Vegetable Research

ARTICLE

Open Access

https://doi.org/10.48130/vegres-0024-0031

Vegetable Research **2024**, 4: e032

De novo assembly of the sweet pitaya (*Stenocereus thurberi*) fruit peel transcriptome and identification of cuticle biosynthesis genes

Heriberto García-Coronado[1], Miguel-Ángel Hernández-Oñate[2], Julio-César Tafolla-Arellano[3], Alexel-Jesús Burgara-Estrella[4] and Martín-Ernesto Tiznado-Hernández[1*]

[1] *Coordinación de Tecnología de Alimentos de Origen Vegetal, Centro de Investigación en Alimentación y Desarrollo A.C., Carretera Gustavo Enrique Astiazarán Rosas 46, Hermosillo 83304, Sonora, México*

[2] *CONAHCYT-Coordinación de Tecnología de Alimentos de Origen Vegetal, Centro de Investigación en Alimentación y Desarrollo A.C., Carretera Gustavo Enrique Astiazarán Rosas 46, Hermosillo 83304, Sonora, México*

[3] *Departamento de Ciencias Básicas, Universidad Autónoma Agraria Antonio Narro, Calzada Antonio Narro 1923, Saltillo 25315, Coahuila, México*

[4] *Departamento de Investigación en Física, Universidad de Sonora, Blvd. Luis Encinas y Rosales S/N, Hermosillo 83000, Sonora, México*

* Corresponding author, E-mail: tiznado@ciad.mx

Abstract

Stenocereus thurberi is a cactus endemic to the Sonoran desert (Mexico), which produces a fruit named sweet pitaya. One trait that allows the cactus to survive in desert ecosystems is its cuticle, which limits water loss in dry conditions. Nevertheless, the mechanism of cuticle biosynthesis has yet to be described for cactus fruits. Also, transcripts from *S. thurberi* published in the databases are scarce. This study reports the *de novo* assembly of the sweet pitaya peel transcriptome. The assembly includes 174,449 transcripts with an N50 value of 2,110 bp. Out of the total transcripts, 43,391 were classified as long non-coding RNA. Functional categorization analysis suggests that mechanisms of response to stress and cuticle biosynthesis are carried out in fruit pitaya peel. The transcripts coding for a cytochrome p450 77A (*StCYP77A*), Gly-Asp-Ser-Leu motif lipase/esterase 1 (*StGDSL1*), and ATP binding cassette G 11 (*StABCG11*), which carried out the synthesis, polymerization, and transport of cuticle components, respectively, were identified. Expression analysis during fruit development suggests an active cuticle biosynthesis at the early stages and the ripe stages, carried out by *StCYP77A*, *StGDSL1*, and *StABCG11*. The dataset generated here will help to improve the elucidation of the molecular mechanism of cuticle biosynthesis in *S. thurberi* and other cactus fruits.

Citation: García-Coronado H, Hernández-Oñate MÁ, Tafolla-Arellano JC, Burgara-Estrella AJ, Tiznado-Hernández ME. 2024. *De novo* assembly of the sweet pitaya (*Stenocereus thurberi*) fruit peel transcriptome and identification of cuticle biosynthesis genes. *Vegetable Research* 4: e032 https://doi.org/10.48130/vegres-0024-0031

Introduction

Columnar cacti are plants of the *Cactaceae* family distributed across arid and semi-arid regions of America, with ecological, economic, and cultural value[1]. One trait that makes it possible for the columnar cactus to survive in the desert ecosystem is its thick epidermis covered by a hydrophobic cuticle, which limits water loss in dry conditions[1]. The cuticle is the external layer that covers the non-woody aerial organs of land plants. The careful control of cuticle biosynthesis could produce drought stress tolerance in relevant crop plants[2]. In fleshy fruits, the cuticle maintains adequate water content during fruit development on the plant and reduces water loss in fruit during postharvest[3]. Efforts to elucidate the molecular pathway of cuticle biosynthesis have been carried out for fleshy fruits such as tomato (*Solanum lycopersicum*)[4], apple (*Malus domestica*)[5], sweet cherry (*Prunus avium*)[6], mango (*Mangifera indica*)[7], and pear (*Pyrus* 'Yuluxiang')[8].

The plant cuticle is formed by the two main layers cutin and cuticular waxes[3]. Cutin is composed mainly of oxygenated long-chain (LC) fatty acids (FA), which are synthesized by cytochrome p450 (CYP) enzymes. CYP family 86 subfamily A (CYP86A) enzymes carry out the terminal (ω) oxidation of LC-FA[9]. Then, CYP77A carries out the mid-chain oxidation to synthesize the main cutin monomers. In *Arabidopsis*, AtCYP77A4 and AtCYP77A6 carry out the synthesis of mid-chain epoxy and mid-chain dihydroxy LC-FA, respectively[10,11]. *AtCYP77A6* is required for the cutin biosynthesis and the correct formation of floral surfaces[10]. The expression of *CYP77A19* (KF410855) and *CYP77A20* (KF410856) from potato (*Solanum tuberosum*) restored the petal cuticular impermeability in *Arabidopsis* null mutant *cyp77a6-1*, tentatively by the synthesis of cutin monomers[12]. In eggplant (*Solanum torvum*), the overexpression of *StoCYP77A2* leads to resistance to *Verticillium dahlia* infection in tobacco plants[13]. Although the function of CYP77A2 in cutin biosynthesis has not yet been tested, gene expression analysis suggests that *CaCYP77A2* (A0A1U8GYB0) could play a role in cutin biosynthesis during pepper fruit development[14].

It has been hypothesized that the export of cuticle precursors is carried out by ATP binding cassette subfamily G (ABCG) transporters. *ABCG11/WBC11*, *ABCG12*, and *ABCG13* are required for the load of cuticle lipids in *Arabidopsis*[15–17], but *ABCG13* function appears to be specific to the flower epidermis[18]. The overexpression of *TsABCG11* (JQ389853) from *Thellungiella salsugineum* increases cuticle amounts and promotes tolerance to different abiotic stresses in *Arabidopsis*[19].

Once exported, the cutin monomers are polymerized on the surface of epidermal cells. *CD1* code for a Gly-Asp-Ser-Leu motif

lipase/esterase (GDSL) from tomato required for the cutin formation through 2-mono(10,16-dihydroxyhexadecanoyl)glycerol esterification[20]. GDSL1 from tomato carries out the ester bond cross-links of cutin monomers located at the cuticle layers and is required for cuticle deposition in tomato fruits[21]. It has been shown that the transcription factor *MIXTA-like* reduces water loss in tomato fruits through the positive regulation of the expression of *CYP77A2*, *ABCG11*, and *GDSL1*[22]. Despite the relevant role of cuticles in maintaining cactus homeostasis in desert environments[1], the molecular mechanism of cuticle biosynthesis has yet to be described for cactus fruits.

Stenocereus thurberi is a columnar cactus endemic from the Sonoran desert (Mexico), which produces an ovoid-globose fleshy fruit named sweet pitaya[23]. In its mature state, the pulp of sweet pitaya contains around 86% water with a high content of antioxidants and natural pigments such as betalains and phenolic compounds, which have nutraceutical and industrial relevance[23]. Due to the arid environment in which pitaya fruit grows, studying its molecular mechanism of cuticle biosynthesis can generate new insights into understanding species' adaptation mechanisms to arid environments. Nevertheless, sequences of transcripts from *S. thurberi* in public databases are scarce.

RNA-sequencing technology (RNA-seq) allows the massive generation of almost all the transcripts from non-model plants, even if no complete assembled genome is available[24]. Recent advances in bioinformatic tools has improved our capacity to identify long non-coding RNA (lncRNA), which have been showed to play regulatory roles in relevant biological processes, such as the regulation of drought stress tolerance in plants[25], fruit development, and ripening[26–29].

In this study, RNA-seq data were obtained for the *de novo* assembly and characterization of the *S. thurberi* fruit peel transcriptome. As a first approach, three transcripts, *StCYP77A*, *StABCG11*, and *StGDSL1*, tentatively involved in cuticle biosynthesis, were identified and quantified during sweet pitaya fruit development. Due to no gene expression analysis having been carried out yet for *S. thurberi*, stably expressed constitutive genes were identified for the first time.

Materials and methods

Plant materials and gene library sequencing

Sweet pitaya fruits (*S. thurberi*) without physical damage were hand harvested from plants in a native conditions field located at Carbó, Sonora, México. They were collocated in a cooler containing dry ice and transported immediately to the laboratory. The superficial part of the peels (~1 mm deep) was removed carefully from the fruits using a scalpel. Peel samples from three fruits were pooled according to their tentative stage of development defined by their visual characteristics, frozen in liquid nitrogen, and pulverized to create a single biological replicate. Four samples belonging to four different plants were analyzed. All fruits harvested were close to the ripening stage. Samples named M1 and M2 were turning from green to ripe [~35–40 Days After Flowering (DAF)], whereas samples M3 and M4 were turning from ripe to overripe (~40–45 DAF).

