), and linoleic (C18:2) acids, as well as ten to twenty others in lower concentrations (Bora *et al*., 2001). Our data indicated that the amounts of C16:0 (palmitic acid), C16:1 (9-*cis*-hexadecenoic acid), C18:1 (oleic acid), and C18:2 (linoleic acid) did not significantly change during the after-ripening process. Nevertheless, C18:3 (α-linolenic acid) increased significantly in the postclimacteric stage. In this study, one objective was to examine the avocado pulp transcriptome for the purpose of identifying the pathways involved in C18:3 promotion. The DEGs associated with the preclimacteric, climacteric, and postclimacteric stages

clustered in six different KEGG pathways related to fatty acid content variation. The "α-linolenic acid metabolism" (ko00592), "fatty acid degradation" (ko00071), and "fatty acid biosynthesis" (ko00061) pathways were clustered in both up- and downregulated DEGs. The "biosynthesis of unsaturated fatty acids pathway" (ko01040) was clustered in upregulated DEGs (Supplementary Table 4). The upregulated DEGs clustered in the KEGG pathway related to fatty acid metabolism had expression patterns differing from those reported in previous research. Earlier studies indicated that the number of fatty acid genes decreased during fruit after-ripening and may influence the α-linolenic acid content changes in avocado after-ripening. Whether different varieties, storage conditions or other factors lead to differences in expression patterns of DEGs remains to be further studied.

The transcription factor involved in avocado after-ripening

The transcription factors were expressed in all three stages of the after-ripening process. Since ethylene production starts at the preclimacteric stage and rapidly increases by the climacteric, ethylene could be a signaling molecule that influences avocado after-ripening. The RIN, NOR, and CNR genes encode MADS-box, NAC-domain, and SBP-box family transcription factors, respectively (Vrebalov *et al*., 2002; Manning *et al*., 2006; Giovannoni, 2007). They act upstream of ethylene biosynthesis and perform key functions involved in fruit ripening control (Elitzur *et al*., 2010; Shan *et al*., 2012). In this study, eleven ethylene-related- and two NAC transcription factors were identified. Five of the former and one of the latter were in subcluster I which upregulated at 6D. There were one NAC- and six ethylene-related transcription factors in subcluster II which upregulated at 3D but downregulated at 6D. The results suggest that the functions of these transcription factors differentiated during after-ripening regulation. Subcluster I, which upregulated at 6D (postclimacteric), might consist of transcription factors involved in senescence. Subcluster II, which upregulated at 3D (climacteric), may play an important role in after-ripening (Fig. 8). MYB, bHLH, and WRKY were also identified, and they might provide additional information as to which transcription factors are involved in avocado after-ripening.

Conclusions

There was no significant difference in total fatty acid content among the preclimacteric, climacteric, and postclimacteric stages of 'Lisa' avocado fruit pulp maturation. Nevertheless, the amount of C18:3 (α-linolenic acid) increased significantly in the climacteric and postclimacteric stages. RNAseq generated 35,082 transcripts and 151,545 unigenes, and 4,324 DEGs were obtained for the three stages. The DEGs indicated that the pathways related to "α-linolenic acid metabolism", "unsaturated fatty acid biosynthesis", "fatty acid degradation", "linoleic acid metabolism fatty acid biosynthesis", "fatty acid elongation", and "linoleic acid metabolism and fatty acid elongation" were identified. They may account for the variation in the fatty acid content of 'Lisa' avocado fruit pulp. KEGG analysis identified ethylene-related transcription factors, MYB, NAC, bHLH, and WRKY. Those transcription factors are thought to regulate avocado after-ripening. In summary, the 'Lisa' avocado transcriptome analyzed in this study provides a molecular basis for characterizing the hidden factors involved in avocado after-ripening.

Acknowledgements

This work was supported by the Science and Technology Program of Guangdong Province (Project No. 2016A020208002) and the Basic Scientific Research Project of Nonprofit Central Research Institutions (No. SSCRI-1630062014006).

