horticulturae

MDPI

Article

De novo Transcriptome Assembly and Comprehensive Annotation of Two Tree Tomato Cultivars (*Solanum betaceum* Cav.) with Different Fruit Color

Juan Pacheco [1], Santiago Vilanova [1], Rubén Grillo-Risco [2], Francisco Garcia-Garcia [2], Jaime Prohens [1] and Pietro Gramazio [3,*]

[1] Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Camino de Vera 14, 46022 Valencia, Spain; juanenriquepacheco1@gmail.com (J.P.); sanvina@upvnet.upv.es (S.V.); jprohens@btc.upv.es (J.P.)

[2] Principe Felipe Research Center (CIPF), Bioinformatics and Biostatistics Unit, Eduardo Primo Yúfera 3, 46012 Valencia, Spain; rubengrillorisco@gmail.com (R.G.-R.); fgarcia@cipf.es (F.G.-G.)

[3] Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-Universitat Politècnica de València, Camino de Vera 14, 46022 Valencia, Spain

* Correspondence: piegra@upv.es

Abstract: The tree tomato (*Solanum betaceum* Cav.) is an underutilized fruit crop native to the Andean region and phylogenetically related to the tomato and potato. Tree tomato fruits have a high amount of nutrients and bioactive compounds. However, so far there are no studies at the genome or transcriptome level for this species. We performed a de novo assembly and transcriptome annotation for purple-fruited (A21) and an orange-fruited (A23) accessions. A total of 174,252 (A21) and 194,417 (A23) transcripts were assembled with an average length of 851 and 849 bp. A total of 34,636 (A21) and 36,224 (A23) transcripts showed a significant similarity to known proteins. Among the annotated unigenes, 22,096 (A21) and 23,095 (A23) were assigned to the Gene Ontology (GO) term and 14,035 (A21) and 14,540 (A23) were found to have Clusters of Orthologous Group (COG) term classifications. Furthermore, 22,096 (A21) and 23,095 (A23) transcripts were assigned to 155 and 161 (A23) KEGG pathways. The carotenoid biosynthetic process GO terms were significantly enriched in the purple-fruited accession A21. Finally, 68,647 intraspecific single-nucleotide variations (SNVs) and almost 2 million interspecific SNVs were identified. The results of this study provide a wealth of genomic data for the genetic improvement of the tree tomato.

Keywords: de novo transcriptome assembly; emerging crop; functional annotation; molecular markers; RNA-Seq; Solanaceae; *Solanum betaceum*; structural annotation

Citation: Pacheco, J.; Vilanova, S.; Grillo-Risco, R.; Garcia-Garcia, F.; Prohens, J.; Gramazio, P. *De novo* Transcriptome Assembly and Comprehensive Annotation of Two Tree Tomato Cultivars (*Solanum betaceum* Cav.) with Different Fruit Color. *Horticulturae* **2021**, *7*, 431. https://doi.org/10.3390/horticulturae7110431

Academic Editors: Luigi De Bellis and Juan Capel

Received: 17 July 2021
Accepted: 15 October 2021
Published: 22 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 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 (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The tree tomato or tamarillo (*Solanum betaceum* Cav.) is a Solanaceae crop native to the Andean region [1,2]. The tree tomato is phylogenetically related to the potato (*S. tuberosum* L.) and tomato (*S. lycopersicum* L.), forming part of the same clade [3]. The tree tomato plant develops into a small tree, even though some cultivars can grow up to four meters in height, with a fast-growing, shallow root system and simultaneous reproductive and vegetative development [4]. In recent years, the tree tomato has caught the attention of growers and the industry due to its attractive, fleshy, edible fruits, which can be consumed either in salads or as a dessert fruit, or processed for making jams, yogurts, juices, or alcoholic beverages, among others [5]. It has developed from being a neglected crop, with a local interest in subsistence farms [6], into a promising fruit crop, having been introduced in several countries of Oceania, Southeast Asia, Europe and Africa [7]. Aside from South American countries, New Zealand is the largest producer and exporter of the

tree tomato, where the marketable word, "tamarillo", was coined from the Maori term "tama", meaning leadership, combined with the Spanish word, "Amarillo", meaning yellow, or the word, "Tomatillo", meaning small tomato [8].

The interest in the tree tomato also lies in the high amounts of antioxidants, vitamins and carotenoids present in the fruit. The standard servings of tree tomato provide 67–75% of the recommended dietary intake (RDI) of ascorbic acid, 16–23% RDI of α-tocopherol and 9-20% RDI of β-carotene [8]. However, the phytochemical profile of the tree tomato varies among cultivars and environmental conditions [8,9]. The main cultivar groups (orange, orange-pointed, purple, red, and red conical) are differentiated by the fruit colour and shape, with different ranges of morphological and genetic variation among them [6,10]. Despite the great potential of tree tomato as a new major fruit crop, there are no high-throughput genetic or genomic studies conducted for this species. Recent advances in RNA next-generation sequencing (RNA-seq) and bioinformatics resources facilitate transcriptomic studies, even for non-model plant species where reference genomes are not available [11,12]. In fact, RNA-Seq is successfully and increasingly performed to decipher the plant transcriptome of neglected plant species [13]. Nevertheless, RNA sequencing offers many other interesting features such as the evaluation of gene expression, polymorphism discovery, small RNA profiling, phylogenomics, and splice variant discovery, among others [14].

In this study, we performed the transcriptome sequencing and assembly of two tree tomato accessions with different fruit colors (purple and orange) followed by their comprehensive structural and functional annotation. In addition, intraspecific polymorphisms between the two cultivars and interspecific ones with tomato and potato were identified. The transcriptomes and the information generated in the present study will be a useful resource for further genomic and molecular studies and will be a key genomic tool in assisting tree tomato breeding programmes.

2. Materials and Methods

2.1. Plant Material

The study was carried out in 2019 at the the Universitat Politècnica de València (UPV). A purple-fruited tree tomato accession (A21, with purple epicarp and mean fruit weight of 108.8 g) and an orange-fruited tree tomato accession (A23, with orange epicarp a mean fruit weight of 75.1 g) [6] (Figure 1), obtained from the UPV germplasm bank, were used for the present study. Seeds from each accession were germinated following the protocol of Ranil et al., (2015) [15]. Subsequently, the plants were grown in a greenhouse at UPV, Spain (GPS coordinates: latitude, 39° 28′ 55″ N; longitude, 0° 20′11″ W; 7 m above sea level). From each accession, tissues were sampled from several young leaves and flower buds and pools were made for each tissue and accession. Unfortunately, the two accessions did not set fruit under greenhouse conditions at our latitude, and thus fruit tissues were not used for the transcriptome assembly. All samples collected were immediately frozen in liquid nitrogen and stored at −80 °C for later use.

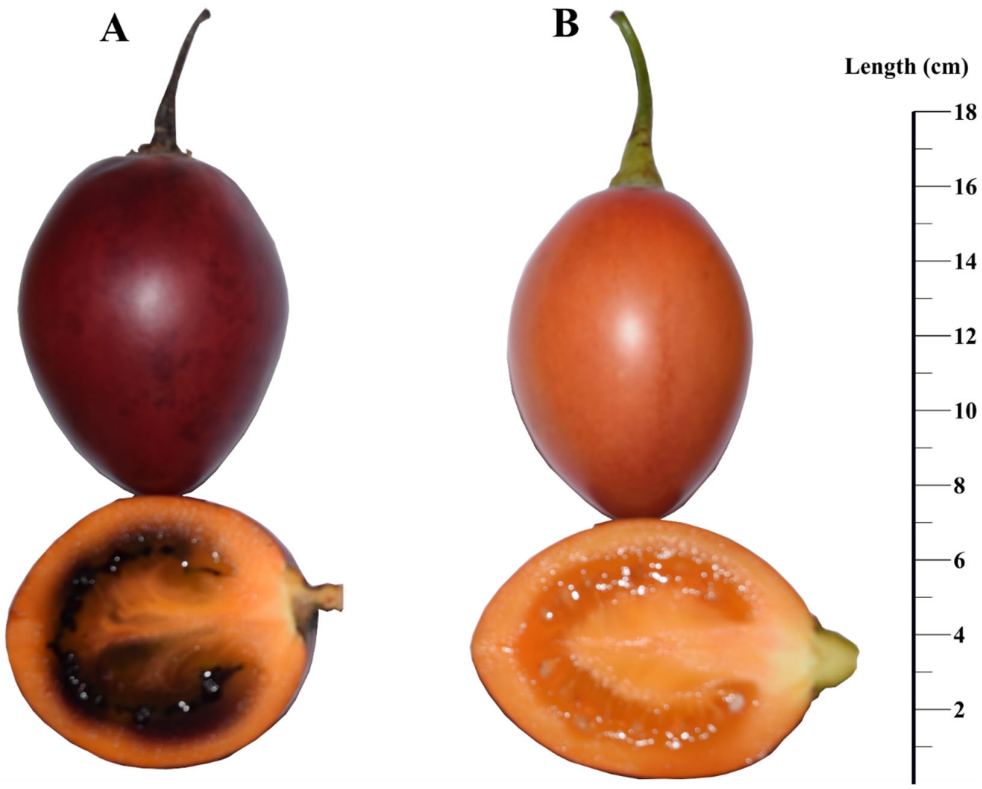

Figure 1. Fruits of tree tomato accessions A21 (**A**) and A23 (**B**).