Total RNA was isolated from the peels through the Hot Borate method[30]. The concentration and purity of RNA were determined in a spectrophotometer Nanodrop 2000 (Thermo Fisher) by measuring the 260/280 and 260/230 absorbance ratios. RNA integrity was evaluated through electrophoresis in agarose gel 1% and a Bioanalyzer 2100 (Agilent). Pure RNA was sequenced in the paired-end mode in an Illumina NextSeq 500 platform at the University of Arizona Genetics Core Facility. Four RNA-seq libraries, each of them from each sample, were obtained, which include a total of 288,199,704 short reads with a length of 150 base pairs (bp). The resulting sequence data can be accessed at the Sequence Read Archive (SRA) repository of the NCBI through the BioProject ID PRJNA1030439. Libraries are named corresponding to the names of samples M1, M2, M3, and M4.

de novo transcriptome assembly and quality analysis

FastQC software (www.bioinformatics.babraham.ac.uk/projects/fastqc) was used for short reads quality analysis. Short reads with poor quality were trimmed or eliminated by Trimmomatic (www.usadellab.org/cms/?page=trimmomatic) with a trailing and leading of 25, a sliding window of 4:25, and a minimum read length of 80 bp. A total of 243,194,888 reads with at least a 25 quality score on the Phred scale were used to carry out the *de novo* assembly by Trinity (https://github.com/trinityrnaseq/trinityrnaseq/wiki) with the following parameters: minimal k-mer coverage of 1, normalization of 50, and minimal transcript length of 200 bp.

Removal of contaminating sequences and ribosomal RNA (rRNA) was carried out through SeqClean. To remove redundancy, transcripts with equal or more than 90% of identity were merged through CD-hit (www.bioinformatics.org/cd-hit/). Alignment and quantification in terms of transcripts per million (TPM) were carried out through Bowtie (https://bowtie-bio.sourceforge.net/index.shtml) and RSEM (https://github.com/deweylab/RSEM), respectively. Transcripts showing a low expression (TPM < 0.01) were discarded. Assembly quality was evaluated by calculating the parameters N50 value, mean transcript length, TransRate score, and completeness. The statistics of the transcriptome were determined by TrinityStats and TransRate (https://hibberdlab.com/transrate/). The transcriptome completeness was determined through a BLASTn alignment (E value < 1×10^{-3}) by BUSCO (https://busco.ezlab.org/) against the database of conserved orthologous genes from *Embryophyte*.

Functional annotation of protein-coding transcripts

To predict the proteins tentatively coded in the *S. thurberi* transcriptome, the best homology match of the assembled transcripts was found by alignment to the Swiss-Prot, RefSeq, nr-NCBI, PlantTFDB, iTAK, TAIR, and ITAG databases using the BLAST algorithm with an E value threshold of 1×10^{-10} for the nr-NCBI database and of 1×10^{-5} for the others[31–34]. An additional alignment was carried out to the protein databases of commercial fruits *Persea americana, Prunus persica, Fragaria vesca, Citrus cinensis,* and *Vitis vinifera* to proteins of the cactus *Opuntia streptacantha*, and the transcriptomes of the cactus *Hylocereus polyrhizus, Pachycereus pringlei,* and *Selenicereus undatus*. The list of all databases and the database websites of commercial fruits and cactus are provided in Supplementary Tables S1 & S2. The open reading frame (ORF) of the transcripts and the protein sequences tentative coded from the sweet pitaya transcriptome was predicted by TransDecoder (https://github.com/TransDecoder/TransDecoder/wiki), considering a minimal ORF length of 75 amino acids (aa). The search for protein domains was carried out by the InterPro database

(www.ebi.ac.uk/interpro). Functional categorization was carried out by Blast2GO based on GO terms and KEGG metabolic pathways[35].

Identification of long non-coding transcripts

LncRNA were identified based on the methods reported in previous studies[25,29,36]. Transcripts without homology to any protein from Swiss-Prot, RefSeq, nr-NCBI, PlantTFDB, iTAK, TAIR, ITAG, *P. americana, P. persica, F. vesca, C. cinensis, V. vinifera*, and *O. streptacantha* databases, without a predicted ORF longer than 75 aa, and without protein domains in the InterPro database were selected to identify tentative lncRNA.

Transcripts coding for signal peptide or transmembrane helices were identified by SignalP (https://services.healthtech.dtu.dk/services/SignalP-6.0/) and TMHMM (https://services.healthtech.dtu.dk/services/TMHMM-2.0/), respectively, and discarded. Further, transcripts corresponding to other non-coding RNAs (ribosomal RNA and transfer RNA) were identified through Infernal by using the Rfam database[37] and discarded. The remaining transcripts were analyzed by CPC[38], and CPC2[39] to calculate their coding potential. Transcripts with a coding potential score lower than −1 for CPC and a coding probability lower than 0.1 for CPC2 were considered lncRNA. To characterize the identified lncRNA, the length and abundance of coding and lncRNA were calculated. Bowtie and RSEM were used to align and quantify raw counts, respectively. The edgeR package[40] was used to normalize raw count data in terms of counts per million (CPM) for both coding and lncRNA.

Identification of tentative reference genes

To obtain the transcript's expression, the aligning of short reads and quantifying of transcripts were carried out through Bowtie and RSEM software, respectively. A differential expression analysis was carried out between the four libraries by edgeR package in R Studio. Only the transcripts with a count equal to or higher than 0.5 in at least one sample were retained for the analysis. Transcripts with $\log_2$ Fold Change ($\log_2$FC) between +1 and −1 and with a False Discovery Rate (FDR) lower than 0.05 were taken as not differentially expressed (NDE).

For the identification of the tentative reference genes two strategies were carried out as described below: i) The NDE transcripts were aligned by BLASTn (E value $< 1 \times 10^{-5}$) to 43 constitutive genes previously reported in fruits from the cactus *H. polyrhizus, S. monacanthus*, and *S. undatus*[41–43] to identify possible homologous constitutive genes in *S. thurberi.* Then, the homologous transcripts with the minimal coefficient of variation (CV) were selected; ii) For all the NDE transcripts, the percentile 95 value of the mean CPM and the percentile 5 value of the CV were used as filters to recover the most stably expressed transcripts, based on previous studies[44]. Finally, transcripts to be tested by quantitative reverse transcription polymerase chain reaction (qRT-PCR) were selected based on their homology and tentative biological function.

Evaluation of reliable reference genes

The fruit harvesting was carried out as described above. Sweet pitaya fruit takes about 43 d to ripen, therefore, open flowers were tagged, and fruits with 10, 20, 30, 35, and 40 DAF were collected to cover the pitaya fruit development process (Supplementary Fig. S1). The superficial part of the peels (~1 mm deep) was removed carefully from the fruits using a scalpel. Peel samples from three fruits were pooled according to their stage of development defined by their DAF, frozen in liquid nitrogen, and pulverized to create a single biological replicate. One biological replicate consisted of peels from three fruits belonging to the same plant. Two to three biological replicates were evaluated for each developmental stage. Two technical replicates were analyzed for each biological replicate. RNA extraction, quantification, RNA purity, and RNA integrity analysis were carried out as described above.

cDNA was synthesized from 100 ng of RNA by QuantiTect Reverse Transcription Kit (QIAGEN). Primers were designed using the PrimerQuest™, UNAFold, and OligoAnalyzer™ tools from Integrated DNA Technologies (www.idtdna.com/pages) and following the method proposed by Thornton & Basu[45]. Transcripts quantification was carried out in a QIAquant 96 5 plex according to the PowerUp™ SYBR™ Green Master Mix protocol (Applied Biosystems), with a first denaturation step for 2 min at 95 °C, followed by 40 cycles of denaturation step at 95 °C for 15 s, annealing and extension steps for 30 s at 60 °C.

The Cycle threshold (Ct) values obtained from the qRT-PCR were analyzed through the algorithms BestKeeper, geNorm, NormFinder, and the delta Ct method[46]. RefFinder (www.ciidirsinaloa.com.mx/RefFinder-master/) was used to integrate the stability results and to find the most stable expressed transcripts in sweet pitaya fruit peel during development. The pairwise variation value (Vn/Vn + 1) was calculated through the geNorm algorithm in R Studio software[47].

Quantification of cuticle biosynthesis-related transcripts

An alignment of 17 reported cuticle biosynthesis genes from model plants were carried out by BLASTx against the predicted proteins from sweet pitaya. Two additional alignments of 17 charaterized cuticle biosynthesis proteins from model plants against the transcripts and predicted proteins of sweet pitaya were carried out by tBLASTn and BLASTp, respectively. An E value threshold of 1×10^{-5} was used, and the unique best hits were recovered for all three alignments. The sequences of the 17 characterized cuticle biosynthesis genes and proteins from model plants are showed in Supplementary Table S3. The specific parameters and the unique best hits for all the alignments carried out are shown in Supplementary Tables S4–S8.