References

Bora PS, Narain N, Rocha RVM, Paulo MQ (2001). Characterization of the oils from the pulp and seeds of avocado (cultivar: Fuerte) fruits. Grasas y Aceites 52:171-174.

Djami-Tchatchou AT, Straker CJ, Allie F (2012). 454 Sequencing for the identification of genes differentially expressed in avocado fruit (cultivar: Fuerte) infected by *Colletotrichum gloeosporioides*. Journal of Phytopathology 160(9):449-460.

Djeridane A, Yousfi M, Nadjemi B, Boutassouna D, Stocker P, Vidal N (2006). Antioxidant activity of some Algerian medicinal plants extracts containing phenolic compounds. Food Chemistry 97(4):654-660.

Dreher ML, Davenport AJ (2013). Hass avocado composition and potential health effects. Critical Reviews in Food Science and Nutrition 53(7):738-750.

Elitzur T, Vrebalov J, Giovannoni JJ, Goldschmidt EE, Friedman H (2010). The regulation of MADS-box gene expression during ripening of banana and their regulatory interaction with ethylene. Journal of Experimental Botany 61(5):1523-1535.

Elitzur T, Yakir E, Quansah L, Zhangjun F, Vrebalov JT, Khayat E, ... Friedman H (2016). Banana MaMADS transcription factors are necessary for fruit ripening and molecular tools to promote shelf-life and food security. Plant Physiology 171(1):380-391.

Gapper NE, McQuinn RP, Giovannoni JJ (2013). Molecular and genetic regulation of fruit ripening. Plant Molecular Biology 82(6):575-591.

Garber M, Grabherr MG, Guttman M, Trapnell C (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. Nature Methods 8(6):469-477.

Gaydou EM, Lozano Y, Ratovohery J (1987). Triglyceride and fatty acid compositions in the mesocarp of *Persea americana* during fruit development. Phytochemistry 26(6):1595-1597.

Giovannoni JJ (2007). Fruit ripening mutants yield insights into ripening control. Current Opinion in Plant Biology 10(3):283-289.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, ... Regev A (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29(7):644-652.

Handa Y, Nishide H, Takeda N, Suzuki Y, Kawaguchi M, Saito K (2015). RNA-seq transcriptional profiling of an arbuscular mycorrhiza provides insights into regulated and coordinated gene expression in *Lotus japonicus* and *Rhizophagus irregularis*. Plant and Cell Physiology

56(8):1490-1511.

Hara A, Radin NS (1978). Lipid extraction of tissues with a low-toxicity solvent. Analytical Biochemistry 90(1):420-426.

Ibarra-Laclette E, Méndez-Bravo A, Pérez-Torres CA, Albert VA, Mockaitis K, Kilaru A, ... Herrera-Estrella L (2015). Deep sequencing of the Mexican avocado transcriptome, an ancient angiosperm with a high content of fatty acids. BMC Genomics 16(1):599.

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, ... Yamanishi Y (2008). KEGG for linking genomes to life and the environment. Nucleic Acids Research 36(Suppl 1):D480-D484.

Katalinic V, Milos M, Kulisic T, Jukic M (2006). Screening of 70 medicinal plant extracts for antioxidant capacity and total phenols. Food Chemistry 94(4):550-557.

Kilaru A, Cao X, Dabbs PB, Sung HJ, Rahman MM, Thrower N, ... Ohlrogge JB (2015). Oil biosynthesis in a basal angiosperm: transcriptome analysis of *Persea americana* mesocarp. BMC Plant Biology 15(1):203.

Li B, Dewey C (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12(1):323.

Mahomed W, Van den Berg N (2011). EST sequencing and gene expression profiling of defence-related genes from *Persea americana* infected with *Phytophthora cinnamomi*. BMC Plant Biology 11(1):167.