2.2. RNA Extraction, Library Construction and RNA Sequencing

Total RNA was isolated from each tissue using the Mini spin kit (Macherey-Nage, Dueren, Germany). RNA integrity was determined by 1.0% (*w*/*v*) agarose gel electrophoresis and RNA quantification was performed by Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). From each accession, tissues were sampled from several young leaves and flower buds and pools were made for each tissue and accession. A total of 2 μg of RNA for each pool was sent to Novogene (Cambridge, UK) for library preparation and sequencing. The cDNA paired-end libraries of 150 bp (250~300 bp insert size) libraries were constructed according to Illumina's instructions. The mRNA of each sample was purified from the total RNA by using Sera-mag Magnetic Oligo (dT), then fragmented into short fragments using the fragmentation buffer. Using these fragments as templates, the first strand of cDNA was synthesized. The second strand of cDNA was synthesized using the buffer containing dNTPs, RNase H, and DNA polymerase I. Short fragments (200 ± 20 bp) were connected to the sequencing adapters and suitable fragments were excised from an agarose gel using a gel extraction kit. Then, the library was sequenced using the Illumina Hiseq-2000 sequencer. The raw reads data are available at NCBI Sequence Read Archive (SRA) with accession number SRR15258852 (A21) and SRR15258851 (A23), within the bioproject number PRJNA749599, available at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA749599 http://www.ncbi.nlm.nih.gov.

2.3. DNA Sequence Processing and de novo *Transcriptome Assembly*

The quality of reads was assessed using FastQC v0.11.8 [16]. The adapter sequences, low-quality reads (Phred score <30) and reads with an average length of less than 135 bp were trimmed using Trimmomatic v0.36 [17]. The two accessions were assembled separately using Trinity software v2.10 [18] with a default k-mer size of 25. Identical or near-identical contigs were clustered into a single contig by CD-HIT-EST tool v 4.8.1 [19] with an identity of more than 80%. The quality and completeness of the assemblies were

first evaluated with Bowtie2 v2.3.2 [20] for assessing the number of paired-end reads that were present in the assembled transcripts, then the Ex90N50 transcript contig length (the contig N50 value based on the set of transcripts representing 90% of the expression data) was computed using contig ExN50 statistic.pl script bundled with Trinity. Finally, the completeness of the assemblies was evaluated using BUSCO v4.1.1 [21,22] using a set of eukaryotic genes as a database (https://busco-data.ezlab.org/v5/data/lineages/eukaryota_odb10.2020-09-10.tar.gz) (accessed on 10 August 2020).

2.4. Structural and Functional Annotation

Gene open reading frames (ORFs) were predicted using Transdecoder v5.5.0 (http://transdecoder.sourceforge.net/) using the assembled unitranscripts as input. After the ORFs were extracted from the assembly, redundant contigs with over 90% identity were eliminated using CD-HIT-EST. Functional annotation of the assembled transcripts was conducted using OmicsBox software v 1.4.11 [23] and the Trinotate v3.2.1 pipeline (https://trinotate.github.io/) [24]. Both nucleotide transcripts and protein sequences were blasted against the UniProtKB/Swiss-Prot database (uniprot_sprot.trinotate_v2.0.pep.gz), using NCBI-BLASTx and BLASTp v2.10.1+ (-evalue 1e−3 -max_target_seqs 1 -outfmt 6). Functional domains were identified using the Pfam domain database (Pfam-A.hmm.gz), which used HMMER v3.3.1 [25]. Potential signal peptides were identified using the SignalP v4.1 tool [26]. The OmicBox program (https://www.biobam.com/omicsbox/) was used to further annotate the transcripts using the functional annotation feature of Blast2GO software to predict Gene Ontology (GO) terms, EC Enzyme Code, identify potential KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways, and orthology relationships using the eggNOG v5.0 databases [23,27,28]. GO enrichment analysis was performed using the "topGO" Bioconductor package (http://www.bioconductor.org/packages/release/bioc/html/topGO.html) (accessed on 5 September 2020).

2.5. Single-Nucleotide Variations (SNVs)

For the intraspecific SNVs, clean reads of each accession were mapped separately to the A23 assembled transcriptome which acted as a reference using BWA v0.7.17 [29], while for interspecific SNVs the reads were mapped against the reference genome of tomato (Heinz 1706 version SL4.0) [30] and potato (DM 1-3 516 R44 v6.1) [31]. Subsequently, SAMtools v1.10 [32] was used to convert SAM to BAM format while duplicate reads were removed from respective alignment sequences using Picard-tools v2.23.8 (http://picard.sourceforge.net) (accessed on 20 January 2021). Variants were called by FreeBayes v1.3.4 [33] to identify intra- and interspecific polymorphisms that were filtered using VcfFilter v0.2 (https://github.com/biopet/vcffilter) (accessed on 20 January 2021) based on a minQualScore of 30, minTotalDepth of 40 and a minSampleDepth of 20. Finally, the variant impact effects were predicted using SnpEff v5.0 [34].

3. Results

3.1. Transcriptome Sequencing and Assembly

The RNA sequencing of the two tree tomato accessions yielded 100,919,310 (14.68 Gb) and 113,802,281 (15.84 Gb) raw paired-end reads for A21 and A23, respectively (Table 1). After the initial trimming and stringent quality filtering to remove adapters and low-quality data, 38,411,167 (4.25 Gb) clean paired-end reads were obtained for A21 and

54,474,055 (5.97 Gb) for A23 (Table 1). The two cohorts of clean reads were assembled independently into transcriptomes using Trinity. For the A21 accession, the assembled transcriptome consisted of 174,252 transcripts and spanned 148,352,996 bp, with an average transcript length of 851.37 bp (Table 1). The N50 value was 1494 bp and the GC content of 38.8% (Table 1). On the other hand, the A23 accession was assembled in 194,417 transcripts with a total length of 165,074,290 bp and an average length of 849.07 bp (Table 1). The N50 value for the latter was 1503 bp and the GC content of 38.6% (Table 1). The assembled sequence lengths ranged from the 200 bp cut-off value to a maximum transcript length of 17,046 bp for A21 and 16,865 bp for A23 (Table 1). The majority of the assembled sequences were in the ranges of 200 bp to 500 bp and 501 to 1000 bp.

Table 1. Summary of raw and clean reads statistics before and after processing, de novo assemblies, and BUSCO completeness for tree tomato accessions A21 and A23.

	Accessions	
Statistics	**A21**	**A23**
Total raw reads	100,919,310	113,802,281
Total raw reads data size (Gb)	14.68	15.84
G/C (%)	42.2	42.2
Total clean reads	38,411,167	54,474,055
Total clean reads data size (Gb)	4.25	5.97
Number of transcripts	174,252	194,417
Total nucleotide length	148,352,996	165,074,290
Average transcript length	851.37	849.07
Maximum transcript length	17,046	16,865
N50	1494	1503
G/C (%)	38.8	38.6
Overall alignment rate (%)	99.09	99.21
BUSCO (%)	98.4	98.8

To evaluate the quality of the assemblies, the clean reads were mapped back to the final assembled transcriptome. The overall alignment rates using the alignment software Bowtie2 were 99.09% for A21 and 99.21% for A23 (Table1). BUSCO was employed to evaluate the accuracy and completeness of our transcriptome assembly, gene set, and transcripts. When comparing the set of genes with the genome, we found that the proportion of complete BUSCO was 98.4% for A21 and 98.8% for A23, which indicated that the integrity of the whole transcriptome was very good (Table 1).

3.2. Structural and Functional Annotation

TransDecoder software was used to identify the open reading frames (ORFs) of the unitranscripts assembled and their associated functions, predicting 27,441 ORFs and 34,636 potential proteins for the A21 and 28,336 ORFs and 36,224 potential proteins for A23 (Table 2).

Table 2. Overview of the functional annotation by homology of transcriptomes for tree tomato accessions A21 and A23.

