

Article

Single-Molecule Long-Read Sequencing of Avocado Generates Microsatellite Markers for Analyzing the Genetic Diversity in Avocado Germplasm

Yu Ge^D, Xiaoping Zang, Lin Tan^D, Jiashui Wang, Yuanzheng Liu, Yanxia Li, Nan Wang, Di Chen, Rulin Zhan * and Weihong Ma *

Haikou Experimental Station, Chinese Academy of Tropical Agricultural Sciences, Haikou 570102, China

* Correspondence: zhanrulin@catas.cn (R.Z.); zjwhma@catas.cn (W.M.); Tel.: +86-898-6679-4563 (R.Z.);

+86-898-6677-3067 (W.M.)

Received: 9 July 2019; Accepted: 1 September 2019; Published: 5 September 2019



Abstract: Avocado (*Persea americana* Mill.) is an important fruit crop commercially grown in tropical and subtropical regions. Despite the importance of avocado, there is relatively little available genomic information regarding this fruit species. In this study, we functionally annotated the full-length avocado transcriptome sequence based on single-molecule real-time sequencing technology, and predicted the coding sequences (CDSs), transcription factors (TFs), and long non-coding RNA (lncRNA) sequences. Moreover, 76,777 simple sequence repeat (SSR) loci detected among the 42,096 SSR-containing transcript sequences were used to develop 149,733 expressed sequence tag (EST)-SSR markers. A subset of 100 EST-SSR markers was randomly chosen for an analysis that detected 15 polymorphicEST-SSR markers, with an average polymorphism information content of 0.45. These 15markers were able to clearly and effectively characterize46 avocado accessions based on geographical origin. In summary, our study is the first to generate a full-length transcriptome sequence and develop and analyze a set of EST-SSR markers in avocado. The application of third-generation sequencing techniques for developing SSR markers is a potentially powerful tool for genetic studies.

Keywords: Persea americana; SMRT sequencing; simple sequence repeat; genetic relationship

1. Introduction

Avocado (*Persea americana* Mill.) belonging to the family Lauraceae of the order Laurales is native to Mexico and Central and South America, and is one of the most economically important subtropical/tropical fruit crops worldwide [1]. Taxonomic treatments differ considerably in terms of the circumscription and defining of infraspecific avocado entities [2–5]. Additionally, researchers have long considered that geographical isolation has likely resulted in the following three ecological races of avocado: Mexican (*P. americana* var. *drymifolia*), Guatemalan (*P. americana* var. *guatemalensis*), and West Indian (*P. Americana* var. *americana*) [1]. The Mexican race adapted to a Mediterranean climate, whereas the Guatemalan race originated in a tropical highland climate, and the West Indian race adapted to humid tropical lowland conditions [1].

Avocado is rich in lipids, sugars, proteins, minerals, vitamins, and other active ingredients [6–8]. Moreover, avocado production has increased worldwide [1]. One factor contributing to the increases in production and consumption is the expansion of avocado products into new global markets where avocado was previously unknown or scarce, includingChina, which is an emerging market for the production and consumption of avocado [1,9]. After avocado was first introduced and cultivatedin China in the late 1950s, selective breeding by some national scientific research bodies and other state farms have resulted in the development of more than 10 superior avocado accessions [9,10].



Additionally, natural crosses among avocado accessions have generated new hybrids on state and private farms, and some native accessions are increasingly produced in somewhat remote areas with distinct local environmental conditions [9,10]. Avocado is broadly grown and exploited in some provinces in southern China, including Hainan, Guangxi, Yunnan, and Taiwan [9,10]. The climatic conditions in these provinces are subtropical to tropical, which are ideal conditions for the cultivation of avocado [9,10].

The avocado germplasm should be precisely characterized to maximize its utility to breeders worldwide [1]. Specifically, a molecular characterization is required for analyses of the genetic relationships among avocado germplasm. Over the past two decades, studies involving various types of molecular markers have examined the genetic relationships among avocado germplasm [11–20]. Of the many available DNA markers, simple sequence repeats (SSRs) are commonly used for investigating plant genetics and breeding because they are widely distributed and abundant in plant genomes. They are also genetically codominant, highly reproducible, multi-allelic, and perfectly suitable for high-throughput genotyping [21–25]. Expressed sequence tag (EST)-derived markers in the genomic coding regions have an advantage over genomic DNA-derived markers, and can be efficiently amplified to reveal conserved sequences among related species [26]. There has recently been increasing interest in developing EST-SSR markers viahigh-throughput transcriptome sequencing. Thus, there has been rapid progress in the development of EST-SSR markers based on transcriptome data produced with second-generation sequencing technology for Lilium brownii var. viridulum Baker [27], crataegus Pinnatifida Bunge [28], Acer miaotaiense P. C. Tsoong [29], and Rosa hybrida hort. ex Lavalle [30]. Among the third-generation sequencing platforms, PacBio RS II, which is regarded as the first commercialized third-generation sequencer, is based on single-molecule real-time (SMRT) technology [31]. The PacBio RS II system can produce much longer reads than second-generation sequencing platforms, and has been applied to effectively capture full-length transcriptsequences for EST-derived marker development [32]. However, there are few reports regarding the application of EST-SSR markers developed with SMRT technology for crop breeding.

Single-molecule real-time technology has the following threemain advantages over second-generation sequencing options: it generates longer reads, it has higher consensus accuracy, and it is less biased [33]. A previous study revealed that SMRT technology can precisely ascertain alternative polyadenylation sites and full-length splice isoforms, and also detect a higher isoform density than that for the reference genome [34]. The application of SMRT technology for nearly 3 years has helped to elucidate the complexity of the transcriptome and molecular mechanism underlying the metabolite synthesisin safflower [31], *Zanthoxylum bungeanum* Maxim. [32], *Trifolium pratense* L. [34], *Saccharum officinarum* L. [35], *Panicum virgatum* L. [36], *Medicago sativa* L. [37], *Zanthoxylum planispinum* Sieb. [38], *Cynodon dactylon* L. Pers. [39], *Camellia sinensis* L. O. Ktze. [40], and *Cassia obtusifolia* L. [41].

In the previous study, we had generated the first full-length transcriptome sequence of avocadobased on SMRT technology andthe short-reads obtained in this previous study involving second-generation transcriptome sequencing were used to correct the transcripts that were obtained with SMRT technology [42]. In this study, we functionally annotated sequences andcompleted SSR mining experiments from SMRT technology in avocado mesocarp. We also predicted the coding sequences (CDSs), transcription factors (TFs), and long non-coding RNA (lncRNA) sequences. Furthermore, we identified a set of EST-SSR markers, and assessed their utility for determining the genetic diversity among 46 selected avocado accessions from various locations in southern China. The generated data enabled the broad and distinct visualization of the genetic diversity in the analyzed avocado germplasm. The results of this study represent useful genetic and transcriptome information to support future research on avocado.