Cuticle biosynthesis-related transcripts tentatively coding for a cytochrome p450 family 77 subfamily A (CYP77A), a Gly-Asp-Ser-Leu motif lipase/esterase 1 (GDSL1), and an ATP binding cassette transporter subfamily G member 11 (ABCG11) were identified by best bi-directional hit according to the functional annotation described above. Protein-conserved domains, signal peptide, and transmembrane helix were predicted through InterProScan, SignalP 6.0, and TMHMM, respectively. Alignment of the protein sequences to tentative orthologous of other plant species was carried out by the MUSCLE algorithm[48]. A neighbor-joining (NJ) phylogenetic tree with a bootstrap of 1,000 replications was constructed by MEGA11[49].

Fruit sampling, primer design, RNA extraction, cDNA synthesis, and transcript quantification were performed as described above. Relative expression was calculated according to the $2^{-\Delta\Delta CT}$ method[50]. The sample corresponding to 10 DAF was used as the calibrator. The transcripts *StEF1a*, *StTUA*, *StUBQ3*, and *StEF1a* + *StTUA* were used as normalizer genes.

Statistical analysis

Normality was assessed according to the Shapiro-Wilk test. Significant differences in the expression of the cuticle biosynthesis-related transcripts between fruit developmental stages

were determined by one-way ANOVA based on a completely randomized sampling design and a Tukey honestly significant difference (HSD) test, considering a p-value < 0.05 as significant. Statistical analysis was carried out through the stats package in R Studio.

Results

Short reads and assembly quality

RNA was extracted from the peels of ripe sweet pitaya fruits (*S. thurberi*) from plants located in the Sonoran Desert, Mexico. Four cDNA libraries were sequenced in an Illumina NextSeq 500 platform at the University of Arizona Genetics Core Facility. A total of 288,199,704 reads with 150 base pairs (bp) in length were sequenced in paired-end mode. After trimming, 243,194,888 (84.38%) cleaned short reads with at least 29 mean quality scores per read in the Phred scale and between 80 to 150 bp in length were obtained to carry out the assembly. After removing contaminating sequences, redundancy, and low-expressed transcripts, the assembly included 174,449 transcripts with an N50 value of 2,110 bp. Table 1 shows the different quality variables of the *S. thurberi* fruit peel transcriptome.

Table 1. Quality metrics of the *Stenocereus thurberi* fruit peel transcriptome.

Metric	Data
Total transcripts	174,449
N50	2,110
Smallest transcript length (bp)	200
Largest transcript length (bp)	19,114
Mean transcript length (bp)	1,198.69
GC (%)	41.33
Total assembled bases	209,110,524
TransRate score	0.05
BUSCO score (%)	C: 85.38 (S:48.22, D:37.16), F: 10.69, M: 3.93.

Values were calculated through the TrinityStats function of Trinity and TransRate software. Completeness analysis was carried out through BUSCO by aligning the transcriptome to the *Embryophyte* database through BLAST with an E value threshold of 1×10^{-3}. Complete (C), single (S), duplicated (D), fragmented (F), missing (M).

BUSCO score showed that 85.4% are completed transcripts, although out of these, 37.2% were found to be duplicated. The resulting sequence data can be accessed at the SRA repository of the NCBI through the BioProject ID PRJNA1030439.

Homology searches

A summary of the homology search in the main public protein database for the *S. thurberi* transcriptome is shown in Supplementary Table S1. From these databases, the higher homologous transcripts were found in RefSeq with 93,993 (53.87 %). Based on the E value distribution, for 41,685 (44%) and 68,853 (49%) of the hits, it was found a strong homology (E value lower than 1×10^{-50}) to proteins in the Swiss-Prot and RefSeq databases, respectively (Supplementary Fig. S2a & b). On the other hand, 56,539 (52.34%) and 99,599 (71.11%) of the matches showed a percentage of identity higher than 60% in the Swiss-Prot and RefSeq databases, respectively (Supplementary Fig. S2c & d).

Figure 1 shows the homology between transcripts from *S. thurberi* and proteins of commercial fruits, as well as proteins and transcripts of cacti. Transcripts from *S. thurberi* homologous to proteins from fruits of commercial interest avocado (*P. americana*), peach (*P. persica*), strawberry (*F. vesca*), orange (*C. sinensis*), and grapefruit (*V. vinifera*) ranged from 77,285 (44.30%) to 85,421 (48.96%), with 70,802 transcripts homologous to all the five fruit protein databases (Fig. 1a).

Transcripts homologous to transcripts or proteins from the cactus dragon fruit (*H. polyrhizus*), prickly pear cactus (*O. streptacantha*), Mexican giant cardon (*P. pringlei*), and pitahaya (*S. undatus*) ranged from 76,238 (43.70%) to 114,933 (65.88%), with 64,009 transcripts homologous to all the four cactus databases (Fig. 1b). Further, out of the total of transcripts, 44,040 transcripts (25.25%) showed homology only to sequences from cactus, but not for model plants *Arabidopsis*, tomato, or the commercial fruits included in this study (Fig. 1c).

A total of 45,970 (26.35%), 58,704 (33.65%), and 48,186 (27.65%) transcripts showed homology to transcription factors, transcriptional regulators, and protein kinases in the PlantTFDB, iTAK-TR, and iTAK-PK databases, respectively (Supplementary Tables S1, S9–S11). For the PlantTFDB, the homologous

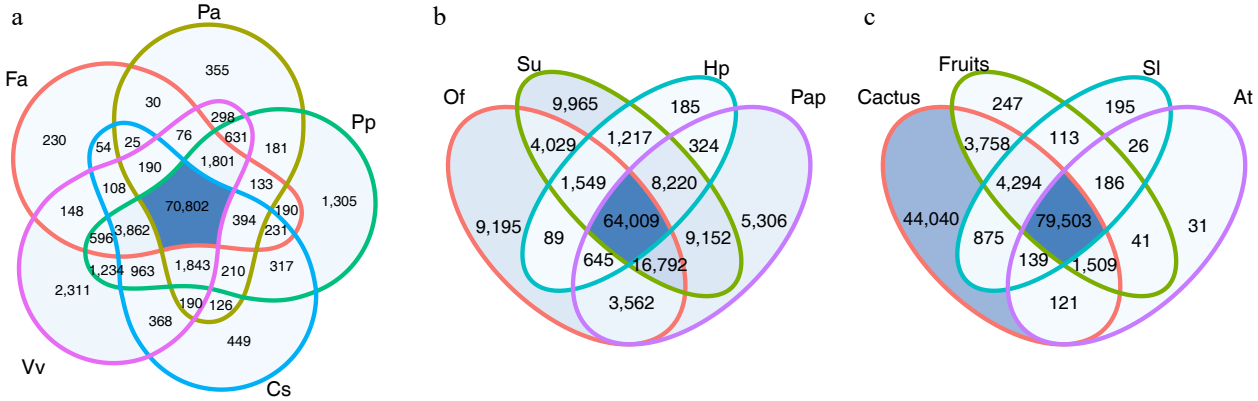

Fig. 1 Venn diagram of the homology search results against model plants databases, commercial fruits, and cactus. The number in the diagram corresponds to the number of transcripts from *S. thurberi* homologous to sequences from that plant species. (a) Homologous to sequences from *Fragaria vesca* (Fa), *Persea americana* (Pa), *Prunus persica* (Pp), *Vitis vinifera* (Vv), and *Citrus sinensis* (Cs). (b) Homologous to sequences from *Opuntia streptacantha* (Of), *Selenicereus undatus* (Su), *Hylocereus polyrhizus* (Hp), and *Pachycereus pringlei* (Pap). (c) Homologous to sequences from *Solanum lycopersicum* (Sl), *Arabidopsis thaliana* (At), from the commercial fruits (Fa, Pa, Pp, Vv, and Cs), or the cactus included in this study (Of, Su, Hp, and Pap). Homologous searching was carried out by BLAST alignment (E value < 1×10^{-5}). The Venn diagrams were drawn by ggVennDiagram in R Studio.

transcripts belong to 57 transcriptional factors (TF) families (Fig. 2 & Supplementary Table S9), from which, the most frequent were the basic-helix-loop-helix (bHLH), myeloblastosis-related (MYB-related), NAM, ATAF, and CUC (NAC), ethylene responsive factor (ERF), and the WRKY domain families (WRKY) (Fig. 2).

Functional categorization

Based on the homology found and the functional domain searches, gene ontology terms (GO) were assigned to 68,559 transcripts (Supplementary Table S12). Figure 3 shows the top 20 GO terms assigned to the *S. thurberi* transcriptome, corresponding to the Biological Processes (BP) and Molecular Function (MF) categories. For BP, organic substance metabolic processes, primary metabolic processes, and cellular metabolic processes showed a higher number of transcripts (Supplementary Table S13). Further, for MF, organic cyclic compound binding, heterocyclic compound binding, and ion binding were the processes with the higher number of transcripts. *S. thurberi* transcripts were classified into 142 metabolic pathways from the KEGG database (Supplementary Table S14). The pathways with the higher number of transcripts recorded were pyruvate metabolism, glycerophospholipid metabolism, glycolysis/gluconeogenesis, and citrate cycle. Further, among the top 20 KEEG pathways, the cutin, suberin, and wax biosynthesis include more than 30 transcripts (Fig. 4).