	Accessions	
Statistics	**A21**	**A23**
Predicted ORFs	27,441	28,336
Predicted proteins	34,636	36,224
sprot_Top_BLASTX_hit	57,422	60,772
sprot_Top_BLASTP_hit	24,311	25,054
Pfam	22,954	23,637

SignalP	1623	1745
TmHMM	6899	7216
GO terms	196,800	204,090
EC numbers	15,828	16,668
Kegg	14,035	14,540

Subsequently, the unique transcripts and the putative proteins identified were annotated by performing Blast searches against several databases using the Trinotate pipeline. A total of 57,422 (33%) and 60,772 unigenes (31.3%) displayed a significant homology when Blastx was performed and 24,311 (14.0%) and 25,054 protein sequences (12.3%) when Blastp searches were performed against the UniProtKB/Swiss-Prot database (cut-off E-value of 1e-3) for A21 and A23, respectively (Table 2). Furthermore, 22,954 and 23,637 unique Pfam protein motifs, 1623 and 1,745 protein sequences with signal peptides (SignalP), and 6899 and 7216 transcripts with at least one transmembrane domain (TmHMM) were predicted for A21 and A23, respectively (Table 2). The species distribution showed that most sequences exhibited a high similarity mainly to those of *Arabidopsis thaliana* (L.) Heynh. (17,602 for A21 and 18,117 for A23), *Oryza sativa* L. japonica group (1039 and 1066), *Nicotiana tabacum* L. (613 and 653), *S. lycopersicum* (487 and 486) and *S. tuberosum* (267 and 276) (Figure 2).

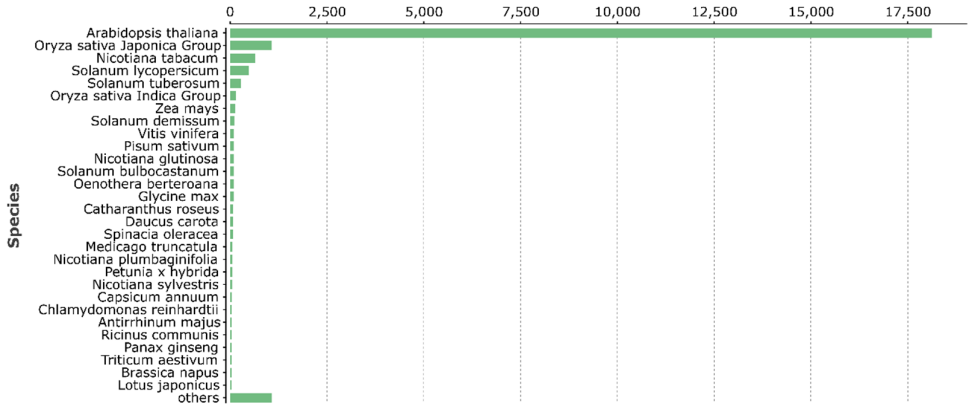

Figure 2. Top species distribution of annotated unigenes for tree tomato accessions A21 and A23.

GO-based functional classification for A21 and A23 transcriptomes assemblies retrieved a total of 196,800 GO terms for A21 and 204,090 for A23 from 22,096 and 23,095 transcripts, respectively (Table 2). The largest number of GO terms (75.2%) was annotated in sequences with a length between 100 and 500 bp (Figure 3).

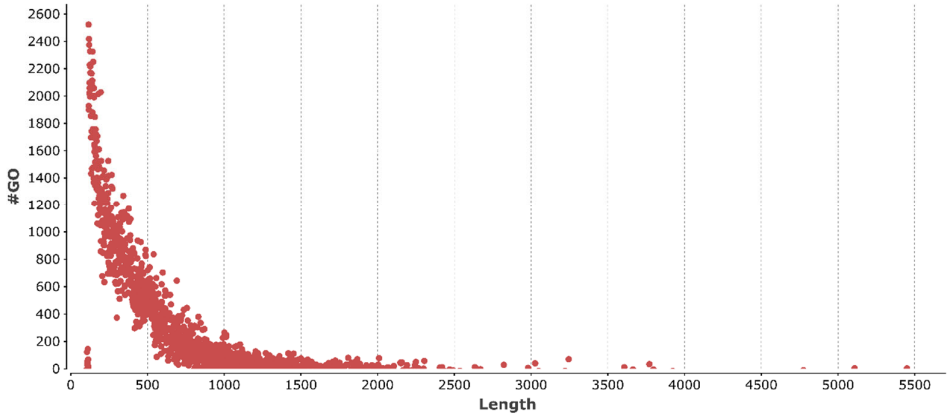

Figure 3. Numbers of GO terms relative to sequence length in the transcriptomes of tree tomato accessions A21 and A23.

Both assemblies had a similar GO distribution for each category; four to nine terms in biological process (BP), three to nine in molecular function (MF) and four to eight in cellular components (CC) category (Figure 4). The GO levels that ranged between 5 and 15, were 88.9% for biological processes, 69.8% for molecular function and 88.2% for cellular components, indicating that the precision of the annotation was accurate (Figure 4) and that a broad diversity of genes was sampled in our transcriptomes.

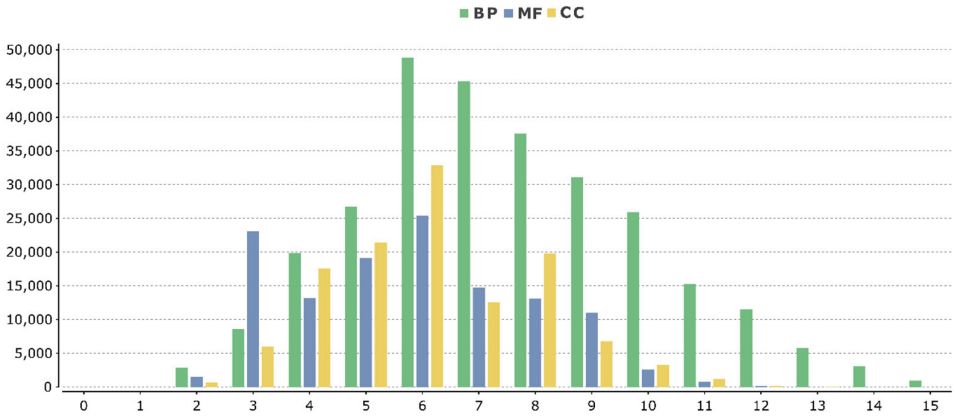

Figure 4. GO level distribution in each category for the annotated tree tomato unigenes. X axis represents the GO level and Y axis the number of annotated unigenes. BP = Biological Process, MF = Molecular Function, CC = Cellular Component.

Among all the GO terms extracted, 137,333 (69.8%) for A21 and 140,193 (68.7%) for A23 were assigned to the biological process category, 35,153 (17.9%) and 38,464 (18.9%) to the molecular function class and 24,314 (12.4%) and 25,233 (12.5%) to the cellular components, respectively (Figure 5).

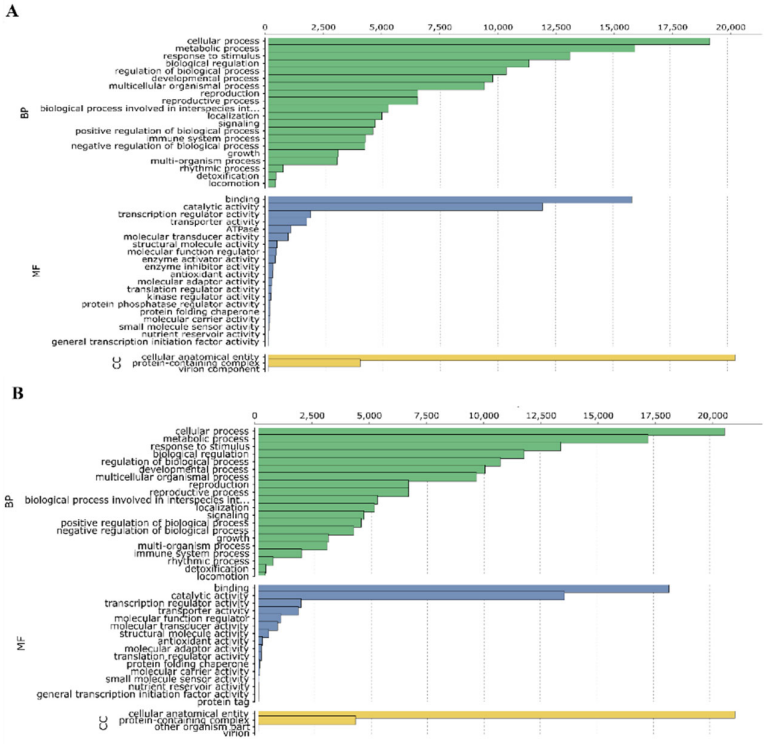

Figure 5. Gene ontology (GO) functional classification of tree tomato A21 (**A**) and A23 (**B**) transcriptomes. Histograms of transcripts annotated to specific GO categories; BP = biological process, MF = molecular functions and CC = cellular components and are represented by green, blue, and yellow bars, respectively.