Manning K, Tör M, Poole M, Hong Y, Thompson AJ, King GJ, ... Seymour GB (2006). A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nature Genetics 38(8):948-952.

Mao X, Cai T, Olyarchuk JG, Wei L (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics 21(19):3787-3793.

Ozdemir F, Topuz A (2004). Changes in dry matter, oil content and fatty acids composition of avocado during harvesting time and post-harvesting ripening period. Food Chemistry 86(1):79-83.

Reeksting BJ, Coetzer N, Mahomed W, Engelbrecht J, Van den Berg N (2014). *De novo* sequencing, assembly, and analysis of the root transcriptome of *Persea americana* (Mill.) in response to *Phytophthora cinnamomi* and flooding. PLoS ONE 9(2):e86399.

Shan W, Kuang J, Chen L, Xie H, Peng H, Xiao Y, ... Lu W (2012). Molecular characterization of banana NAC transcription factors and their interactions with ethylene signalling component EIL during fruit ripening. Journal of Experimental Botany 63(14):5171-5187.

Shu B, Li WC, Liu LQ, Wei YZ, Shi SY (2016). Transcriptomes of arbuscular mycorrhizal fungi and litchi host interaction after tree girdling. Frontiers in Microbiology 7(636):408.

Singleton VL, Rossi JA (1965). Colorimetry of total phenolics with phosphomolybdic-phosphotungstic acid reagents. American Journal of Enology and Viticulture 16(3):144-158.

Storey JD (2003). The positive false discovery rate: a Bayesian interpretation and the q-value. The Annals of Statistics 31(6):2013-2035.

Villa-Rodriguez J, Molina-Corral J, Ayala-Zavala F, Olivas G, Gonzalez-Aguilar G (2011). Effect of maturity stage on the content of fatty acids and antioxidant activity of 'Hass' avocado. Food Research International 44(5):1231-1237.