2. Materials and Methods

2.1. Sample Collection, DNA Extraction, and RNA Extraction

For transcriptome analyses, avocado fruits (cultivar 'Hass') were harvested from April to September 2018 from six 10-year-old trees (grafted onto Zutano clonal rootstock) growing at the Chinese Academy of Tropical Agricultural Sciences (CATAS; Danzhou, Hainan, China; latitude 19°31' N, longitude 109°34' E, and 20 m above sea level). Each biological replicate comprised samples from two trees. Specifically, fruits that developed during the main flowering season (i.e., February 2018) were marked, after which samples were collected at five time-points (75, 110, 145, 180, and 215 days after full bloom) until the fruits reached physiological maturity (i.e., able to ripen after harvest). The fruits were randomly collected for each biological replicate during each developmental stage. Fruits were quickly brought to the laboratory, after which the mesocarp (pulp) was separated from the seedand then immediately frozen at -80 °C for subsequent transcriptome analyses. Total RNA was extracted with a Plant RNA Kit (OMEGA Bio-Tek, Norcross, GA, USA).

For kompetitive allele-specific PCR (KASP) genotyping and EST-SSR detection, seven commercial cultivars and 39 native accessions were selected. These native accessions were obtained from the CATAS (Danzhou, Hainan, China; latitude 19°31' N, longitude 109°34' E, and 20 m above sea level), Daling State Farm (DLSF; Baisha, Hainan, China; latitude 19°14' N, longitude 109°14' E, and 60 m above sea level), Mengmao State Farm (MMSF; Ruili, Yunnan, China; latitude 24°00' N, longitude 97°50' E, and 240 m above sea level), and Guangxi Vocational and Technical College (GVTC; Nanning, Guangxi, China; latitude 22°29' N, longitude 108°11' E, and 79 m above sea level). Details regarding the avocado germplasm are provided in Table S1. Genomic DNA was extracted from fresh leaves as described by Ge [43].

2.2. PacBiocDNA Library Construction and Sequencing

Poly-T oligo-attached magnetic beads were used to purify the mRNA from the total RNA extracted from 15 mesocarp (pulp) samples collected at each analyzed developmental stage. The mRNA from all five developmental stages was combined to serve as the template to synthesize cDNA with the SMARTer PCR cDNA Synthesis Kit (Clontech, Mountain View, CA, USA). After a PCR amplification, quality control check, and purification, full-length cDNA fragments were acquired according to the BluePippin Size Selection System protocol, ultimately resulting in the construction of a cDNA library (1–6 kb). Selected full-length cDNA sequences were ligated to the SMRT bell hairpin loop. The concentration of the cDNA library was then determined with the Qubit 2.0 fluorometer, whereas the quality of the cDNA library was assessed with the 2100 Bioanalyzer (Agilent). Finally, one SMRT cell was sequenced with the PacBio RSII system (Pacific Biosciences, Menlo Park, CA, USA).

2.3. IlluminacDNA Library Construction and Sequencing

Oligo-(dT) magnetic beads were used to purify the mRNA from the total RNA extracted from 15 mesocarp (pulp) samples from five developmental stages. Three replicates were analyzed for each developmental stage. Samples from each developmental stage underwent an RNA-sequencing analysis, with three biological replicates per sample. The fragmentation step was completed with divalent cations in heated 5× NEBNext First Strand Synthesis Reaction Buffer. First-strand cDNA was synthesized with a series of random hexamer primers and reverse transcriptase, after which the second-strand cDNA was generated with DNA polymerase I and RNase H. The cDNA libraries were constructed by ligating the cDNA fragments to sequencing adapters and amplifying the fragments by PCR. The libraries were then sequenced with the Illumina HiSeq 2000 platform (Nanxin Bioinformatics Technology Co., Ltd., Guangzhou, China).

2.4. Quality Filtering and Correction of PacBio Long-Reads

Raw reads were processed into error-corrected reads of insert (ROIs) using an isoform sequencing pipeline, with minimum full pass = 0.00 and minimum predicted accuracy = 0.80. Next, full-length, non-chimeric transcripts were detected by searching for the poly-A tail signal and the 5' and 3' cDNA primer sequences in the ROIs. Iterative clustering for error correction was used to obtain high-quality consensus isoforms, which were then polished with QuiverVersion 1.0. The low-quality full-length transcript isoforms were corrected based on Illumina short-reads with the default setting of the Proovread program. High-quality and corrected low-quality transcript isoforms were confirmed as nonredundant with the CD-HIT software.

2.5. Functional Annotation

Genes were functionally annotated based on a BLASTX search (*E*-value threshold of 10⁻⁵) of the following databases: Clusters of Orthologous Groups of proteins (KOG/COG) (available online: http://www.ncbi.nlm.nih.gov/COG/), Non-supervised Orthologous Groups (eggNOG) (available online: http://eggnogdb.embl.de/#/app/home), Swiss-Prot (a manually annotated and reviewed protein sequence database, available online: http://www.uniprot. org/), Pfam (assigned with the HMMER3.0 package, available online: http://pfam.xfam.org/), and NCBI nonredundant protein sequence (Nr) (availableonline: http://www.ncbi.nlm.nih.gov/). Additionally, the KEGG Automatic Annotation Server [44] was used to assign these genes to Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathways (available online: http://www.genome.jp/kegg/). The unigenes were annotated with gene ontology (GO) terms (available online: http://www.geneontology. org/) with the Blast2GO (version 2.5) program [45] based on the BLASTX matches in the Pfam and Nr databases (*E*-value threshold of 10⁻⁶).

2.6. Mining of EST-SSR Markers

The MISA (version 1.0) program, with the following default settings, was used to locate SSRs: a minimum of five repeats; a minimum motif length of 5 for tri- and hexanucleotides, 6 for dinucleotides, and 10 for single nucleotides.

2.7. Analyses of Detected Coding Sequences, Transcription Factors, and Long Non-Coding RNA Features

The open reading frames (ORFs) detected with the TransDecoder (version 3.0.0) program were designated as putative CDSs if they satisfied the following criteria: (1) An ORF was detected in a transcript sequence; (2) the log-likelihood score was >0, and was similar to what was calculated with the GeneID software; (3) the score was higher when the ORF was in the first reading frame than when the ORF was in the other five reading frames; (4) if a candidate ORF was within another candidate ORF, the longer one was reported. However, a single transcript could be associated with multiple ORFs (because of operons and chimeras); and (5) the putative encoded peptide matched a Pfam domain.

Transcription factor gene families were identified based on categorically defined TF families and criteria from the KO, KOG, GO, Swiss-Prot, Pfam, Nr, and Nt databases. Specifically, the default parameters of the iTAK (version 1.2) program were used. The methods used to identify and classify TFs were previously described by Perez-Rodriguez [46].

The following four computational tools were combined to sort non-protein-coding RNA candidates from putative protein-coding RNAs among the transcripts: the Coding Potential Calculator (CPC), Coding-Non-Coding Index (CNCI), Coding Potential Assessment Tool (CPAT), and Pfam database. Transcripts longer than 200 nt, with more than two exons, were selected as lncRNA candidates and were further screened with CPC/CNCI/CPAT/Pfam, which distinguished the protein-coding genes from the non-coding genes.