For the biological process category, the top three subcategories were the cellular process with 19,220 (14.0%) sequences for A21 and 20,660 (14.8%) for A23, the metabolic process with 15,954 (11.6%) and 17,273 (12.3%) and the response to stimulus with 13,134 (9.6%) and 13,400 (9.6%) sequences (Figure 5). For the molecular function category, the vast majority of sequences belonged to two subcategories: binding sequences (15,824; 45% for A21 and 18,204; 47.3% for A23) and catalytic activity (11,940; 34% and 13,566; 35.3%) (Figure 5). Finally, for the cellular component category, most sequences were classified into two sub-categories: cellular anatomical entity (20,315 sequences; 83.6% and 21,109; 83%) and the protein-containing complex (4315; 16.4% and 3998; 17%) (Figure 5). For A21, the GO term enrichment analysis indicated significant GO terms associated with a defense response (GO:0006952), proteolysis (GO:0006508), cellular and lipid metabolic processes (GO:0006629, GO:0044255), carotenoid metabolic processes (GO:0016116), and carotenoid biosynthetic processes (GO:0016117) (Figure 6, Supplementary Table S1). Different to A21, the significantly enriched GO terms of A23 were protein localization (GO:0008104), root morphogenesis (GO:0010015), root development (GO:0048364), post-embryonic plant organ morphogenesis (GO:0090697), the regulation of catabolic process (GO:0009894), the regulation of the cellular catabolic process (GO:0031329) (Figure 6, Supplementary Table S1).

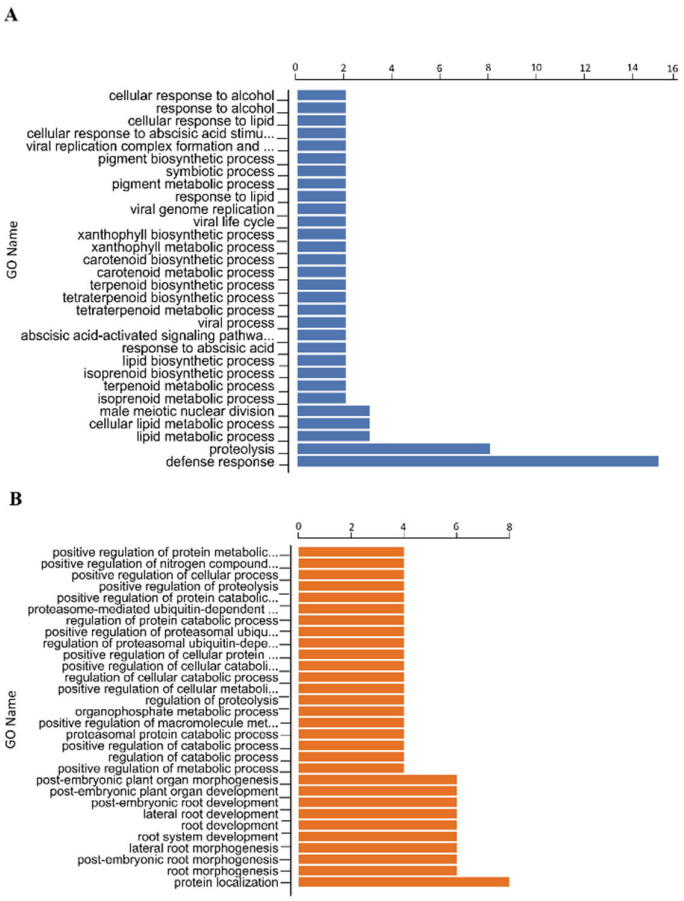

Figure 6. GO enrichment analysis in tree tomato A21 (**A**) and A23 (**B**) transcriptomes.

Several candidate regulatory genes of the carotenoid biosynthetic pathway were identified in the assembled transcriptomes from *S. lycopersicum* and *A. thaliana*. The protein query sequences used for mining the transcriptomic data were the *S. lycopersicum* prolycopene isomerase (*CRTISO*), 9-cis-epoxycarotenoid dioxygenase (*NCED1*), lycopene epsilon cyclase (*Lcy-e*), neoxanthin synthase (*NSY*) and the *A. thaliana* protein ORANGE (*OR*) (Supplementary Table S2). All of them were found to be expressed in both cultivars, where *CRTISO* and *Lcy-e* homologues exhibited a high identity (96%), followed by *NCED1* with 95%, *NSY* with 93%, and finally, *OR*, which showed a higher identity in A23 (74%) than A21 (71%) (Supplementary Table S2).

The enzyme commission (EC) numbers were assigned to 15,828 for the A21 and 16,668 for the A23 unigenes (Table 2). The most represented enzymes were hydrolases (5489 unigenes in A21 and 5562 in A23), transferases (5430 and 5857), oxidoreductases (1766 and 2031) and translocases (1708 and 1680) (Figure 7). Other enzyme classes such as lyases, isomerases, and ligases were represented to a lesser degree.

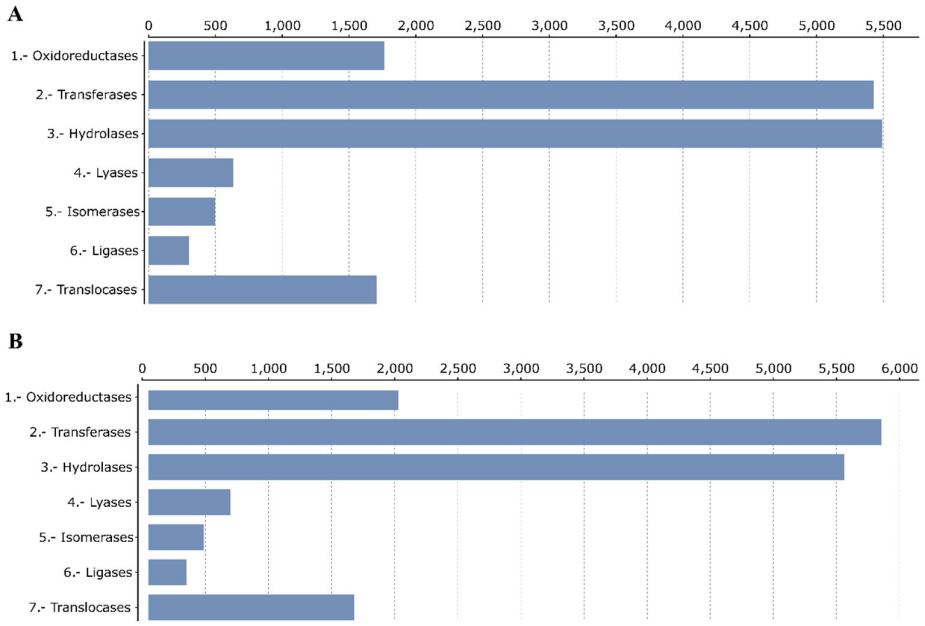

Figure 7. Number of unigenes for each enzyme commission (EC) category for tree tomato A21 (**A**) and A23 (**B**) transcriptomes.

KEGG analysis was performed to identify the potential mechanisms and pathways represented in the identified unigenes. A total of 14,035 unigenes for A21 and 14,540 unigenes for A23 were assigned to the 155 and 161 KEGG pathways, respectively (Table 2). The most represented pathways in terms of the number of homologous transcripts were purine metabolism (map00230, 58 sequences), cysteine and methionine metabolism (map00270, 58 sequences), amino sugar and nucleotide sugar metabolism (map00520, 46 sequences), terpenoid backbone biosynthesis (map00900, 33 sequences), drug metabolism (map00983, 20 sequences), flavonoid biosynthesis (map00941, 17 sequences), and carotenoid biosynthesis (map00906, 17 sequences).

3.3. COG Classification

A Cluster Orthologous Group (COG) is defined as a cluster of three or more homologous sequences that diverge from the same speciation event. Orthologous groups were functionally annotated using the EggNog (evolutionary genealogy of genes: Non-supervised Orthologous Groups) database. In total 97,437 for the A21 and 99,471 for the A23 GO were assigned to 14,530 and 14,928 unique sequences, respectively (Figure 8). The largest group is represented by the cluster for cellular processes and signaling (CPS) (6311; 21.4% and 6443; 21%), followed by metabolism (MB) (6052; 20.5% and 6,417; 20.9%), information storage and processing (ISP) (6040; 20.4% and 6,396; 20.9%) (Figure 8). Within the CPS category, the largest proportion was assigned to signal transduction mechanisms (T) (2359 for A21 and 2392 for A23) and post-translational modification, protein turnover, and chaperones (O) (2051 and 2104). Within the MB category, the largest proportion was assigned to amino acid transport and metabolism (E) (1232 and 1,342), and carbohydrate transport and metabolism (G) (1249 and 1314), and within the ISP category, the majority were assigned to replication, recombination, repair (L) (2043 and 2254), and transcription (K) (1997 and 2043) (Figure 8).