Identification of lncRNA

Out of the total of transcripts, 43,391 (24.87%) were classified as lncRNA (Supplementary Tables S15 & S16). Figure 5

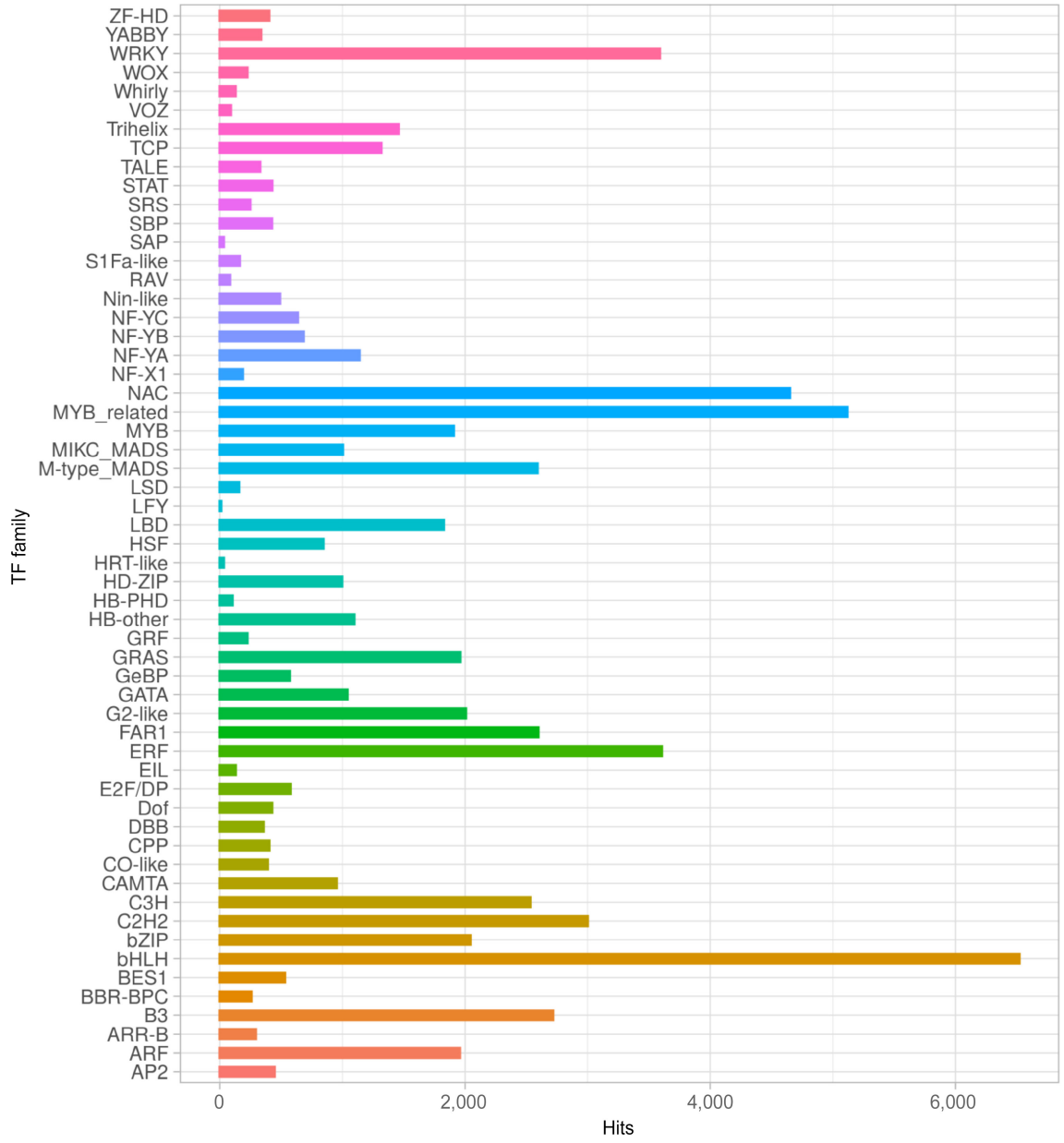

Fig. 2 Transcription factor (TF) families distribution of *S. thurberi* fruit peel transcriptome. The X-axis indicates the number of transcripts with hits to each TF family. Alignment to the PlantTFDB database by BLASTx was carried out with an E value threshold of 1×10^{-5}. The bar graph was drawn by ggplot2 in R Studio.

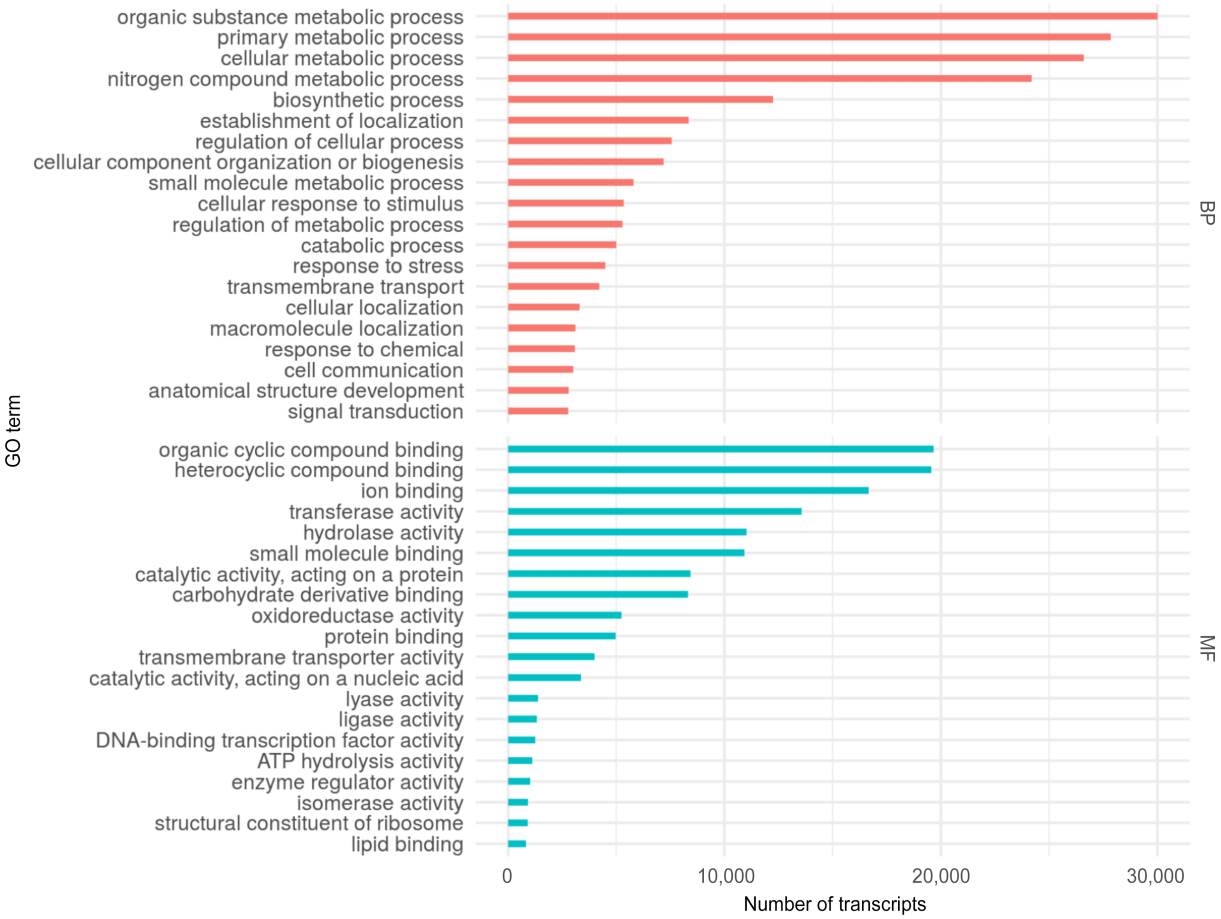

Fig. 3 Top 20 Gene Ontology (GO) terms assigned to the *S. thurberi* fruit peel transcriptome. Bars indicate the number of transcripts assigned to each GO term. Assignment of GO terms was carried out by Blast2GO with default parameters. BP and MF mean Biological Processes and Molecular Functions GO categories, respectively. The graph was drawn by ggplot2 in R Studio.

shows a comparison of the length (Fig. 5a) and expression (Fig. 5b) of lncRNA and coding RNA. Both length and expression values were higher in coding RNA than in lncRNA. In general, coding RNA ranged from 201 to 18,629 bp with a mean length of 1,507.18, whereas lncRNA ranged from 200 to 5,198 bp with a mean length of 481.51 (Fig. 5a). The higher expression values recorded from coding RNA and lncRNA were 12.83 and 9.45 $\log_2$(CPM), respectively (Fig. 5b).