2.8. Assignment of the Native Avocado Accessions with an Unknown Race

To validate the origins of the 33 native accessions with anunknown race, six primers for race-specific single nucleotide polymorphism (SNP) loci were used for KASP genotyping listed in Table S2 [47]. The primer mix, which was prepared and used as described by KBioscience (http://www.kbioscience.co.uk), comprised 46 μ L dH₂O, 30 μ L common primer (100 μ M), and 12 μ L each tailed primer (100 μ M). The SNPs were amplified by PCR in a thermal cycler with a 5- μ L solution consisting of 1× KASP Master mix, 10 ng genomic DNA, and the SNP-specific KASP assay mix. The following PCR amplification conditions were the same as those used for each SNP assay: 94 °C for 15 min; 10 touchdown cycles of 94 °C for 20 s, and 58–61 °C for 60 s (decreasing by 0.8 °C per cycle); 35 cycles of 94 °C for 20 s and 57 °C for 60 s. The resulting data were analyzed with the Roche LightCycler 480 (version 1.50.39) program.

2.9. Identification of EST-SSR Markers

To screen the EST-SSR loci, primers based on the sequences flanking the selected microsatellite loci were designed with the Primer3 program; the PCR products ranged from 100 to 300 bp. All assigned marker names included Pa-eSSR to indicate their association with *P. Americana* and EST-SSRs. A subset of 100 EST-SSR primer pairs was randomly selected for validation by a PCR amplification with the same conditions as those described by Ge [43]. The PCR products were analyzed with the 96-capillary 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA, USA). The detection system included 8.9 μ L HIDI (Applied Biosystems), 0.1 μ L LIZ (Applied Biosystems), and 1 μ L PCR products (1:10 dilution). A lack of amplification was considered indicative of a null allele.

2.10. Data Analysis

The number of observed alleles (Na), effective number of alleles (Ne), observed heterozygosity (Ho), expected heterozygosity (He), and polymorphism information content (PIC) of each EST-SSR was assessed with the POPGEN (version 1.32) program [48]. A cluster analysis was performed with PowerMarker (version 3.25) [49]. The cophenetic correlation coefficient was computed for the dendrogram after the construction of a cophenetic matrix to measure the goodness of fit between the original similarity matrix and the dendrogram. Bootstrap support values were obtained from 1000 replicates. A neighbor-joining tree was constructed based on shared alleles, and visualized with the MEGA6.0 software [50].

3. Results

3.1. General Properties and Functional Annotations Based on Public Databases of Single-Molecule Long-Reads

Figure 1 presents the length distribution of 651,260 reads of insert in avocado mesocarp, and the classification of the reads of insert in avocado mesocarpis listed in Figure 2. The SMRT and Illumina HiSeq 2000 sequencing data were deposited in the GenBank database (accession numbersPRJNA551932 and PRJNA541745, respectively). Gene annotations according to a BLASTX algorithm indicated that the 71,627 avocado transcripts significantly matched sequences in the COG, GO, KEGG, KOG, Pfam, Swiss-Prot, eggNOG, and Nr databases, respectively (Table S3). The species with the most matches for the transcripts were Nelumbo nucifera Gaertn. (41.18% of transcripts), Vitis vinifera L. (10.76% of transcripts), Elaeis guineensis Jacq. (8.88% of transcripts), and Phoenix dactylifera L. (6.90% of transcripts). The homology with the other species was relatively low (1.14%–2.54% of transcripts; Figure 3). To further predict and classify the functions of the annotated transcripts, we analyzed their matching GO terms, eggNOG classifications, and KEGG pathway assignments. A total of 45,134 transcripts were assigned to 51 subcategories of the three main GO functional categories as follows: 106,390 transcripts for biological processes, 45,931 transcripts for cellular components, and 69,120 for molecular functions (Figure 4a, Table S4). Next, 70,205 transcripts were functionally classified into 25eggNOG categories (Figure 4b, Table S5). Among the 26 categories, the most heavily represented group was posttranslational modification, protein turnover, chaperones (6410 transcripts, 8.94%), followed by

signal transduction mechanisms (4189 transcripts, 5.84%) and transcription (3868 transcripts, 5.39%). Only 20 and 6 transcripts belonged to the cell motility and nuclear structure categories, respectively. Finally, 33,310 transcripts were assigned to 129 KEGG pathways (Table S6). The most represented pathways were related to carbon metabolism (1678 transcripts), protein processing in endoplasmic reticulum (1649 transcripts), and biosynthesis of amino acids (1503 transcripts).



Figure 1. Length distribution of 651,260 reads of insert in avocadomesocarp.



Figure 2. Classification of reads of insert in avocadomesocarp.



Figure 3. Species most closely related to avocado based on the NCBI nonredundant protein sequence database.



(a) Figure 4. Cont.



Figure 4. Functional classification of transcripts. The predicted functions were based on Gene Ontology (a) and Non-supervised Orthologous Groups (b) databases.

3.2. Predictions of ORFs, TFs, and lncRNAs

A total of 73,946 ORFs were predicted, 61,523 of which were complete CDSs. The number and length distribution of proteins encoded by the CDS regions are presented in Figure 5 and Additional file 1. A total of 7969 putative avocado TFs distributed in 203 families were identified (Table S7). The most abundant TF categories included RLK-Pelle_DLSV (241) and C3H (240). Additionally, the CPC, CNCI, CPAT, and Pfam database were combined to distinguish lncRNA candidates from putative protein-coding RNAs among the unannotated transcripts. Analyses with the CPC, CNCI, CPAT, and Pfam database revealed 7869, 6444, 16,464, and 15,579 transcripts longer than 200 nt with more than two exons as lncRNA candidates. A total of 3596 lncRNA transcripts were predicted (Figure 6).



Figure 5. Distribution of 61,523 complete coding sequences for the avocado open reading frames.



Figure 6. The number of long non-coding RNA transcripts predicted in avocado based on the Coding Potential Calculator, Coding-Non-Coding Index, Coding Potential Assessment Tool, and Pfam database.

3.3. Frequency and Distribution of Various Types of EST-SSR Loci

The 75,946 transcript sequences comprising 170,959,769 bp detected in this study included 42,096 sequences containing 76,777 SSR loci (Table 1). Of these SSR-containing transcript sequences, 19,825 harbored more than one SSR locus. Mononucleotide motifs were the most abundant (44,800, 58.35%), followed by di- (18,903; 24.62%), tri- (11,724, 15.27%), tetra- (788, 0.01%), hexa- (321, 0.00%), and pentanucleotide (241, 0.00%) motif repeats (Table 2).

Table 1. Details regarding the simple seq	aence repeats (SSRs) identified from single-molecule real-time
(SMRT) sequencing in avocado mesocar).

Source	Number		
Total number of sequences examined	75,956		
Total size of examined sequences (bp)	170,959,769		
Total number of identified SSRs	76,777		
Number of SSR containing sequences	42,096		
Number of sequences containing more than 1 SSR	19,825		
Number of SSRs present in compound formation	12,675		

Table 2. Details regarding the number of repeating units at avocado expressed sequence tag-simple sequence repeat (EST-SSR) loci.