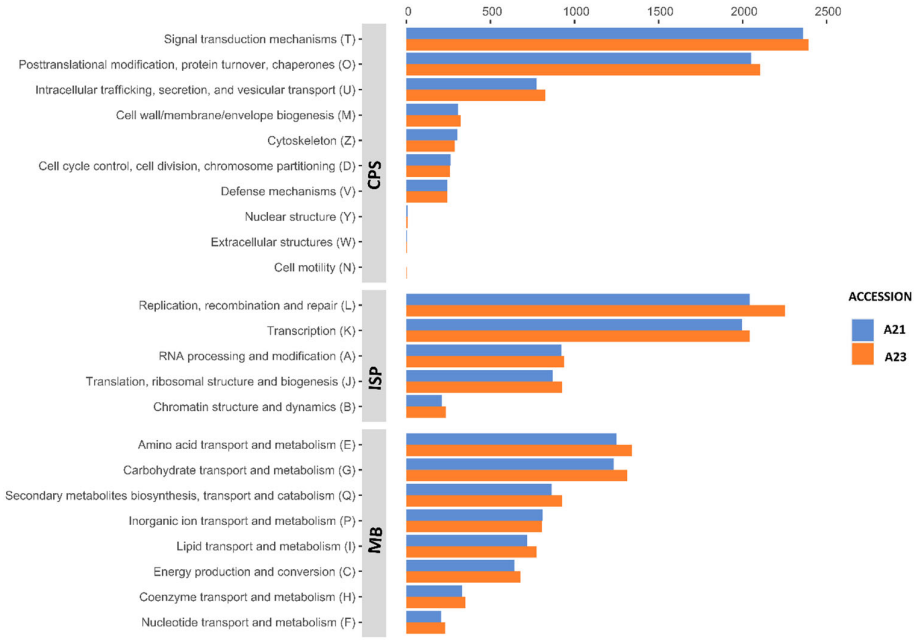

Figure 8. COG categories in the transcriptomes of tree tomato accessions A21 and A23.

3.4. Identification and Characterizacion of SNVs

Intra- and interspecific polymorphisms were identified in both accessions and between the tomato and potato genomes. The number of intraspecific SNVs was significantly higher in the A23 (49,530) than in the A21 accession (19,117) (Table 3). Of these, 14,837 (77.6%) in A21 and 38,183 (77.1%) in A23 were SNPs, 3283 (17.2%) and 8213 (16.6%) were multiple-nucleotide polymorphisms (MNP), 767 (4%) and 2391 (4.8%) were InDels, and 227 (1.2%) and 726 (1.5%) were multiple-nucleotide and an InDel (MIXED) (Table 3). Among the SNPs, the number of transitions (10,687 in A21 and 25,925 in A23) was higher than the number of transversions (5758 and 13,810), with a transition/transversion (Ts/Tv) ratio of 1.86 and 1.88, respectively (Supplementary Table S2). For transition substitution, the most abundant were C/T (17.4% in A21 and 17% A23), followed by G/A (16.8% and 16.6%), A/G (15.7% and 16.1%), and T/C (14.2% and 15.4%) (Supplementary Table S3). In the case of the transversion substitution, the frequency of occurrence of the SNPs was A/T, (6.2% and 5.4%) followed by T/A (5.6% and 5.4%), G/T (4.6% and 4.8%), C/A (4.3% and 4.5%), A/C (4.1% and 4.3%), G/C (3.0% and 2.9%) and C/G (2.7% and 2.6%) (Supplementary Table S3). The average genomic SNPs and InDels variation frequency were 1 in 242 bp in A21 and 1 in 204 bp in A23. In both accessions, the number and proportion of heterozygous variants were higher (74% in A21 and 79% in A23) than the homozygous variants (Supplementary Table S3).

Table 3. Polymorphism statistics for the tree tomato A21 and A23 transcriptomes.

Statistics	SNPs	MNP	INDELs	MIXED	Total SNVs
SNVs intraspecific variations					
A21	14,837	3283	767	227	19,117
A23	38,183	8213	2391	726	49,530
SNVs interspecific variations					
A21 and *S. tuberosum*	619,626	174,982	28,2835	23,115	1,973,023
A23 and *S. tuberosum*	805,997	242,484	42,142	36,352	
A21 and *S. lycopersicum*	624,503	194,857	23,407	20,788	1,809,264
A23 and *S. lycopersicum*	684,775	218,205	27,102	24,627	

The vast majority of the variants (12,095; 50.4% in A21 to 38,632; 61.3% in A23), classified according to SNPeff, were predicted as a "modifier", i.e., the variants were located in intergenic or intronic regions, or in an exon from a non-coding transcript, which indicates that there is no evidence of their impact or that their predictions are difficult to assess (Supplementary Table S4). The second most abundant impact effects predicted were "low" (6507; 27.1% and 11,732; 18.7%), which were mostly harmless variants or unlikely to change protein behaviour (Supplementary Table S4). The third ones were those predicted as having "moderate" impact effects (5139; 21.4% and 11,637; 18.6%), i.e., nondisruptive variants, such as codon insertion/deletion or codon substitution, which might change protein effectiveness (Supplementary Table S4). Finally, the less abundant impact class corresponded to the "high" variation effects (262; 1.1% and 808; 1.3%), which were considered to have a disruptive impact on the protein-like truncation or loss of function caused by exon deletion/deletion (Supplementary Table S4). The top variant categories were in the exon regions (48% for A21 and 37% for A23), intergenic regions (21% and 30%), 3′ UTR variant (17%), 5′ UTR variant (14% and 16%) and the synonymous variant (25% and 17%) (Supplementary Table S5). Regarding the effects on protein function, on average, 58% of the variants in A21 and 52% in A23 were predicted to produce a silent effect (41% and 47%), a missense impact (1%) and a nonsense protein product (Supplementary Table S5).

Regarding the interspecific SNVs, the highest number of SNVs were identified with potato (1,973,023) and a little less with tomato (1,809,264). Of those, 1,425,623 (72.3%) with potato and 1,309,278 (72.0%) with tomato were SNPs; 417,416 (21.2%) and 413,062 (22.7%) were MNP; 70,427 (3.6%) and 50,509 (2.8%) were InDels; and 59,507 (3.0%) and 45,415 (2.5%) were MIXED (Table 3). The accession A23 exhibited a higher number of interspecific variants than A21 (Table 3). In contrast to the intraspecific SNVs, the proportion of homozygous variants was higher (over 95%) than the heterozygous ones. Considerable differences were observed in the average number of polymorphisms among the chromosomes, with differences of over two-fold between chromosome 1 (259,267 in potato and 237,098 in tomato) and chromosome 12 (127,595 and 113,202) in both accessions (Table 4).

Table 4. Chromosome distribution of tree tomato variants with potato (*S. tuberosum*) and tomato (*S. lycopersicum*).

	Species	
Chromosome	***S. tuberosum***	***S. lycopersicum***
1	259,267	237,496
2	200,827	90,558
3	205,561	92,045
4	176,113	78,104
5	133,442	59,508
6	164,694	72,583

7	152,942	67,774
8	140,397	62,569
9	148,610	64,651
10	130,071	58,183
11	133,138	59,040
12	127,595	54,634

The impact of 3,186,724 SNVs (58.7%) in potato and 2,095,805 (59.9%) in tomato was classified as a "modifier"; 1,192,042 (21.9%) and 728,209 (22.3%) were classified as "low"; 1,029,629 (19%) and 611,869 (17.5%) were classified as "moderate"; and the impact of the remaining 23,696 (0.4%) and 12,268 (0.4%) SNPs were classified as "high" (Supplementary Table S4). The majority of variant categories were in the exons (39% to 43%), downstream gene variant intergenic regions (25% and 28%), upstream gene variant (15% and 19%), 3′ UTR variant (5% and 8%), intron variant (2% and 6%), intergenic region (2% and 3%), and the 5′ UTR variant (2% and 3%).

We further analyzed the sequences of the candidate genes that played an important role in the carotenoids biosynthesis, identifying a total of 1548 SNVs in the two cultivars assessed when compared to the tomato reference genome (Table 5). Of them, 478 SNPs were found in the coding region of the prolycopene isomerase (*CRTISO*) gene, 372 in 9-cis-epoxycarotenoid dioxygenase (*NCED1*), 194 in lycopene epsilon cyclase (*Lcy-e*), 164 in neoxanthin synthase (*NSY*), and 340 in protein ORANGE (OR) (Table 3). The impact of the majority of variants (42.2%) was classified as a "modifier", 31.6% as "low", 17.9% as "moderate" and 1.9% as "high". Regarding the effects on protein function, on average, 51% of the variants were synonymous mutations, while the remaining variants were missense mutations.

Table 5. Single-nucleotide variations (SNVs) identified in candidate genes of the carotenoids biosynthesis.