Identification of tentative reference genes

To identify the transcripts without significant changes in expression between the four RNA-seq libraries, a differential expression analysis was carried out. Of the total of transcripts, 4,980 were not differentially expressed (NDE) at least in one paired comparison between the libraries (Supplementary Tables S17–S20). Mean counts per million of reads (CPM) and coefficient of variation (CV)[44] were calculated for these NDE transcripts. Transcripts with a CV value lower than 0.113, corresponding with the percentile 5 of the CV, and a mean CPM higher than 1,138.06, corresponding with the percentile 95 of the mean CPM were used as filters to identify the most stably expressed transcripts (Supplementary Table S21). Based on its homology and its tentative biological function, five transcripts were selected to be tested as tentative reference genes. Besides, three NDE transcripts homologous to previously identified stable expressed reference genes in other species of cactus fruit[41–43] were selected (Supplementary Table S22). Homology metrics for the eight tentative reference genes selected are shown in Supplementary Table S23. The primer sequences used to amplify the transcripts by qRT-PCR and their nucleotide sequence are shown in Supplementary Tables S24 & S25, respectively.

Expression stability of tentative reference genes

The amplification specificity of the eight candidate reference genes determined by melting curves analysis is shown in Supplementary Fig. S3. For the eight tentative reference transcripts selected, the cycle threshold (Ct) values were recorded during sweet pitaya fruit development by qRT-PCR (Supplementary Table S26). The Ct values obtained ranged from 16.85 to 30.26 (Fig. 6a). Plastidic ATP/ADP-transporter (*StTLC1*) showed the highest Ct values with a mean of 27.34 (Supplementary Table S26). Polyubiquitin 3 (*StUBQ3*) showed the lowest Ct values in all five sweet pitaya fruit developmental stages (Fig. 6a).

The best stability values calculated by NormFinder were 0.45, 0.51, 0.97, and 0.99, corresponding to the transcripts elongation factor 1-alpha (*StEF1a*), alpha-tubulin (*StTUA*), plastidic ATP/ADP-transporter (*StTLC1*), and actin 7 (*StACT7*), respectively (Supplementary Table S27). For BestKeeper, the most stable expressed transcripts were *StUBQ3*, *StTUA*, and *StEF1a*, with values of 0.72, 0.75, and 0.87, respectively. In the case of

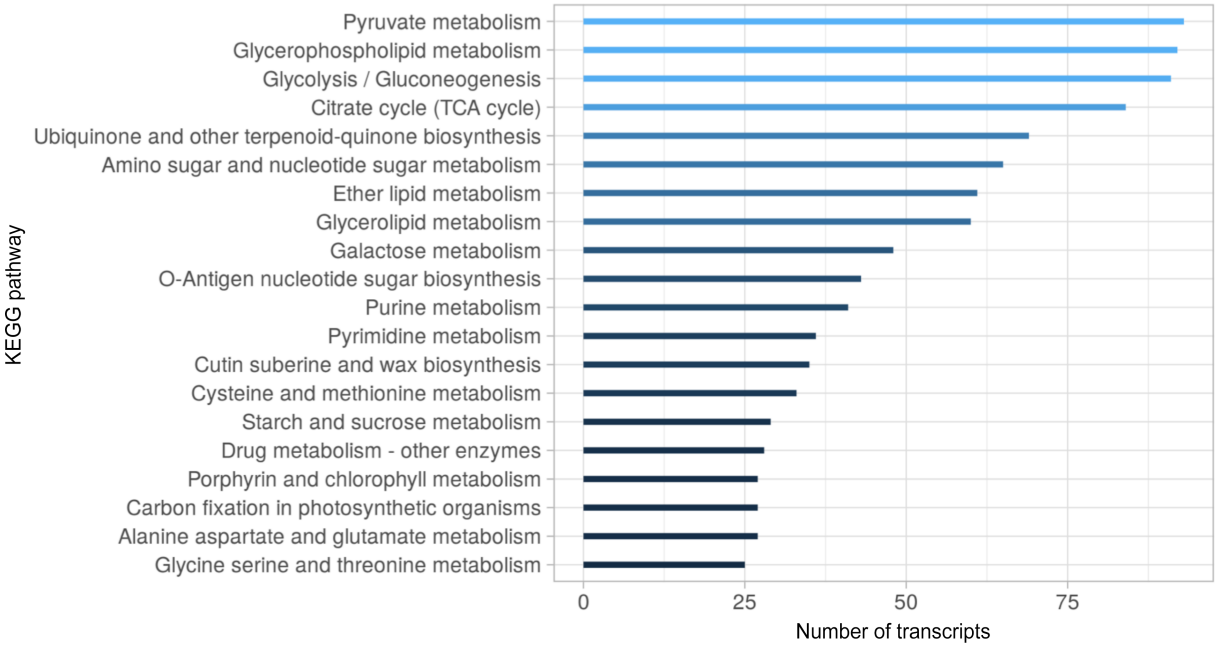

Fig. 4 Top 20 KEGG metabolic pathways distribution in the *S. thurberi* fruit peel transcriptome. Bars indicate the number of transcripts assigned to each KEGG pathway. Assignment of KEGG pathways was carried out in the Blast2GO suite. The bar graph was drawn by ggplot2 in R Studio.

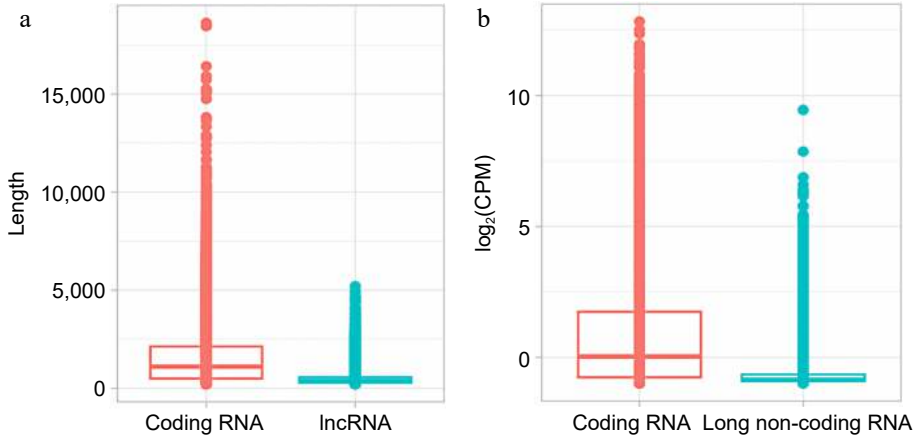

Fig. 5 Comparison of coding RNA and long non-coding RNA (lncRNA) from *S. thurberi* transcriptome. (a) Box plot of transcript length distribution. The Y-axis indicates the length of each transcript in base pairs. (b) Box plot of expression levels. The Y-axis indicates the $\log_2$ of the count per million of reads ($\log_2$(CPM)) recorded for each transcript. Expression levels were calculated by the edgeR package in R studio. (a), (b) The lines inside the boxes indicate the median. The higher and lower box limits represent the 75th and 25th percentiles, respectively. The box plots were drawn by ggplot2 in R Studio.

the delta Ct method[51], the transcripts *StEF1a*, *StTUA*, and *StTLC1* showed the best stability.

According to geNorm analysis, the most stable expressed transcripts were *StTUA*, *StEF1a*, *StUBQ3*, and *StACT7*, with values of 0.74, 0.74, 0.82, and 0.96, respectively. All the pairwise variation values (Vn/Vn + 1) were lower than 0.15, ranging from 0.019 for V2/V3 to 0.01 for V6/V7 (Fig. 6c). The V value of 0.019 obtained for V2/V3 indicates that the use of the best two reference genes *StTUA* and *StEF1a* is reliable enough for the accurate normalization of qRT-PCR data, therefore no third reference gene is required[47]. Except for BestKeeper analysis, *StEF1a* and *StTUA* were the most stable transcripts for all of the methods carried out in this study (Supplementary Table S27). The comprehensive ranking analysis indicates that *StEF1a*, followed by *StTUA* and *StUBQ3*, are the most stable expressed genes and are stable enough to be used as reference genes in qRT-PCR analysis during sweet pitaya fruit development (Fig. 6b).