SSR Motif	Repeat Unit Number								
Length	5	6	7	8	9	10	>10	Total	%
Mono-	-	-	-	-	-	8951	35,849	44,800	58.35
Di-	-	4017	2773	2489	2013	1547	6064	18,903	24.62
Tri-	6129	2838	1237	720	403	193	204	11,724	15.27
Tetra-	541	175	41	21	8	-	2	788	0.01
Penta-	172	67	1	1	-	-	-	241	0.00
Hexa-	228	72	15	3	2	-	1	321	0.00
Total	7070	7169	4067	3234	2426	10,691	42,120	76,777	
%	9.21	9.34	5.30	4.21	3.16	13.92	45.14		

There were 5–1343 SSRs per locus. Moreover, SSRs with more than 10 repeats were the most abundant, followed by those with 10, 6, and 5 random repeats. Among the 139 different repeat types, $(A/T)_n$ was the most common (56.63%). The six other main motif types were $(AG/CT)_n$ (19.14%), $(AAG/CTT)_n$ (5.97%), $(AT/AT)_n$ (3.35%), $(AGC/CTG)_n$ (2.18%), and $(AC/GT)_n$ (2.04%) (Table S8).

3.4. Development of Polymorphic EST-SSR Markers, Analysis of Genetic Diversity, and KASP genotyping

Using Primer3, we developed 149,733 EST-SSR markers from the 49,911 SSR loci (Table S9). To verify the amplification of the EST-SSR markers, a subset of 100 EST-SSR markers was randomly chosen and tested with seven accessions from various regions in southern China (Table S10). The primers for 30 of the tested markers generated amplification products, whereas 37 primer pairs amplified nonpolymorphic products and 33 did not produce clear amplicons. The 30 polymorphic EST-SSR markers, which included 15 di-, 5 tri-, 5 tetra-, 2 penta-, and 3hexanucleotidemotif-based markers, were further verified with 46 avocado accessions. Finally, 15 polymorphic EST-SSR markers, with missing allele frequencies <10% for all 46 avocado accessions, were selected for subsequent analyses of genetic diversity (Table S11). A total of 71 alleles in the 46 avocado accession-specific and the other 63 alleles were generally found in multiple accessions (Table S11). The eight accession-specific alleles were from the following accessions: Renong No. 4, Renong, No. 5; Renong No. 6, Guiyan No. 8, Daling No. 5, Daling No. 6, RL chang, and RL yuan.

The 15polymorphic EST-SSRs were applied to evaluate diversity parameters (Table 3). The Na amplified per SSR locus varied from 2 to 10, with a mean of 4.73. The Ne varied from 1.04 to 4.39, with an average of 2.31, and Ho ranged from 0.04 to 0.93, with an average of 0.49. The He ranged from 0.04 to 0.77, with an average of 0.50, and PIC values ranged from 0.04 to 0.74, with an average of 0.45.

Marker Name	Transcript ID	Na ¹	Ne ²	Ho ³	He ⁴	PIC ⁵
Pa-eSSR-17	F01_cb7709_c10/f1p0/2063	8	3.02	0.61	0.67	0.62
Pa-eSSR-18	F01_cb7876_c2/f1p0/2226	10	3.09	0.61	0.68	0.65
Pa-eSSR-19	F01_cb1803_c26/f1p0/2838	6	2.04	0.63	0.51	0.46
Pa-eSSR-20	F01_cb10663_c1/f1p0/2458	3	1.87	0.50	0.46	0.40
Pa-eSSR-21	F01_cb15691_c2/f1p0/2049	3	2.41	0.50	0.58	0.50
Pa-eSSR-22	F01_cb3034_c12/f2p0/2705	5	2.85	0.67	0.65	0.60
Pa-eSSR-23	F01_cb12182_c0/f6p2/1774	3	1.47	0.28	0.32	0.29
Pa-eSSR-24	F01_cb13109_c0/f3p0/1635	5	2.80	0.48	0.64	0.58
Pa-eSSR-25	F01_cb1901_c3/f1p1/2722	2	1.04	0.04	0.04	0.04
Pa-eSSR-26	F01_cb7204_c7/f10p1/2700	3	2.65	0.93	0.62	0.55
Pa-eSSR-27	F01_cb10594_c1/f1p0/4058	3	1.40	0.33	0.29	0.27
Pa-eSSR-28	F01_cb9432_c36/f1p2/1811	5	1.56	0.43	0.36	0.33
Pa-eSSR-29	F01_cb15387_c0/f3p0/1548	8	4.39	0.49	0.77	0.74
Pa-eSSR-30	F01_cb12814_c24/f1p0/3423	4	2.67	0.53	0.62	0.55
Pa-eSSR-31	F01_cb10835_c0/f4p0/2019	3	1.33	0.28	0.25	0.22
	Total	71				
	Mean	4.73	2.31	0.49	0.50	0.45

Table 3. Diversity parameters associated with 15 polymorphic EST-SSRs analyzed in 46 avocado accessions.

¹ Number of observed alleles; ² effective number of alleles; ³ observed heterozygosity; ⁴ expected heterozygosity; ⁵ polymorphism information content.

Six race-specificKASP markers were used to determine the race of 33 avocado accessions with an unknown race. The KASP genotyping results demonstrated that all 33 avocado accessions were Guatemalan \times West Indian hybridsbased on the corresponding genotype of each racial avocado (Table S2).

A cluster analysis grouped the 46 accessions into two major sections (Figure 7). The dendrogram revealed a clear separation between the native avocado accessions from Hainan province and those from Guangxi and Yunnan provinces. In cluster I, 19 Guatemalan × West Indian hybrids were clustered into two sub-sections. Sub-cluster I-I consisted of 13native Guatemalan × West Indian hybrids from Guangxi province. Sub-cluster I-II contained two native Guatemalan × West Indian hybrids from Yunnan province.Cluster II comprised 27 Guatemalan × West Indian hybrids from Hainan province. Among these hybrids, 15 and 6were obtained from the CATAS and DLSF, respectively.



Genetic distance

Figure 7. Neighbor-joining consensus tree of 1000 bootstrap replicates revealing the phylogenetic relationships among the 46 analyzed avocado accessions based on the shared alleles for the 15 EST-SSR markers. GVTC, native avocado accessions from Guangxi Vocational and Technical College; MMSF, native avocado accessions from Mengmao State Farm; CATAS, native avocado accessions from the Chinese Academy of Tropical Agricultural Sciences; and DLSF, native avocado accessions from Daling State Farm. The native avocado accessionslabeled withan asterisk originated from other regions.

of a principal coordinate analysis (PCoA). On the basis of the first coordinate, which accounted for 21.71% of the total variation, the accessions were generally distributed in two groups. The native avocado accessions from Hainan and Yunnan provinceswere basically grouped separately from the native avocado accessions from Guangxi province. The second coordinate accounted for 10.06% of the total variation.Finally, we observed that the native avocado accessions were generally grouped according to their geographical origins.



Principal coordinate 1 (21.71%)

Figure 8. Principal coordinate analysis of 46 avocado accessions based on the 15 EST-SSR markers. POP1, avocado accessions fromFlorida, USA; POP2, native avocado accessions from theChinese Academy of Tropical Agricultural Sciences; POP3, native avocado accessions from Mengmao State Farm; POP4, native avocado accessions from Daling State Farm; and POP5, native avocado accessionsfrom Guangxi Vocational and Technical College.