Statistics	A21	A23	Total SNVs
CRTISO	233	245	478
NCED1	174	198	372
Lcy-e	91	103	194
NSY	79	85	164
OR	139	201	340

4. Discussion

Although tree tomato is one of the most promising fruit crops in the Mediterranean and temperate regions [4], its genomic landscape has not yet been explored yet. Other unexploited crops similar to the tree tomato, such as the cape gooseberry (*Physalis peruviana* L.) and *amaranth (Amaranthus cruentus* L.), have greatly benefited from genomic studies, which have fostered the dissection of multiple agronomic traits and breeding programs [35–37]. In this study, we conducted the de novo transcriptome assembly of two tree tomato cultivars to provide useful genomic data for the improvement of this unexploited but emerging crop. Through RNA sequencing, a total of 174,252 (for A21) and 194,417 (for A23) transcripts were assembled from 38 and 54 million filtered reads and with an average length of 851 and 849 bp, respectively. The number of transcripts of these accessions was slightly higher than those obtained in previous transcriptome studies in other related Solanaceae species such as tomato, potato or pepino (*Solanum muricatum Aiton*) [14,38,39], but it was similar to others obtained in plant species of the same family,

such as *S. commersonii* Dunal and *S. aculeatissimum* [40,41], suggesting the high quality and reliability of our assemblies. Furthermore, the assembly and annotation completeness was quantitatively confirmed by the high percentage values (>98%) of the BUSCO assessment, values that were comparable or even higher than those of other recent *Solanum* transcriptomes, which exhibited values of 97% for *S. tuberosum* and 93% for *S. chilense* [42,43].

The functional annotation of the assembled unigenes is essential for understanding the role of the represented genes [44]. Even though the number of protein-coding genes is unknown in tree tomato, the prediction of the potential ORFs (27,441 in A21 and 28,336 in A23) and proteins (34,636 in A21 and 36,224 in A23) was in agreement with those observed for protein-coding genes in other *Solanum* species, such as tomato (35,535), potato (39,290), eggplant (*S. melongena* L.) (30,630 and 34,231) [45–47]. Similarly, signal peptides, transmembrane and Pfam domains were assigned to around 5%, 20%, and 65% of the identified proteins, respectively. These percentages were higher than those obtained in other plant species of the Solanaceae family such as *S. trilobatum* and *S. sisymbriifolium* [48,49]. The GO annotation revealed that unigenes could be categorized into three major functional categories: biological processes (68%), molecular functions (18%) and cellular components (12%). The top two subcategories were the cellular and metabolic processes in the biological processes, binding, and catalytic activity of the molecular function; and the cellular anatomical entity and protein-containing complex in the cellular component, which suggests that many novel genes involved in metabolic activities could play important roles during the growth and development stages of the plant.

The KEGG annotation allows for the functional analysis and interpretation of transcriptomic data and exhibits how the assembled transcripts are integrated into metabolic pathways and biological systems [50]. A total of 155 pathways in A21 and 161 in A23 involving 14,035 and 14,540 unigenes were annotated, including pathways of great interest that could be used to improve the quality of breeding programs for the breeding of tree tomato such as purine metabolism, drug metabolism, terpenoid backbone biosynthesis, and the biosynthesis of flavonoid and carotenoids. The increased accumulation of flavonoids and carotenoids in fruit crops improves their commercial and health values [51]. Among the biological features, the most renowned property of flavonoids and carotenoids is their antioxidant effects, which are often much higher than those of vitamin E and vitamin C [52,53]. Our transcriptional results confirmed the presence of known genes and enzymes in pathways related to the synthesis of flavonoids and carotenoids. These results are in agreement with previous studies that reported the tree tomato as an abundant source of carotenoids, anthocyanins, flavonoids, and phenolic compounds and has higher antioxidant activity than other antioxidant-rich fruits such as kiwifruit or grape [7]. The carotenoid concentration in tree tomato may be under the control of several genes that are associated with the structure and function of the genes in the carotenoid pathway. In the accession A21, our data showed that the carotenoid biosynthetic process GO terms were significantly enriched. This is in agreement with the previous results of [54,55] who reported that the purple cultivar had higher levels of carotenoids compared to the yellow or orange cultivars. Our results suggested that the flavonoid and carotenoid biosynthesis pathway-related genes were well conserved in the tree tomato when compared with the tomato [56]. The sequence variants in these genes among tree tomato varieties could be used as functional markers for marker-assisted breeding to obtain new varieties of tree tomato with improved nutritional values.

We obtained a total of 68,647 SNVs between both accessions, suggesting a high level of polymorphisms for tree tomato. The SNVs reported here were higher than the cohorts identified in other transcriptomic studies of Solanaceae, such as the 17,000 SNVs found in tomato [39]; however, in the case of potato a similar number of SNPs, 69,011, were reported [57]. The A21 accession exhibited a higher number of SNVs and, interestingly, the vast majority of the detected variants were heterozygous. The latter might be due to the fact that, even though some tree tomato cultivars are considered self-compatible and

autogamous, the flowers are frequently visited by pollinator insects, which can lead to cross-pollination [4]. The annotated SNP effects located in the exon and intergenic regions and a transition to transversion ratio of 1.86 agree with previous findings in tomato and eggplant [39,58]. On the other hand, the number of interspecific variants detected with potato was significantly higher than those of tomato, confirming that tree tomato is phylogenetically closer to the latter [59]. The data also indicated that the differences in variant number between the tree tomato and its closely related species were evident, particularly for chromosomes 1 and 12, which were highly related to the physical length between them. Regarding the SVNs found in the candidate genes involved in the carotenoids biosynthesis pathway, our results showed that the *CRTISO* gene exhibited the highest number of SNPs, which could be due to mutations in its coding sequence. The carotenoid analysis in tomato ripe fruits showed that a mutation in CRTISO leads to a prolycopene accumulation instead of all-trans-lycopene compounds, resulting in a fruit color change from red to orange [60]. In addition, most of the SNVs within the genes involved in carotenoid metabolism resulted in synonymous substitutions. These results were consistent with previous studies in tomato [61] where protein expression and protein folding may be influenced by synonymous SNPs as they are involved in regulating microRNA-mediated genes [62,63]. Hence, the synonymous SNPs identified in the tree tomato cultivars in this work may have potential functional significance in carotenoid biosynthesis.

The identification of the intraspecific and interspecific variants will foster several applications including genetic mapping, genotype identification, marker-assisted selection, breeding, comparative genomics, and understanding the genetic control of adaptive traits in the tree tomato [64].

5. Conclusions

In this work, we assembled high-quality transcriptome sequences of two tree tomato cultivars, a fruit crop closely related to tomato and potato, with great potential in subtropical regions. The comprehensive annotation provided extensive and detailed information that facilitates the dissection of traits of agronomic interest, such as the content in bioactive compounds or the response to stresses, among others. In addition, this is the first study in tree tomato where a high number of polymorphisms have been identified, both intraspecifically and with closely related species that could be used in genetic diversity analysis, qualitative and quantitative trait mapping, and breeding programs in tree tomato. This information constitutes a valuable resource for tree tomato breeding programs and genetic diversity studies and will help in the enhancement of tree tomato and its successful introduction in other regions and countries.

Supplementary Materials: The following are available online at www.mdpi.com/article/10.3390/horticulturae7110431/s1, Table S1: GO enrichment analysis in A21 and A23, Table S2: Identification of regulatory genes of the carotenoid biosynthetic pathway in A21 and A23 from *Solanum lycopersicum* and *Arabidopsis thaliana,* Table S3: Number of transitions/transversions, variant rate, and homozygous and heterozygous variants, Table S4: Number of predicted effects by impact, Table S5: Percentage of effects by region and functional class.

Author Contributions: Conceptualization, S.V., J.P. (Juan Pacheco) and P.G.; methodology, J.P. (Juan Pacheco), S.V, R.G.-R., F.G.-G., and P.G.; formal analysis, J.P. (Juan Pacheco) , S.V., R.G.-R., F.G.-G., and P.G.; investigation, J.P. (Juan Pacheco); resources, J.P. (Jaime Prohens); data curation, J.P. (Juan Pacheco), S.V., R.G.-R., F.G.-G., J.P. (Jaime Prohens) and P.G.; writing—original draft preparation, J.P. (Juan Pacheco); writing—review and editing, J.P. (Juan Pacheco), S.V., R.G.-R., F.G.-G., J.P. (Jaime Prohens) and P.G. All authors have read and agreed to the published version of the manuscript.

Funding: Pietro Gramazio is grateful to the Spanish Ministerio de Ciencia e Innovación for a Juan de la Cierva-Formación post-doctoral grant FJC2019-038921-I funded by MCIN/AEI/ 10.13039/501100011033)

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw reads data are available at NCBI Sequence Read Archive (SRA) with accession number SRR15258852 (A21) and SRR15258851 (A23), within the bioproject number PRJNA749599, available at http://www.ncbi.nlm.nih.gov. VCF files are available upon request from the corresponding author.