Identification of cuticle biosynthesis-related transcripts

Three cuticle biosynthesis-related transcripts TRINITY_DN17030_c0_g1_i2, TRINITY_DN15394_c0_g1_i1, and TRINITY_DN23528_c1_g1_i1 tentatively coding for the enzymes cytochrome p450 family 77 subfamily A (CYP77A), Gly-Asp-Ser-Leu motif lipase/esterase 1 (GDSL1), and an ATP binding cassette transporter subfamily G member 11 (ABCG11/WBC11), respectively, were identified and quantified. The nucleotide sequence

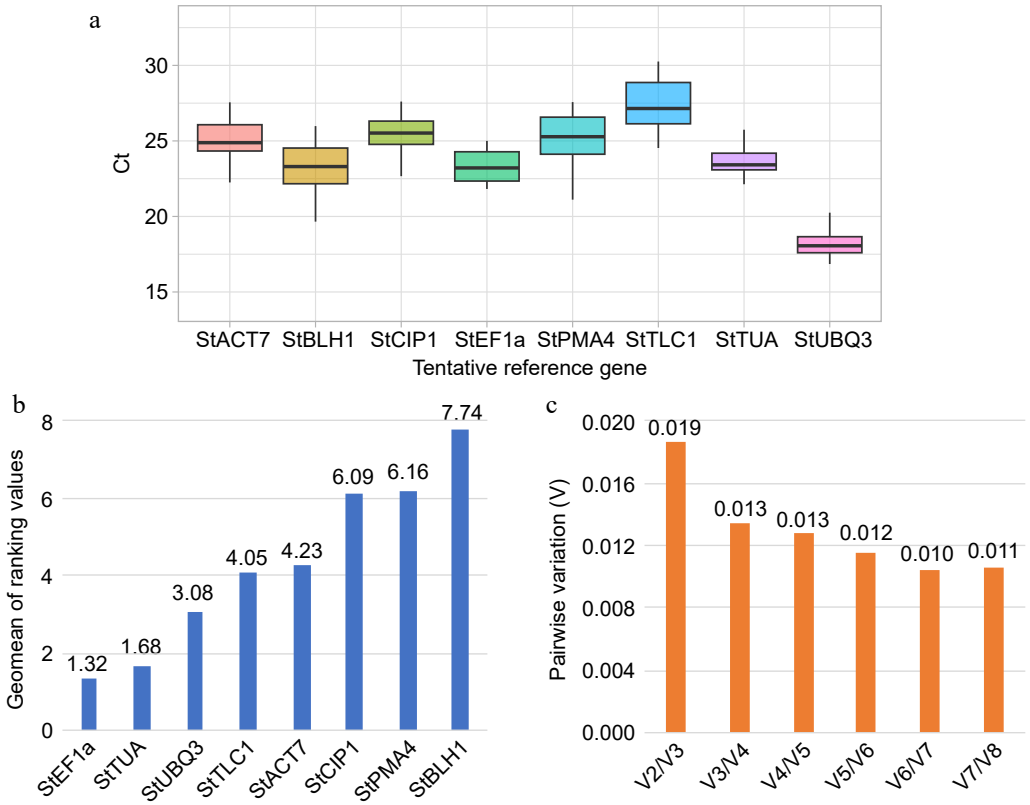

Fig. 6 Expression stability analysis of tentative reference genes. (a) Box plot of cycle threshold (Ct) distribution of candidate reference genes during sweet pitaya fruit development (10, 20, 30, 35, and 40 d after flowering). The black line inside the box indicates the median. The higher and lower box limits represent the 75th and 25th percentiles, respectively. (b) Bar chart of the geometric mean (geomean) of ranking values calculated by RefFinder for each tentative reference gene (X-axis). The lowest values indicate the best reference genes. (c) Bar chart of the pairwise variation analysis and determination of the optimal number of reference genes by the geNorm algorithm. A pairwise variation value lower than 0.15 indicates that the use of Vn/Vn + 1 reference genes is reliable for the accurate normalization of qRT-PCR data. The Ct data used in the analysis were calculated by qRT-PCR in a QIAquant 96 5 plex (QIAGEN) according to the manufacturer's protocol. The box plot and the bar graphs were drawn by ggplot2 and Excel programs, respectively. Abbreviations: Actin 7 (*StACT7*), alpha-tubulin (*StTUA*), elongation factor 1-alpha (*StEF1a*), COP1-interactive protein 1 (*StCIP1*), plasma membrane ATPase 4 (*StPMA4*), BEL1-like homeodomain protein 1 (*StBLH1*), polyubiquitin 3 (*StUBQ3*), and plastidic ATP/ADP-transporter (*StTLC1*).

and predicted amino acid sequences of the three transcripts are shown in Supplementary File 1. The best homology match for *StCYP77A* (TRINITY_DN17030_c0_g1_i2) was for AtCYP77A4 (AT5G04660) from *Arabidopsis* and SmCYP77A2 (P37124) from eggplant (*Solanum melongena*) in the TAIR and Swiss-Prot databases, respectively (Supplementary Table S23).

TransDecoder, InterPro, and TMHMM analysis showed that *StCYP77A* codes a polypeptide of 518 amino acids (aa) in length that comprises a cytochrome P450 E-class domain (IPR002401) and a transmembrane region (residues 10 to 32). The phylogenetic tree constructed showed that *StCYP77A* is grouped in a cluster with all the CYP77A2 proteins included in this analysis, being closer to CYP77A2 (XP_010694692) from *B. vulgaris* and Cgig2_012892 (KAJ8441854) from *Carnegiea gigantean* (Supplementary Fig. S4).

StGDSL1 (TRINITY_DN15394_c0_g1_i1) alignment showed that it is homologous to a GDSL esterase/lipase from *Arabidopsis* (Q9LU14) and tomato (Solyc03g121180) (Supplementary Table S23). TransDecoder, InterPro, and SignalP analysis showed that *StGDSL1* codes a polypeptide of 354 aa in length that comprises a GDSL lipase/esterase domain IPR001087 and a signal peptide with a cleavage site between position 25 and 26 (Supplementary Fig. S5).

Supplementary Figure S6 shows the analysis carried out on the predicted amino acid sequence of *StABCG11* (TRINITY_DN23528_c1_g1_i1). The phylogenetic tree constructed shows three clades corresponding to the ABCG13, ABCG12, and ABCG11 protein classes with bootstrap support ranging from 40% to 100% (Supplementary Fig. S6a). *StABCG11* is grouped with all the ABCG11 transporters included in this study in a well-separated clade, being closely related to its tentative ortholog from *C. gigantean* Cgig2_004465 (KAJ8441854). InterPro and TMHMM results showed that the StABCG11 sequence contains an ABC-2 type transporter transmembrane domain (IPR013525; PF01061.27) with six transmembrane helices (Supplementary Fig. S6b).

The predicted protein sequence of StABCG11 is 710 aa in length, holding the ATP binding domain (IPR003439; PF00005.30) and the P-loop containing nucleoside triphosphate hydrolase domain (IPR043926; PF19055.3) of the ABC transporters of the G family. Multiple sequence alignment shows that the Walker A and B motif sequence and the ABC signature[15] are conserved between the ABCG11 transporters from *Arabidopsis*, tomato, *S. thurberi,* and *C. gigantean* (Supplementary Fig. S6c).

Evaluation of reliable reference genes and quantification of cuticle biosynthesis-related transcripts

According to the results of the expression stability analysis (Fig. 6), four normalization strategies were tested to quantify the three cuticle biosynthesis-related transcripts during sweet pitaya fruit development. The four strategies consist of normalizing by *StEF1a*, *StTUA*, *StUBQ3*, or *StEF1a*+*StTUA*. Primer sequences used to quantify the transcripts *StCYP77A* (TRINITY_DN17030_c0_g1_i2), *StGDSL1* (TRINITY_DN15394_c0_g1_i1), and *StABCG11* (TRINITY_DN23528_c1_g1_i1) by qRT-PCR during sweet pitaya fruit development are shown in Supplementary Table S24.

The three cuticle biosynthesis-related transcripts showed differences in expression during sweet pitaya fruit development (Supplementary Table S28). The same expression pattern was recorded for the three cuticle biosynthesis transcripts when normalization was carried out by *StEF1a, StTUA*, *StUBQ3*, or *StEF1a* + *StTUA* (Fig. 7). A higher expression of *StCYP77A* and *StGDSL1* are shown at the 10 and 20 DAF, showing a decrease at 30, 35, and 40 DAF. *StABCG11* showed a similar behavior, with a higher expression at 10 and 20 DAF and a reduction at 30 and 35 DAF. Nevertheless, unlike *StCYP77A* and *StGDSL1*, a significant increase at 40 DAF, reaching the same expression as compared with 10 DAF, is shown for *StABCG11* (Fig. 7).

Discussion

S. thurberi transcriptome quality is similar to that reported for *de novo* assembly of non-model plants

Characteristics of a well-assembled transcriptome include an N50 value closer to 2,000 bp, a high percentage of conserved transcripts completely assembled (> 80%), and a high proportion of reads mapping back to the assembled transcripts[52]. In the present study, the first collection of 174,449 transcripts from *S. thurberi* fruit peel are reported. The generated transcriptome showed an N50 value of 2,110 bp, a TransRate score of 0.05, and a GC percentage of 41.33 (Table 1), similar to that reported for other *de novo* plant transcriptome assemblies[53]. According to BUSCO, 85.4% of the orthologous genes from the *Embryophyta* databases completely matched the *S. thurberi* transcriptome, and only 3.9% were missing (Table 1). These results show that the *S. thurberi* transcriptome generated is not fragmented, and it is helpful in predicting the sequence of almost all the transcripts expressed in sweet pitaya fruit peel[24].