4. Discussion

Transcriptome sequencing is a useful technique for obtaining a large number of transcripts for organisms lacking a reference sequence, at least partly because it is inexpensive and can be completed rapidly [51–53]. To date, several short-read next-generation sequencing (NGS) transcriptome databases have been developed for avocado mesocarp samples [54,55] and avocado mixed tissue samples [18,56]. However, both the number and length of the transcript sequences derived from these short-read NGS studies have hampered their application ingenetics and molecular biology research [41]. One of the advances in sequencing technology has been the development of the long-read SMRT sequencing

technique, which enables researchers to obtain a substantial number of full-length sequences from a cDNA library [32]. In the current study, we applied the PacBio SMRT system to generate and analyze the full-length transcriptome of avocado mixed mesocarp samples collected at various developmental stages. The 25.79 Gb SMRT data produced in this study provide the first comprehensive insights into the avocado mesocarp, which is the most economically valuable organ of this fruit species, and might serve as the genetic basis for future research on avocado. Interestingly, the full-length transcriptome sequence described herein is also the first such sequence for a plant species from the family Lauraceae.

In this study, 93.82% (71,627 of 76,345) of the nonredundant transcripts were annotated based on similarities with sequences in public databases. Thus, a greater proportion of transcripts were annotated in this study than in previous investigations involving NGS data for various avocado races (49.00%) [18] and for avocado mesocarp samples (57.50%) [55]. We determined that the mean length of the avocado nonredundant transcripts was2330 bp, implying that our sequences were long enough to represent full-length transcripts. Additionally, this mean length was in between the mean lengths obtained for other species, including *Z. bungeanum* (3414 bp) [32], *T. pretense* (2789 bp) [34], *M. sativa* (1706 bp) [37], *Z. planispinum* (1781 bp) [38], *C. sinensis* (1781 bp) [40], and *Arabidopsis pumila* (2194 bp) [57]. Moreover, the 76,345 nonredundant transcripts derived from the 25.79 Gb clean PacBio SMRT data produced in this study may facilitate future research on the physiology, biochemistry, and molecular genetics of avocado and related species.

A previous study indicated that lncRNAs may be important for the gene regulation in eukaryotic cells, especially during some key biological processes [58]. However, the number of lncRNAs encoded in genomes as well as their characteristics remains largely unknown [59]. Predicting and functionally annotating lncRNAs is challenging, but valuable because they are not orthologous and there is a lack of homologous sequences between closely related species [38]. Unfortunately, very few of the lncRNA functions have been elucidated [60,61]. Hence, the lncRNA information for one species is not suitable for predicting the lncRNAs in another species. In this study, 3596 avocado transcript sequences (accounting for 4.71% of the total number of nonredundant transcripts) were putatively predicted aslncRNAs. This almost completely uncharacterized gene pool may include genes associated with agronomically relevant traits related to the most economically valuable organ (mesocarp).

The accurate identification of avocado germplasm races is needed to ensure that germplasm collections are optimally used by plant breeders and farmers worldwide [1]. The traditional assignment of avocado races based on morphological traits is imprecise because of environmental effects and a limited number of applicable characteristics [17]. Molecular-based characterizations are more consistent and valid for assigning avocado genotypes. We previously confirmed the universality of six race-specific KASP markers [47]. These markers were used in the current study to identify avocado accessions with an unknown race, with implications for the application of available avocado germplasms for breeding and resource conservation. Interestingly, the KASP genotyping results revealed that all of the native avocado accessions included in this study are Guatemalan \times West Indian hybrids. The reason for this observation might be related to theintroduction of avocado cultivars and the climates of the sample collection regions. First, the major avocado cultivars grown commercially are typically hybrids of three races (i.e., mainly Guatemalan × West Indian and Guatemalan × Mexican hybrids) [1]. Since the late 1950s, Guatemalan × West Indian and Guatemalan × Mexican hybrids have been brought into China from other countries for cultivation in Southern China [9]. Second, the native avocado accessions included in the present studyare mainly from three geographical regions, namely Nanning located in the central and southern region of Guangxi province, Danzhou and Baisha located in the central and western region of Hainan province, and Ruili located in the western region of Yunnan province. These locations are characterized by a warm and humid oceanic climatewith a relatively low altitude in the central and southern region of Guangxi province and the central and western region of Hainan province. Although Ruili is located in the western region of Yunnan province and far from the ocean, it still has a subtropical monsoon climate. The climates of these three regions resemble that of the areas in which the West Indian races originated, and are favorable for the growth of Guatemalan ×

West Indian hybrids. Therefore, Guatemalan × West Indian hybridsmay have graduallybecome the dominant native avocado accessions because of artificial selection or via naturally occurring crosses.

The 100 EST-SSR markers randomly selected for validation in the present study had an amplification rate of 67%, and 30 were determined to be polymorphic. This polymorphism level is generally consistent with that of our previous study [18]. In subsequent analyses of the genetic diversity of these polymorphic EST-SSR markers among 46 avocado accessions, 15 markers produced 4.73 alleles per locus, which was fewer than the 6.13 alleles per locus of Ge [18], the 11.40 alleles per SSR locus of Gross-German and Viruel [17], the 18.8 alleles per SSR locus of Schnell [16], and the 9.75 alleles per SSR locus of Alcaraz and Hormaza [62]. Additionally, a PIC value > 0.5 is generally considered to represent a high polymorphism rate [63]. In this study, 7 of 15 polymorphic EST-SSRs had a PIC value < 0.5. This result may have been because the 46 avocado accessions in this study are genotypically the same (Guatemalan × West Indian hybrids), with relatively low genetic diversity.

In this study, a cluster analysis and a PCoA grouped the native avocado accessions according to where they originated. Additionally, some of the native avocadoaccessions derived from different regions was included in the same sub-cluster. For example, Renong No. 13 from Hainan province clustered with the native accessions from Guangxi province. One factor leading to this promiscuous clustering is the fact that avocado germplasm resources have been exchanged among researchers and breeders since the late 1980s. The CATAS, which is a national scientific research unit, was commissioned to popularize superior avocado accessions among breeders at adjacent state farms or at other national scientific research units. Some superior native accessions from the CATAS may be the male or female parent of other native accessions from various state farms orother national scientific research units, which is consistent with our study results. Furthermore, a cluster analysis grouped two native avocado accessions from Yunnan province with the native avocado accessions from Guangxi province. In contrast, our PCoA indicated that these two native avocado accessions from Yunnan province belong to the same groupas the native avocado accessions from Hainan province. We speculate that the relatively few native avocado accessions from Yunnan province (i.e., two) may have led to these contradictory results based on two statistical analyses. At many avocado plantations in Yunnan province, the local avocado accessions have been replaced by"Hass," which is the most economically valuable avocado cultivar, ultimately making it difficult to collect local avocado accessions. Thus, maximizing the economic benefits of cultivating specific avocado cultivars, while ensuring avocado genetic resources are conserved will need to be addressed.