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

References

1. Duarte, O.; Paull, R.E. Solanaceae. In *Exotic Fruits and Nuts of the New World*, 1st ed.; Duarte, O., Paull, R.E., Eds.; CABI: Wallingford, UK, 2015; pp. 136–144.
2. Orqueda, M.E.; Zampini, I.C.; Torres, S.; Alberto, M.R.; Ramos, L.L.P.; Schmeda-Hirschmann, G.; Isla, M.I. Chemical and functional characterization of skin, pulp and seed powder from the Argentine native fruit mistol (*Ziziphus mistol*). Effects of phenolic fractions on key enzymes involved in metabolic syndrome and oxidative stress. *J. Funct. Foods* **2017**, *37*, 531–540. https://doi.org/10.1016/j.jff.2017.08.020.
3. Särkinen, T.; Bohs, L.; Olmstead, R.G.; Knapp, S. A phylogenetic framework for evolutionary study of the nightshades (*Solanaceae*): A dated 1000-tip tree. *BMC Evol. Biol.* **2013**, *13*, 214–214. https://doi.org/10.1186/1471-2148-13-214.
4. Ramírez, F.; Kallarackal, J. Tree tomato (*Solanum betaceum* Cav.) reproductive physiology: A review. *Sci. Hortic.* **2019**, *248*, 206–215. https://doi.org/10.1016/j.scienta.2019.01.019.
5. Chen, X.; Quek, S.Y.; Fedrizzi, B.; Kilmartin, P.A. Characterization of free and glycosidically bound volatile compounds from tamarillo (*Solanum betaceum* Cav.) with considerations on hydrolysis strategies and incubation time. *LWT* **2020**, *124*, 109178. https://doi.org/10.1016/j.lwt.2020.109178.
6. Acosta-Quezada, P.G.; Martínez-Laborde, J.B.; Prohens, J. Variation among tree tomato (*Solanum betaceum* Cav.) accessions from different cultivar groups: Implications for conservation of genetic resources and breeding. *Genet. Resour. Crop Evol.* **2011**, *58*, 943–960. https://doi.org/10.1007/s10722-010-9634-9.
7. Diep, T.T.; Rush, E.C.; Yoo, M.J.Y. Tamarillo (*Solanum betaceum* Cav.): A review of physicochemical and bioactive properties and potential applications. *Food Rev. Int.* **2020**, *6*, 1–25. https://doi.org/10.1080/87559129.2020.1804931.
8. Diep, T.T.; Pook, C.; Rush, E.C.; Yoo, M.J.Y. Quantification of carotenoids, α-tocopherol, and ascorbic acid in amber, mulligan, and laird's large cultivars of New Zealand tamarillos (*Solanum betaceum* Cav.). *Foods* **2020**, *9*, 769. https://doi.org/10.3390/foods9060769.
9. Mertz, C.; Brat, P.; Caris-Veyrat, C.; Gunata, Z. Characterization and thermal lability of carotenoids and vitamin C of tamarillo fruit (*Solanum betaceum* Cav.). *Food Chem.* **2010**, *119*, 653–659. https://doi.org/10.1016/j.foodchem.2009.07.009.
10. Acosta-Quezada, P.G.; Vilanova, S.; Martínez-Laborde, J.B.; Prohens, J. Genetic diversity and relationships in accessions from different cultivar groups and origins in the tree tomato (*Solanum betaceum* Cav.). *Euphytica* **2012**, *187*, 87–97. https://doi.org/10.1007/s10681-012-0736-7.
11. Huang, X., Chen, X.-G.; Armbruster, P.A. Comparative performance of transcriptome assembly methods for non-model organisms. *BMC Genom.* **2016**, *17*, 523. https://doi.org/10.1186/s12864-016-2923-8.
12. Ward, J.A.; Ponnala, L.; Weber, C.A. Strategies for transcriptome analysis in nonmodel plants. *Am. J. Bot.* **2012**, *99*, 267–276. https://doi.org/10.3732/ajb.1100334.
13. Xia, Z.; Xu, H.; Zhai, J.; Li, D.; Luo, H.; He, C.; Huang, X. RNA-Seq analysis and de novo transcriptome assembly of Hevea brasiliensis. *Plant Mol. Biol.* **2011**, *77*, 299–308. https://doi.org/10.1007/s11103-011-9811-z.
14. Herraiz, F.J.; Blanca, J.; Ziarsolo, P.; Gramazio, P.; Plazas, M.; Anderson, G.J.; Prohens, J.; Vilanova, S. The first de novo transcriptome of pepino (*Solanum muricatum*): Assembly, comprehensive analysis and comparison with the closely related species S. caripense, potato and tomato. *BMC Genom.* **2016**, *17*, 321. https://doi.org/10.1186/s12864-016-2656-8.
15. Ranil, R.H.G.; Niran, H.M.L.; Plazas, M.; Fonseka, R.M.; Fonseka, H.H.; Vilanova, S.; Andújar, I.; Gramazio, P.; Fita, A.; Prohens, J. Improving seed germination of the eggplant rootstock Solanum torvum by testing multiple factors using an orthogonal array design. *Sci. Hortic.* **2015**, *193*, 174–181. https://doi.org/10.1016/j.scienta.2015.07.030.
16. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (accessed on 18 July 2020).
17. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
18. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.; A Thompson, D.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. https://doi.org/10.1038/nbt.1883.
19. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012**, *28*, 3150–3152. https://doi.org/10.1093/bioinformatics/bts565.
20. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

21. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.
22. Waterhouse, R.M.; Seppey, M.; Simao, F.A.; Manni, M.; Ioannidis, P.; Klioutchnikov, G.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **2018**, *35*, 543–548. https://doi.org/10.1093/molbev/msx319.
23. Götz, S.; Garcia-Gomez, J.M.; Terol, J.; Williams, T.; Nagaraj, S.H.; Nueda, M.J.; Robles, M.; Talón, M.; Dopazo, J.; Conesa, A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **2008**, *36*, 3420–3435. https://doi.org/10.1093/nar/gkn176.
24. Bryant, D.M.; Johnson, K.; DiTommaso, T.; Tickle, T.; Couger, M.B.; Payzin-Dogru, D.; Lee, T.J.; Leigh, N.; Kuo, T.-H.; Davis, F.G.; et al. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Rep.* **2017**, *18*, 762–776. https://doi.org/10.1016/j.celrep.2016.12.063.
25. Wheeler, T.J.; Eddy, S.R. Nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **2013**, *29*, 2487–2489. https://doi.org/10.1093/bioinformatics/btt403.
26. Petersen, T.N.; Brunak, S.; Von Heijne, G.; Nielsen, H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* **2011**, *8*, 785–786. https://doi.org/10.1038/nmeth.1701.
27. Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernández-Plaza, A.; Forslund, S.K.; Cook, H.V.; Mende, D.R.; Letunic, I.; Rattei, T.; Jensen, L.J.; et al. EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **2019**, 47, D309–D314. https://doi.org/10.1093/nar/gky1085.
28. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. https://doi.org/10.1093/nar/28.1.27.
29. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760.
30. Hosmani, P.S.; Flores-Gonzales, M.; van de Geest, H.; Maumus, F.; Bakker, L.V.; Schijlen, E.; van Haarst, J.; Cordewener, J.; Sanchez-Perez, G.; Peters, S.; et al. An Improved de Novo Assembly and Annotation of the Tomato Reference Genome using Single-Molecule Sequencing, Hi-C Proximity Ligation and Optical Maps. BioRxiv under review. Available online: https://www.biorxiv.org/content/10.1101/767764v1 (accessed on day month year).
31. Pham, G.M.; Hamilton, J.P.; Wood, J.C.; Burke, J.T.; Zhao, H.; Vaillancourt, B.; Ou, S.; Jiang, J.; Buell, C.R. Construction of a chromosome-scale long-read reference genome assembly for potato. *GigaScience* **2020**, 9, giaa100. https://doi.org/10.1093/gigascience/giaa100.
32. Li, H.; Handsaker, R.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and samtools. *Bioinformatics* **2009**, *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
33. Garrison, E.; Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv* **2012**, arXiv:1207.3907. Available online: http://arXiv.org/abs/1207.3907.
34. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **2012**, *6*, 80–92. https://doi.org/10.4161/fly.19695.
35. Barabaschi, D.; Tondelli, A.; Desiderio, F.; Volante, A.; Vaccino, P.; Valè, G.; Cattivelli, L. Next generation breeding. *Plant Sci.* **2015**, *242*, 3–13. https://doi.org/10.1016/j.plantsci.2015.07.010.
36. Enciso-Rodriguez, F.; Osorio-Guarín, J.A.; Garzón-Martínez, G.A.; Delgadillo-Duran, P.; Barrero, L.S. Optimization of the genotyping-by-sequencing SNP calling for diversity analysis in cape gooseberry (*Physalis peruviana* L.) and related taxa. *PLoS ONE* **2020**, 15, e0238383. https://doi.org/10.1371/journal.pone.0238383.
37. Ma, X.; Vaistij, F.E.; Li, Y.; van Rensburg, W.S.J.; Harvey, S.; Bairu, M.W.; Venter, S.L.; Mavengahama, S.; Ning, Z.; Graham, I.A.; et al. A chromosome-level Amaranthus cruentus genome assembly highlights gene family evolution and biosynthetic gene clusters that may underpin the nutritional value of this traditional crop. *Plant J.* **2021**, 107, 613–628. https://doi.org/10.1111/tpj.15298.
38. Moon, K.-B.; Ahn, D.-J.; Park, J.-S.; Jung, W.Y.; Cho, H.S.; Kim, H.-R.; Jeon, J.-H.; Park, Y.-I.; Kim, H.-S.Transcriptome profiling and characterization of drought-tolerant potato plant (*Solanum tuberosum* L.). *Mol. Cells* **2018**, *41*, 979–992. https://doi.org/10.14348/molcells.2018.0312.
39. Scarano, D.; Rao, R.; Corrado, G. In Silico identification and annotation of noncoding RNAs by RNA-seq and de novo assembly of the transcriptome of tomato fruits. *PLoS ONE* **2017**, 12, e0171504. https://doi.org/10.1371/journal.pone.0171504.
40. Yang, X.; Liu, F.; Zhang, Y.; Wang, L.; Cheng, Y.-F. Cold-responsive miRNAs and their target genes in the wild eggplant species Solanum aculeatissimum. *BMC Genom.* **2017**, *18*, 1–13. https://doi.org/10.1186/s12864-017-4341-y.
41. Zuluaga, A.P.; Solé, M.; Lu, H.; Góngora-Castillo, E.; Vaillancourt, B.; Coll, N.; Buell, C.R.; Valls, M. Transcriptome responses to Ralstonia solanacearum infection in the roots of the wild potato Solanum commersonii. *BMC Genom.* **2015**, *16*, 1–16. https://doi.org/10.1186/s12864-015-1460-1.
42. Petek, M.; Zagorščak, M.; Ramšak, .; Sanders, S.; Tomaž, .; Tseng, E.; Zouine, M.; Coll, A.; Gruden, K. Cultivar-specific transcriptome and pan-transcriptome reconstruction of tetraploid potato. *Sci. Data* **2020**, *7*, 1–15. https://doi.org/10.1038/s41597-020-00581-4.