S. thurberi transcriptome shows higher homology to sequences from species of the same family

The percentage of transcripts homologous found, E values, and identity distribution (Supplementary Tables S1 & S2; Supplementary Fig. S2) were similar to that reported in the *de novo* transcriptome assembly for non-model plants and other cactus fruits[41–43,54] and further suggests that the transcriptome assembled of *S. thurberi* peel is robust[52]. Of the total of transcripts, 70,802 were common to all the five commercial fruit protein databases included in this study, which is helpful for the search for conserved orthologous involved in fruit development and ripening (Fig. 2a). A total of 34,513 transcripts (20%) show homology only to sequences in the cactus's databases, but not in the others (Supplementary Tables S1 & S2; Fig. 1c). This could suggest that a significant conservation of sequences among plants of the *Cactaceae* family exists which most likely are to have a function in this species adaptation to desert ecosystems.

Functional categorization analysis suggests a response to stress and an active cuticle biosynthesis in fruit pitaya peel

To infer the biological functionality represented by the *S. thurberi* fruit peel transcriptome, gene ontology (GO) terms and KEGG pathways were assigned. Of the main metabolic pathways assigned, 'glycerolipid metabolism' and 'cutin, suberine, and wax biosynthesis' suggests an active cuticle biosynthesis in pitaya fruit peel (Fig. 4). In agreement with the above, the main GO terms assigned for the molecular function (MF) category were 'organic cyclic compound binding', 'transmembrane transporter activity', and 'lipid binding' (Fig. 3). For the biological

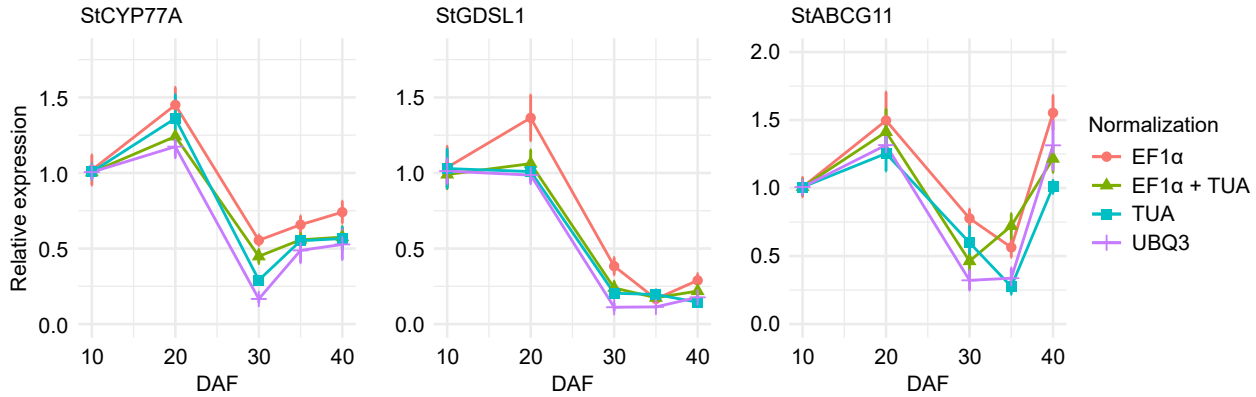

Fig. 7 Expression analysis of cuticle biosynthesis-related transcripts *StCYP77A*, *StGDSL1*, and *StABCG11* during sweet pitaya (*Stenocereus thurberi*) fruit development. Relative expression was calculated through the $2^{-\Delta\Delta CT}$ method using elongation factor 1-alpha (*StEF1a*), alpha-tubulin (*StTUA*), polyubiquitin 3 (*StUBQ3*), or *StEF1a* + *StTUA* as normalizing genes at 10, 20, 30, 35, and 40 d after flowering (DAF). The Y-axis and error bars represent the mean of the relative expression ± standard error (n = 4–6) for each developmental stage in DAF. The Ct data for the analysis was recorded by qRT-PCR in a QIAquant 96 5 plex (QIAGEN) according to the manufacturer's protocol. The graph line was drawn by ggplot2 in R Studio. Abbreviations: cytochrome p450 family 77 subfamily A (*StCYP77A*), Gly-Asp-Ser-Leu motif lipase/esterase 1 (*StGDSL1*), and ATP binding cassette transporter subfamily G member 11 (*StABCG11*).

processes (BP) category, the critical GO terms for the present research are 'cellular response to stimulus', 'response to stress', 'anatomical structure development', and 'transmembrane transport', which could suggest the active development of the fruit epidermis and cuticle biosynthesis for protection to stress.

The most frequent transcription factors (TF) families found in *S. thurberi* transcriptome were NAC, WRKY, bHLH, ERF, and MYB-related (Fig. 2), which had been reported to play a function in the tolerance to abiotic stress in plants[55,56]. Although the role of NAC, WRKY, bHLH, ERF, and MYB TF in improving drought tolerance in relevant crop plants has been widely documented[57,58], their contribution to the adaptation of cactus to arid ecosystems has not yet been elucidated and further experimental pieces of evidence are needed.

It has been reported that the heterologous expression of ERF TF from *Medicago truncatula* induces drought tolerance and cuticle wax biosynthesis in *Arabidopsis* leaf[59]. In tomato fruits, the gene *SlMIXTA-like* which encodes a MYB transcription factor avoids water loss through the positive regulation of genes related to the biosynthesis and transport of cuticle compounds[22]. Despite the relevant role of cuticles in maintaining cactus physiology in desert environments, experimental evidence showing the role of the different TF-inducing cuticle biosynthesis has yet to be reported for cactus fruits.

lncRNA from pitaya are similar in length and expression to that reported from other plants

Out of the transcripts, 43,391 were classified as lncRNA (Supplementary Tables S15 & S16). This is the first report of lncRNA identification for the species *S. thurberi*. In fruits, 3,679 lncRNA has been identified from tomato[26], 3,330 from peach (*P. persica*)[29], 3,857 from melon (*Cucumis melo*)[28], 2,505 from hot pepper (*Capsicum annuum*)[27], and 3,194 from pomegranate (*Punica granatum*)[36]. Despite the stringent criteria to classify the lncRNA of sweet pitaya fruit (*S. thurberi*), a higher number of lncRNAs are shown when compared with previous reports. This finding is most likely due to the higher level of redundancy found during the transcriptome analysis. To reduce this redundancy, further efforts to achieve the complete genome assembly of *S. thurberi* are needed.

Previous studies showed that lncRNA is shorter and has lower expression levels than coding RNA[60–62]. In agreement with those findings, both the length and expression values of lncRNA from *S. thurberi* were lower than coding RNA (Fig. 5). It has been suggested that lncRNA could be involved in the biosynthesis of cuticle components in cabbage[61] and pomegranate[36] and that they could be involved in the tolerance to water deficit through the regulation of cuticle biosynthesis in wild banana[60]. Nevertheless, the molecular mechanism by which lncRNA may regulate the cuticle biosynthesis in *S. thurberi* fruits has not yet been elucidated.

Previously reported housekeeping genes were found to have the most stable expression during pitaya fruit development

A relatively constant level of expression characterizes housekeeping genes because they are involved in essential cellular functions. These genes are not induced under specific conditions such as biotic or abiotic stress. Because of this, they are very useful as internal reference genes for qRT-PCR data normalization[63]. Nevertheless, their expression could change depending on plant species, developmental stages, and experimental conditions[64]. Reliable reference genes for a specific experiment in a given species must be identified to carry out an accurate qRT-PCR data normalization[63]. An initial screening of the transcript expression pattern through RNA-seq improves the identification of stably expressed transcripts by qRT-PCR[44,64].

Identification of stable expressed reference transcripts during fruit development has been carried out in blueberry (*Vaccinium bracteatum*)[65], kiwifruit (*Actinidia chinensis*)[66], peach (*P. persica*)[67], apple (*Malus domestica*)[68], and soursop (*Annona muricata*)[69]. These studies include the expression stability analysis through geNorm, NormFinder, and BestKeeper algorithms[68,69], some of which are supported in RNA-seq data[65,66]. Improvement of expression stability analysis by RNA-seq had led to the identification of non-previously reported reference genes with a more stable expression during fruit development than commonly known housekeeping genes in grapevine (*V. vinifera*)[44], pear (*Pyrus pyrifolia* and *P. calleryana*)[64], and pepper (*C. annuum*)[70].

For fruits of the *Cactaceae* family, only a few studies identifying reliable reference genes have been reported[41–43]. Mainly because gene expression analysis has not been carried out previously for sweet pitaya (*S. thurberi*), the RNA-seq data generated in this work along with geNorm, NormFinder, BestKeeper, and RefFinder algorithms were used to identify reliable reference genes. The comprehensive ranking analysis showed that out of the eight candidate genes tested, *StEF1a* followed by *StTUA* and *StUBQ3* were the most stable (Fig. 6b). All the pairwise variation values (Vn/Vn + 1) were lower than 0.15 (Fig. 6c), which indicates that *StEF1a*, *StTUA*, and *StUBQ3* alone or the use of *StEF1a* and *StTUA* together are reliable enough to normalize the gene expression data generated by qRT-PCR.