5. Conclusions

We annotated SMRT sequencing data based on the COG, GO, KEGG, KOG, Pfam, Swiss-Prot, eggNOG, and Nr databases. Among 71,627 transcripts, 45,134, 52,125, and 33,310 were annotated according to GO, eggNOG, and KEGG classifications, respectively. We detected 76,777 SSR loci in 42,096 transcript sequences and used them to develop 149,733 EST-SSR markers. From a randomly selected subset comprising 100 EST-SSR markers, we finally identified 15 polymorphic EST-SSR markers on 71 alleles, which had 2–10 of these markers per locus. A cluster analysis and a PCoA separated the 46 avocado accessions according to their geographical origins. These 15 newly developed EST-SSR markers markers may be useful for future analyses of avocado accessions and may contribute to the improved management of avocado resources for germplasm conservation and breeding programs.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4395/9/9/512/s1. Table S1. Sources of the 46 avocado accessions evaluated in this study. Table S2. KASP primer information and KASP genotyping results. Table S3. Gene annotations of the 71,627 avocado transcripts. Table S4. Characteristics of the GO annotation of avocado transcripts. Table S5. Characteristics of eggNOG classifications of avocado transcripts. Table S6. Characteristics of KEGG pathways ofavocado transcripts. Table S7. Transcription factors identified in the avocado transcripts. Table S8. Frequencies of different repeat motifs in EST-SSRs from avocado. Table S9. Characteristics ofavocado EST-SSR markers in this study. Table S10. Summary of 100 EST-SSR markers used for amplification. Table S11. Summary of 15 EST-SSRs in 46 avocado accessions.Additional file 1. Coding sequences predicted with TransDecoder.

Author Contributions: Y.G., R.Z., and W.M. conceived and designed the experiments; J.W., Y.L. (Yuanzheng Liu), and N.W. performed the experiments; L.T. and D.C. analyzed the data; Y.L. (Yanxia Li) helped complete the experiments; X.Z. contributed materials; and Y.G. wrote the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant number 31701883) and the Natural Science Foundation of Hainan Province of China (grant number 319QN266).

Acknowledgments: We gratefully acknowledge Pingzhen Lin from the Haikou Experimental Station of the Chinese Academy of Tropical Agricultural Sciences for supporting the collection of avocado resources. We thank Yajima for editing the English text of a draft of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Schaffer, B.; Wolstenholme, B.N.; Whiley, A.W. *The Avocado: Botany, Production and Uses*, 2nd ed.; CPI Group (UK) Ltd.: Croydon, UK, 2012.
- 2. Kopp, L.E. A taxonomic revision of the genus *Persea* in the western hemisphere (*Persea-Lauraceae*). *Mem. N. Y. Bot. Gard.* **1966**, *14*, 1–120.
- 3. Williams, L.O. The avocado, a synopsis of the genus *Persea*, subg. *Persea Econ. Bot.* **1977**, *31*, 315–320. [CrossRef]
- 4. Schaffer, B.; Wolstenholme, B.N. *The Avocado: Botany, Production and Uses*; CAB International: Wallingford, UK, 2002.
- 5. Van der Werff, H. A synopsis of Persea (Lauraceae) in Central America. Novon 2002, 12, 575–586. [CrossRef]
- Dreher, M.L.; Davenport, A.J. Hass avocado composition and potential health effects. *Crit. Rev. Food Sci.* 2013, 53, 738–750. [CrossRef]
- 7. Galvão, M.D.S.; Narain, N.; Nigam, N. Influence of different cultivars on oil quality and chemical characteristics of avocado fruit. *Food Sci. Technol.* **2014**, *34*, 539–546. [CrossRef]
- 8. Ge, Y.; Si, X.Y.; Cao, J.Q.; Zhou, Z.X.; Wang, W.L.; Ma, W.H. Morphological characteristics, nutritional quality, and bioactive constituents in fruits of two avocado (*Perseaamericana*) varieties from hainan province, China. *J. Agric. Sci.* **2017**, *9*, 8–17. [CrossRef]
- 9. Ge, Y.; Si, X.Y.; Lin, X.E.; Wang, J.S.; Zang, X.P.; Ma, W.H. Advances in avocado (*Perseaamericana* Mill.). *South China Fruit* **2017**, *46*, 63–70. [CrossRef]
- 10. Zhang, L.; Zhang, D.S.; Liu, K.D. Environmental analysis and countermeasures for industrial development of Hainan avocado. *Chin. J. Agric. Resour. Reg. Plan.* **2015**, *36*, 78–84.
- 11. Fiedler, J.; Bufler, G.; Bangerth, F. Genetic relationships of avocado (*Perseaamericana* Mill.) using RAPD markers. *Euphytica* **1998**, *101*, 249–255. [CrossRef]
- Mhameed, S.; Sharon, D.; Kaufman, D.; Lahav, E.; Hillel, J.; Degani, C.; Lavi, U. Genetic relationships within avocado (*Perseaamericana* Mill.) cultivars and between *Persea* species. *Theor. Appl. Genet.* 1997, 94, 279–286. [CrossRef]
- 13. Furnier, G.R.; Cummings, M.P.; Clegg, M.T. Evolution of the avocados as revealed by DNA restriction site variation. *J. Hered.* **1990**, *81*, 183–188. [CrossRef]
- 14. Davis, J.; Henderson, D.; Kobayashi, M. Genealogical relationships among cultivated avocado as revealed through RFLP analysis. *J. Hered.* **1998**, *89*, 319–323. [CrossRef]
- 15. Ashworth, V.E.T.M.; Clegg, M.T. Microsatellite markers in avocado (*Perseaamericana* Mill.). genealogical relationships among cultivated avocado genotypes. *J. Hered.* **2003**, *94*, 407–415. [CrossRef]
- 16. Schnell, R.J.; Brown, J.S.; Olano, C.T.; Power, E.J.; Krol, C.A.; Kuhn, D.N.; Motamayor, J.C. Evaluation of avocado germplasm using microsatellite markers. *J. Am. Soc. Hortic. Sci.* 2003, *128*, 881–889. [CrossRef]
- 17. Gross-German, E.; Viruel, M.A. Molecular characterization of avocado germplasm with a new set of SSR and EST-SSR markers: Genetic diversity, population structure, and identification of race-specific markers in a group of cultivated genotypes. *Tree Genet. Genomes* **2013**, *9*, 539–555. [CrossRef]
- 18. Ge, Y.; Tan, L.; Wu, B.; Wang, T.; Zhang, T.; Chen, H.; Zou, M.; Ma, F.; Xu, Z.; Zhan, R. Transcriptome sequencing of different avocado ecotypes: De novo transcriptome assembly, annotation, identification and validation of EST-SSR markers. *Forests* **2019**, *10*, 411. [CrossRef]
- 19. Chen, H.; Morrel, P.L.; Ashwoth, V.E.T.M.; De la Cruz, M.; Clegg, M.T. Nucleotide diversity and linkage disequilibrium in wild avocado (*Perseaamericana* Mill.). *J. Hered.* **2008**, *99*, 382–389. [CrossRef]