43. Stam, R.; Nosenko, T.; Hörger, A.C.; Stephan, W.; Seidel, M.; Kuhn, J.M.M.; Haberer, G.; Tellier, A. The de novo reference genome and transcriptome assemblies of the wild tomato species Solanum chilense highlights birth and death of NLR genes between tomato species. *G3 Genes Genomes Genet.* **2019**, *9*, 3933–3941.
44. Hu, H.; Scheben, A.; Edwards, D. Advances in integrating genomics and bioinformatics in the plant breeding pipeline. *Agriculture* **2018**, *8*, 75. https://doi.org/10.3390/agriculture8060075.
45. Aversano, R.; Contaldi, F.; Ercolano, M.R.; Grosso, V.; Iorizzo, M.; Tatino, F.; Xumerle, L.; Molin, A.D.; Avanzato, C.; Ferrarini, A.; et al. The Solanum commersonii genome sequence provides insights into adaptation to stress conditions and genome evolution of wild potato relatives. *Plant Cell* **2015**, *27*, 954–968. https://doi.org/10.1105/tpc.114.135954.
46. Gramazio, P.; Blanca, J.; Ziarsolo, P.; Herraiz, F.J.; Plazas, M.; Prohens, J.; Vilanova, S. Transcriptome analysis and molecular marker discovery in Solanum incanum and S. aethiopicum, two close relatives of the common eggplant (*Solanum melongena*) with interest for breeding. *BMC Genom.* **2016**, *17*, 1–17. https://doi.org/10.1186/s12864-016-2631-4.
47. Wang, X.; Gao, L.; Jiao, C.; Stravoravdis, S.; Hosmani, P.S.; Saha, S.; Zhang, J.; Mainiero, S.; Strickler, S.R.; Catala, C.; et al. Genome of *Solanum pimpinellifolium* provides insights into structural variants during tomato breeding. *Nat. Commun.* **2020**, *11*, 1–11. https://doi.org/10.1038/s41467-020-19682-0.
48. Lateef, A.; Prabhudas, S.K.; Natarajan, P. RNA sequencing and de novo assembly of Solanum trilobatum leaf transcriptome to identify putative transcripts for major metabolic pathways. *Sci. Rep.* **2018**, *8*, 1–13. https://doi.org/10.1038/s41598-018-33693-4.
49. Wu, L.; Du, G.; Bao, R.; Li, Z.; Gong, Y.; Liu, F. De novo assembly and discovery of genes involved in the response of *Solanum sisymbriifolium* to *Verticillium dahlia*. *Physiol. Mol. Biol. Plants* **2019**, *25*, 1009–1027. https://doi.org/10.1007/s12298-019-00666-4.
50. 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, D480–D484. https://doi.org/10.1093/nar/gkm882.
51. Karasawa, M.M.G.; Mohan, C. Fruits as prospective reserves of bioactive compounds: A review. *Nat. Prod. Bioprospecting* **2018**, *8*, 335–346. https://doi.org/10.1007/s13659-018-0186-6.
52. Britton, G. Carotenoid research: History and new perspectives for chemistry in biological systems. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* **2020**, *1865*, 158699. https://doi.org/10.1016/j.bbalip.2020.158699.
53. Nabavi, S.M.; Šamec, D.; Tomczyk, M.; Milella, L.; Russo, D.; Habtemariam, S.; Suntar, I.; Rastrelli, L.; Daglia, M.; Xiao, J.; et al. Flavonoid biosynthetic pathways in plants: Versatile targets for metabolic engineering. *Biotechnol. Adv.* **2020**, *38*, 107316. https://doi.org/10.1016/j.biotechadv.2018.11.005.
54. Acosta-Quezada, P.G.; Raigón, M.D.; Riofrío-Cuenca, T.; García-Martínez, M.D.; Plazas, M.; Burneo, J.I.; Figueroa, J.G.; Vilanova, S.; Prohens, J. Diversity for chemical composition in a collection of different varietal types of tree tomato (*Solanum betaceum* Cav.), an Andean exotic fruit. *Food Chem.* **2015**, *169*, 327–335. https://doi.org/10.1016/j.foodchem.2014.07.152.
55. Vasco, C.; Avila, J.; Ruales, J.; Svanberg, U.; Kamal-Eldin, A. Physical and chemical characteristics of golden-yellow and purple-red varieties of tamarillo fruit (*Solanum betaceum* Cav.). *Int. J. Food Sci. Nutr.* **2009**, *60*, 278–288. https://doi.org/10.1080/09637480903099618.
56. Ye, J.; Hu, T.; Yang, C.; Li, H.; Yang, M.; Ijaz, R.; Ye, Z.; Zhang, Y. Transcriptome profiling of tomato fruit development reveals transcription factors associated with ascorbic acid, carotenoid and flavonoid biosynthesis. *PLoS ONE* **2015**, 10, e0130885. https://doi.org/10.1371/journal.pone.0130885.
57. Hamilton, J.P.; Hansey, C.N.; Whitty, B.R.; Stoffel, K.; Massa, A.N.; Van Deynze, A.; De Jong, W.S.; Douches, D.S.; Buell, C.R. Single nucleotide polymorphism discovery in elite north American potato germplasm. *BMC Genom.* **2011**, *12*, 302. https://doi.org/10.1186/1471-2164-12-302.
58. Gramazio, P.; Yan, H.; Hasing, T.; Vilanova, S.; Prohens, J.; Bombarely, A. Whole-genome resequencing of seven eggplant (*Solanum melongena*) and one wild relative (*S. incanum*) accessions provides new insights and breeding tools for eggplant enhancement. *Front. Plant Sci.* **2019**, *10*, 1220. https://doi.org/10.3389/fpls.2019.01220.
59. Olmstead, R.G.; Bohs, L.; Migid, H.A.; Santiago-Valentin, E.; Garcia, V.F.; Collier, S.M. A molecular phylogeny of the Solanaceae. *Taxon* **2008**, *57*, 1159–1181. https://doi.org/10.1002/tax.574010.
60. Isaacson, T.; Ronen, G.; Zamir, D.; Hirschberg, J. Cloning of *tangerine* from tomato reveals a carotenoid isomerase essential for the production of β-carotene and xanthophylls in plants. *Plant Cell* ***2002****, 14, 333–342*. https://doi.org/10.1105/tpc.010303.
61. Livingstone, K.; Anderson, S. Patterns of variation in the evolution of carotenoid biosynthetic pathway enzymes of higher plants. *J. Hered.* ***2009****, 100, 754–761*. https://doi.org/10.1093/jhered/esp026.
62. Hunt, R.C.; Simhadri, V.L.; Iandoli, M.; Sauna, Z.E.; Kimchi-Sarfaty, C. Exposing synonymous mutations. *Trends Genet.* **2014**, *30*, 308–321.
63. Gu, W.; Wang, X.; Zhai, C.; Xie, X.; Zhou, T. Selection on synonymous sites for increased accessibility around miRNA binding sites in plants. *Mol. Biol. Evol.* ***2020****, 29, 3037–3044*. https://doi.org/10.1093/molbev/mss109
64. He, J.; Zhao, X.; Laroche, A.; Lu, Z.-X.; Liu, H.K.; Li, Z. Genotyping-by-sequencing (GBS), An ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Front. Plant Sci.* **2014**, *5*, 484. https://doi.org/10.3389/fpls.2014.00484.