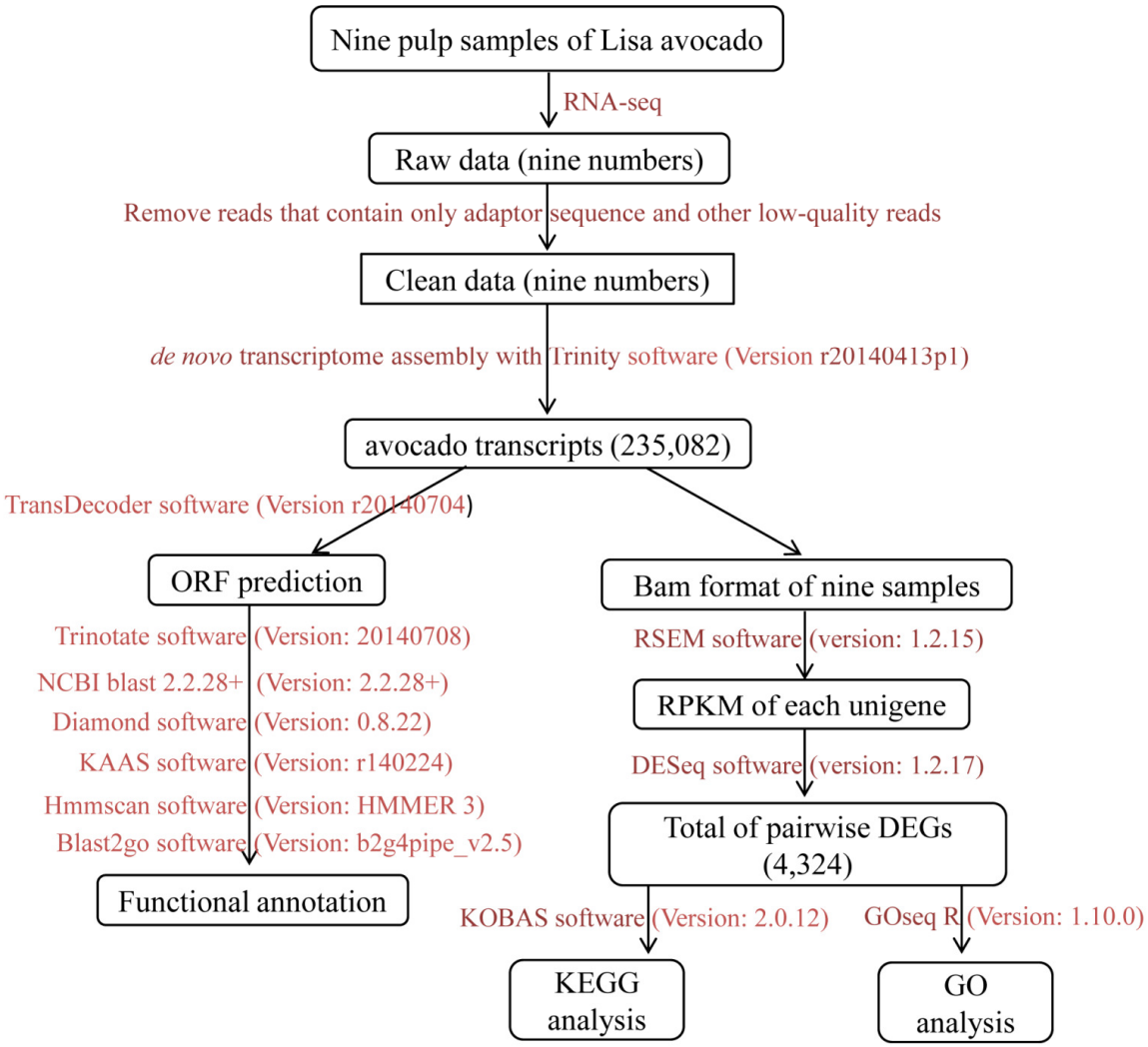
Vrebalov J, Ruezinsky D, Padmanabhan V, White R, Medrano D, Drake R, ... Giovannoni J (2002). A MADS-box gene necessary for fruit ripening at the tomato *ripening-inhibitor* (*rin*) locus. Science 296(5566):343-346.

Wang C, Gong B, Bushel PR, Thierry-Mieg J, Thierry-Mieg D, Xu J, ... Tong W (2014). The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. Nature Biotechnology 32(9):926-932.

Werman MJ, Neeman I (1987). Avocado oil production and chemical characteristics. Journal of the American Oil Chemists' Society 64(2):229-232.

Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biology 11(2):R14.

Zhang L, Zhang DS, Liu KD (2015). Environment analysis and policy for development of avocado industry in Hainan. Chinese Journal of Agricultural Resources and Regional Planning 36(4):78-84.



Supplementary Fig. 1. Flowchart of RNAseq analysis

Supplementary Table 1. Summary of read numbers based on the RNAseq data from 'Lisa' avocado fruit pulp during after-ripening

Sample name	Total reads(no)	Total bases(G)	GC content (%)	Q20 (%)	Q30 (%)
LS0D-1	42,199,452	6.33	46.36	97.31	93.13
LS0D-2	43,149,086	6.47	46.44	94.46	86.69
LS0D-3	40,039,878	6.01	46.65	97.03	92.51
LS3D-1	39,759,018	5.96	46.16	96.94	92.32
LS3D-2	46,378,784	6.96	46.46	97.04	92.54
LS3D-3	38,892,694	5.83	46.6	97.06	92.55
LS6D-1	42,154,638	6.32	46.43	96.84	92.12
LS6D-2	55,445,588	8.32	46.37	96.87	92.29
LS6D-3	46,039,402	6.91	46.32	96.66	91.68

The LS represents 'Lisa' avocado; 0D, 3D, 6D represent the time in the after-ripening process; 1, 2, 3 represents biological repeats

See attached files:

Supplementary Table 2

Supplementary Table 3

Supplementary Table 4