- 20. Chen, H.; Morrel, P.L.; Ashwoth, V.E.T.M.; De la Cruz, M.; Clegg, M.T. Tracing the geographic origins of mayor avocado cultivars. *J. Hered.* **2009**, *100*, 56–65. [CrossRef]
- Hou, M.Y.; Mu, G.J.; Zhang, Y.J.; Cui, S.L.; Yang, X.L.; Liu, L.F. Evaluation of total flavonoid content and analysis of related EST-SSR in Chinese peanut germplasm. *Crop Breed. Appl. Biotechnol.* 2017, 17, 221–227. [CrossRef]
- 22. Azevedo, A.O.N.; Azevedo, C.D.O.; Santos, P.H.A.D.; Ramos, H.C.C.; Boechat, M.S.B.; Arêdes, F.A.S.; Ramos, S.R.R.; Mirizola, L.A.; Perera, L.; Aragão, W.M.; et al. Selection of legitimate dwarf coconut hybrid seedlings using DNA fingerprinting. *Crop Breed. Appl. Biotechnol.* **2018**, *18*, 409–416. [CrossRef]
- 23. Ahmad, A.; Wang, J.D.; Pan, Y.B.; Sharif, R.; Gao, S.J. Development and use of simple sequence repeats (SSRs) markers for sugarcane breeding and genetic studies. *Agronomy* **2018**, *8*, 260. [CrossRef]
- 24. Ferreira, F.; Scapim, C.A.; Maldonado, C.; Mora, F. SSR-based genetic analysis of sweet corn inbred lines using artificial neural networks. *Crop Breed. Appl. Biotechnol.* **2018**, *18*, 309–313. [CrossRef]
- 25. Ge, Y.; Hu, F.C.; Tan, L.; Wu, B.; Wang, T.; Zhang, T.; Ma, F.N.; Cao, J.Q.; Xu, Z.N.; Zhan, R.L. Molecular diversity in a germplasm collection of avocado accessions from the tropical and subtropical regions of China. *Crop Breed. Appl. Biotechnol.* **2019**, *19*, 153–160. [CrossRef]
- 26. Wu, J.; Cai, C.F.; Cheng, F.Y.; Cui, H.L.; Zhou, H. Characterization and development of EST-SSR markers in tree peony using transcriptome sequences. *Mol. Breed.* **2014**, *34*, 1853–1866. [CrossRef]
- 27. Biswas, M.K.; Nath, U.K.; Howlader, J.; Bagchi, M.; Natarajan, S.; Kayum, M.A.; Kim, H.T.; Park, J.I.; Kang, J.G.; Nou, I.S. Exploration and exploitation of novel SSR markers for candidate transcription factor genes in Lilium species. *Genes* **2018**, *9*, 97. [CrossRef]
- 28. Ma, S.L.Y.; Dong, W.S.; Lyu, T.; Lyu, Y.M. An RNA sequencing transcriptome analysis and development of EST-SSR markers in Chinese hawthorn through Illumina sequencing. *Forests* **2019**, *10*, 82. [CrossRef]
- 29. Li, X.; Li, M.; Hou, L.; Zhang, Z.Y.; Pang, X.M.; Li, Y.Y. De novo transcriptome assembly and population genetic analyses for an endangered Chinese endemic *Acer miaotaiense* (Aceraceae). *Genes* **2018**, *9*, 378. [CrossRef]
- 30. Qi, W.C.; Chen, X.; Fang, P.H.; Shi, S.C.; Li, J.J.; Liu, X.T.; Cao, X.Q.; Zhao, N.; Hao, H.Y.; Li, Y.J.; et al. Genomic and transcriptomic sequencing of *Rosa hybrida* provides microsatellite markers for breeding, flower trait improvement and taxonomy studies. *BMC Plant Biol.* **2018**, *18*, 119. [CrossRef]
- 31. Chen, J.; Tang, X.H.; Ren, C.X.; Wei, B.; Wu, Y.Y.; Wu, Q.H.; Pei, J. Full-length transcriptome sequences and the identification of putative genes for flavonoid biosynthesis in safflower. *BMC Genom.* **2018**, *19*, 548. [CrossRef]
- Tian, J.Y.; Feng, S.J.; Liu, Y.L.; Zhao, L.L.; Tian, L.; Hu, Y.; Yang, T.X.; Wei, A.Z. Single-molecule long-read sequencing of *Zanthoxylumbungeanum* maxim. transcriptome: Identification of aroma-related genes. *Forests* 2018, 9, 765. [CrossRef]
- 33. Roberts, R.J.; Carneiro, M.O.; Schatz, M.C. The advantages of SMRT sequencing. *Genome Biol.* **2013**, *14*, 405–409. [CrossRef]
- Chao, Y.H.; Yuan, J.B.; Li, S.F.; Jia, S.Q.; Han, L.B.; Xu, L.X. Analysis of transcripts and splice isoforms in red clover (*Trifoliumpratense* L.) by single-molecule long-read sequencing. *BMC Plant Biol.* 2018, 18, 300. [CrossRef]
- 35. Hoang, N.V.; Furtado, A.; Mason, P.J.; Marquardt, A.; Kasirajan, L.; Thirugnanasambandam, P.P.; Botha, F.C.; Henry, R.J. A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. *BMC Genom.* **2017**, *18*, 395. [CrossRef]
- 36. Zuo, C.M.; Blow, M.; Sreedasyam, A.; Kuo, R.C.; Ramamoorthy, G.K.; Torres-Jerez, I.; Li, G.F.; Wang, M.; Dilworth, D.; Barry, K.; et al. Revealing the transcriptomic complexity of switchgrass by PacBio long-read sequencing. *Biotechnol. Biofuels* **2018**, *11*, 170. [CrossRef]
- Chao, Y.H.; Yuan, J.B.; Guo, T.; Xu, L.X.; Mu, Z.Y.; Han, L.B. Analysis of transcripts and splice isoforms in Medicago sativa L. by single-molecule long-read sequencing. Plant Mol. Biol. 2019, 99, 219–235. [CrossRef]
- Kim, J.A.; Roy, N.S.; Lee, I.H.; Choi, A.Y.; Choi, B.S.; Yu, Y.S.; Park, N.I.; Park, K.C.; Kim, S.; Yang, H.S.; et al. Genome-wide transcriptome profiling of the medicinal plant *Zanthoxylumplanispinum* using a single-molecule direct RNA sequencing approach. *Genomics* 2019, 111, 973–979. [CrossRef]
- 39. Zhang, B.; Liu, J.X.; Wang, X.S.; Wei, Z.W. Full-length RNA sequencing reveals unique transcriptome composition in bermudagrass. *Plant Physiol. Biochem.* **2018**, *132*, 95–103. [CrossRef]

- 40. Xu, Q.S.; Zhu, J.Y.; Zhao, S.Q.; Hou, Y.; Li, F.D.; Tai, Y.L.; Wan, X.C.; Wei, C.L. Transcriptome profiling using single-molecule direct RNA sequencing approach for in-depth understanding of genes in secondary metabolism pathways of *Camellia sinensis*. *Front. Plant Sci.* **2017**, *8*, 1205. [CrossRef]
- 41. Deng, Y.; Zheng, H.; Yan, Z.C.; Liao, D.Y.; Li, C.L.; Zhou, J.Y.; Liao, H. Full-length transcriptome survey and expression analysis of *Cassia obtusifolia* to discover putative genes related to aurantio-obtusin biosynthesis, seed formation and development, and stress response. *Int. J. Mol. Sci.* **2018**, *19*, 2476. [CrossRef]
- 42. Ge, Y.; Cheng, Z.H.; Si, X.Y.; Ma, W.H.; Tan, L.; Zang, X.P.; Wu, B.; Xu, Z.N.; Wang, N.; Zhou, Z.X.; et al. Transcriptome profiling provides insight into the genes in carotenoid biosynthesis during the mesocarp and seed developmental stages of avocado (*Persea Americana*). *Int. J. Mol. Sci.* **2019**, *20*, 4117. [CrossRef]
- 43. Ge, Y.; Ramchiary, N.; Wang, T.; Liang, C.; Wang, N.; Wang, Z.; Choi, S.R.; Lim, Y.P.; Piao, Z.Y. Development and linkage mapping of unigene-derived microsatellite markers in *Brassica rapa* L. *Breed. Sci.* **2011**, *61*, 160–167. [CrossRef]
- Kanehisa, M.; Araki, M.; Goto, S.; Hattori, M.; Hirakawa, M.; Itoh, M.; Katayama, T.; Kawashima, S.; Okuda, S.; Tokimatsu, T.; et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 2008, *36*, 480–484. [CrossRef]
- 45. Götz, S.; García-Gómez, J.M.; Terol, J.; Williams, T.D.; Nagaraj, S.H.; Nueda, M.J.; Robles, M.; Talon, M.; Dopazo, J.; Conesa, A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **2008**, *36*, 3420–3435. [CrossRef]
- Perez-Rodriguez, P.; Riano-Pachon, D.M.; Correa, L.G.; Rensing, S.A.; Kersten, B.; Mueller-Roeber, B. PlnTFDB: Updated content and new features of the plant transcription factor database. *Nucleic Acids Res.* 2010, *38*, 822–827. [CrossRef]
- 47. Ge, Y.; Zhang, T.; Wu, B.; Tan, L.; Ma, F.N.; Zou, M.H.; Chen, H.H.; Pei, J.L.; Liu, Y.Z.; Chen, Z.H.; et al. Genome-wide assessment of avocado germplasm determined from specific length amplified fragment sequencing and transcriptomes: Population structure, genetic diversity, identification, and application of race-specific markers. *Genes* **2019**, *10*, 215. [CrossRef]
- 48. Krawczak, M.; Nikolaus, S.; von Eberstein, H.; Croucher, P.J.; El Mokhtari, N.E.; Schreiber, S. PopGen: Population based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet.* **2006**, *9*, 55–61. [CrossRef]
- 49. Liu, K.; Muse, S.V. PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics* **2005**, *21*, 2128–2129. [CrossRef]
- 50. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725–2729. [CrossRef]
- 51. Du, M.; Li, N.; Niu, B.; Liu, Y.; You, D.; Jiang, D.; Ruan, C.Q.; Qin, Z.Q.; Song, T.W.; Wang, W.T. De novo transcriptome analysis of *Bagariusyarrelli* (Siluriformes: Sisoridae) and the search for potential SSR markers using RNA-Seq. *PLoS ONE* **2018**, *13*, e0190343. [CrossRef]
- Liu, F.M.; Hong, Z.; Yang, Z.J.; Zhang, N.N.; Liu, X.J.; Xu, D.P. De Novo transcriptomeanalysis of Dalbergiaodorifera T. Chen (Fabaceae) and transferability of SSR markers developed from the transcriptome. Forests 2019, 10, 98. [CrossRef]
- 53. Li, W.; Zhang, C.P.; Jiang, X.Q.; Liu, Q.C.; Liu, Q.H.; Wang, K.L. De Novotranscriptomicanalysis and development of EST–SSRs for *Styrax japonicas*. *Forests* **2018**, *9*, 748. [CrossRef]
- 54. Kilaru, A.; Cao, X.; Dabbs, P.B.; Sung, H.J.; Rahman, M.M.; Thrower, N.; Zynda, G.; Podicheti, R.; Ibarra-Laclette, E.; Herrera-Estrella, L.; et al. Oil biosynthesis in a basal angiosperm: Transcriptome analysis of *Persea Americana* mesocarp. *BMC Plant Biol.* **2015**, *15*, 203. [CrossRef]
- Vergara-Pulgar, C.; Rothkegel, K.; González-Agüero, M.; Pedreschi, R.; Campos-Vargas, R.; Defilippi, B.G.; Meneses, C. De novo assembly of *Perseaamericana* cv. 'Hass' transcriptome during fruit development. *BMC Genom.* 2019, 20, 108. [CrossRef]
- 56. Ibarra-Laclette, E.; Méndez-Bravo, A.; Pérez-Torres, C.A.; Albert, V.A.; Mockaitis, K.; Kilaru, A.; López-Gómez, R.; Cervantes-Luevano, J.I.; Herrera-Estrell, L. Deep sequencing of the Mexican avocado transcriptome, an ancient angiosperm with a high content of fatty acids. *BMC Genom.* 2015, 16, 599. [CrossRef]
- 57. Yang, L.F.; Jin, Y.H.; Huang, W.; Sun, Q.; Liu, F.; Huang, X.Z. Full-length transcriptome sequences of ephemeral plant *Arabidopsis pumila* provides insight into gene expression dynamics during continuous salt stress. *BMC Genom.* **2018**, *19*, 717. [CrossRef]

- 58. Liu, J.; Wang, H.; Chua, N.H. Long noncoding RNA transcriptome of plants. *Plant Biotechnol. J.* **2015**, *13*, 319–328. [CrossRef]
- 59. Yandell, M.; Ence, D. A beginner's guide to eukaryotic genome annotation. *Nat. Rev. Genet.* **2012**, *13*, 329–342. [CrossRef]
- 60. Liu, J.; Jung, C.; Xu, J.; Wang, H.; Deng, S.; Bernad, L.; Arenas-Huertero, C.; Chua, N.H. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell* **2012**, *24*, 4333–4345. [CrossRef]
- 61. Ochogavía, A.; Galla, G.; Seijo, J.G.; González, A.M.; Bellucci, M.; Pupilli, F.; Barcaccia, G.; Albertini, E.; Pessino, S. Structure, target-specifificity and expression of *PN_LNC_N13*, a long non-coding RNA differentially expressed in apomictic and sexual *Paspalumnotatum*. *Plant Mol. Biol.* **2018**, *96*, 53–67. [CrossRef]
- 62. Alcaraz, M.L.; Hormaza, J.I. Molecular characterization and genetic diversity in an avocado collection of cultivars and local Spanish genotypes using SSRs. *Heredity* **2007**, *144*, 244–253. [CrossRef]
- 63. Botstein, D.; White, R.L.; Skolnick, M.; Davis, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **1980**, *32*, 314–331.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).