The genes *StEF1a*, *StTUA*, and *StUBQ3* are homologous to transcripts found in the cactus species known as dragonfruit (*Hylocereus monacanthus* and *H. undatus*)[41], which have been tested as tentative reference genes during fruit development. *EF1a* has been proposed as a reliable reference gene in the analysis of changes in gene expression of dragon fruit (*H. monacanthus* and *H. undatus*)[41], peach (*P. persica*)[67], apple (*M. domestica*)[68], and soursop (*A. muricata*)[69]. According to the expression stability analysis carried out in the present study (Fig. 6) four normalization strategies were designed. The same gene expression pattern was recorded for the three target transcripts evaluated when normalization was carried out by the genes *StEF1a*, *StTUA*, *StUBQ3*, or *StEF1a* + *StTUA* (Fig. 7). Further, these data indicates that these reference genes are reliable enough to be used in qRT-PCR experiments during fruit development of *S. thurberi*.

Cutin biosynthesis could have a relevant role in the first stages of pitaya fruit development

The plant cuticle is formed by two main layers: the cutin, composed mainly of mid-chain oxygenated LC fatty acids, and the cuticular wax, composed mainly of very long-chain (VLC) fatty acids, and their derivates VLC alkanes, VLC primary alcohols, VLC ketones, VLC aldehydes, and VLC esters[3]. In *Arabidopsis* CYP77A4 and CYP77A6 catalyze the synthesis of midchain epoxy and hydroxy ω-OH long-chain fatty acids, respectively[10,11], which are the main components of fleshy fruit cuticle[3].

The functional domain search carried out in the present study showed that *StCYP77A* comprises a cytochrome P450

E-class domain (IPR002401) and a membrane-spanning region from residues 10 to 32 (Supplementary Fig. S4). This membrane-spanning region has been previously characterized in CYP77A enzymes from *A. thaliana* and *Brassica napus*[11,71]. It suggests that the protein coded by *StCYP77A* could catalyze the oxidation of fatty acids embedded in the endoplasmic reticulum membrane of the epidermal cells of *S. thurberi* fruit. Phylogenetic analysis showed that StCYP77A was closer to proteins from its phylogenetic-related species *B. vulgaris* (BvCYP772; XP_010694692) and *C. gigantea* (Cgig2_012892) (Supplementary Fig. S4). StCYP77A, BvCYP77A2, and Cgig2_012892 were closer to SlCYP77A2 and SmCYP77A2 than to CYP77A4 and CYP77A6 proteins, suggesting that StCYP77A (TRINITY_DN17030_c0_g1_i2) could correspond to a CYP77A2 protein.

Five *CYP77A* are present in the *Arabidopsis* genome, named *CYP77A4*, *CYP77A5*, *CYP77A6*, *CYP77A7*, and *CYP77A9*, but their role in cuticle biosynthesis has only been reported for *CYP77A4* and *CYP77A6*[72]. It has been suggested that *CYP77A2* from eggplant (*S. torvum*) could contribute to the defense against fungal phytopathogen infection by the synthesis of specific compounds[13]. In pepper fruit (*C. annuum*), the expression pattern of *CYP77A2* (A0A1U8GYB0) and *ABCG11* (LOC107862760) suggests a role of CYP77A2 and ABCG11 in cutin biosynthesis at the early stages of pepper fruit development[14].

In the case of the protein encoded by *StGDSL1* (354 aa), the length found in this work is similar to the length of its homologous from *Arabidopsis* (AT3G16370) and tomato (Solyc03g121180) (Supplementary Fig. S5). A GDSL1 protein named CD1 polymerizes midchain oxygenated ω-OH long-chain fatty acids to form the cutin polyester in the extracellular space of tomato fruit peel[20,21]. It has been suggested that the 25-amino acid N-signal peptide found in StGDSL1 (Supplementary Fig. S5), previously reported in GDSL1 from *Arabidopsis*, *B. napus*, and tomato, plays a role during the protein exportation to the extracellular space[21,73].

A higher expression of *StCYP77A*, *StGDSL1*, and *StABCG11* is shown at the 10 and 20 DAF of sweet pitaya fruit development (Fig. 7), suggesting the active cuticle biosynthesis at the early stages of sweet pitaya fruit development. In agreement with that, two genes coding for GDSL lipase/hydrolases from tomato named *SGN-U583101* and *SGN-U579520* are highly expressed in the early stages and during the expansion stages of tomato fruit development, but their expression decreases in later stages[74]. It has been shown that the expression of *GDSL* genes, like *CD1* from tomato, is higher in growing fruit[20,21]. Like tomato, the increase in expression of *StCYP77A* and *StGDSL1* shown in pitaya fruit development could be due to an increase in cuticle deposition during the expansion of the fruit epidermis[20].

StABCG11 could be playing a relevant role in the transport of cuticle components during sweet pitaya fruit ripening

The phylogenetic analysis, the functional domains, and the six transmembrane helices found in the StABCG11 predicted protein (Supplementary Fig. S6), suggests that it is an ABCG plasma membrane transporter of sweet pitaya[15]. Indeed, an increased expression of *StABCG11* at 40 DAF was recorded in the present study (Fig. 7). Further, this data strongly suggests that it could be playing a relevant role in the transport of cuticle components at the beginning and during sweet pitaya fruit ripening.

In *Arabidopsis*, ABCG11 (WBC11) exports cuticular wax and cutin compounds from the plasma membrane[15,75]. It has been reported that a high expression of the ABC plasma membrane transporter from mango *MiWBC11* correlates with a higher cuticle deposition during fruit development[7]. The expression pattern for *StABCG11*, *StCYP77A*, and *StGDSL1* suggests a role of StABCG11 as a cutin compound transporter in the earlier stages of sweet pitaya fruit development (Fig. 7). Further, its increase at 40 DAF suggests that it could be transporting cuticle compounds other than oxygenated long-chain fatty acids, or long-chain fatty acids that are not synthesized by StCYP77A and StGDSL1 in the later stages of fruit development.

Like sweet pitaya, during sweet cherry fruit (*Prunus avium*) development, the expression of *PaWCB11*, homologous to *AtABCG11* (AT1G17840), increases at the earlier stages of fruit development decreases at the intermediate stages, and increases again at the later stages[76]. *PaWCB11* expression correlated with cuticle membrane deposition at the earlier and intermediate stages of sweet cherry fruit development but not at the later[76]. The increased expression of *StABCG11* found in the present study could be due to the increased transport of cuticular wax compounds, such as VLC fatty acids and their derivates, in the later stages of sweet pitaya development[15,75].

Cuticular waxes make up the smallest amount of the fruit cuticle. Even so, they mainly contribute to the impermeability of the fruit's epidermis[3]. An increase in the transport of cuticular waxes at the beginning of the ripening stage carried out by ABCG transporters could be due to a greater need to avoid water loss and to maintain an adequate amount of water during the ripening of the sweet pitaya fruit. Nevertheless, further expression analysis of cuticular wax biosynthesis-related genes, complemented with chemical composition analysis of cuticles could contribute to elucidating the molecular mechanism of cuticle biosynthesis in cacti and their physiological contribution during fruit development.

Conclusions

In this study, the transcriptome of the sweet pitaya (*S. thurberi*) fruit peel was assembled for the first time. The reference genes found here are a helpful tool for further gene expression analysis in sweet pitaya fruit. Transcripts tentatively involved in cuticle compound biosynthesis and transport are reported for the first time in sweet pitaya. The results suggest a relevant role of cuticle compound biosynthesis and transport at the early and later stages of fruit development. The information generated will help to improve the elucidation of the molecular mechanism of cuticle biosynthesis in *S. thurberi* and other cactus species in the future. Understanding the cuticle's physiological function in the adaptation of the *Cactaceae* family to harsh environmental conditions could help design strategies to increase the resistance of other species to face the increase in water scarcity for agricultural production predicted for the following years.

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Tiznado-Hernández ME, Tafolla-Arellano JC, García-Coronado H, Hernández-Oñate MÁ; data collection: Tiznado-Hernández ME, Tafolla-Arellano JC, García-Coronado H, Hernández-Oñate MÁ; analysis and

interpretation of results: Tiznado-Hernández ME, García-Coronado H, Hernández-Oñate MÁ, Burgara-Estrella AJ; draft manuscript preparation: Tiznado-Hernández ME, García-Coronado H. All authors reviewed the results and approved the final version of the manuscript.

Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files. The sequence data can be accessed at the Sequence Read Archive (SRA) repository of the NCBI through the BioProject ID PRJNA1030439.

Acknowledgments

The authors wish to acknowledge the financial support of Consejo Nacional de Humanidades, Ciencias y Tecnologías de México (CONAHCYT) through project number 579: Elucidación del Mecanismo Molecular de Biosíntesis de Cutícula Utilizando como Modelo Frutas Tropicales. We appreciate the University of Arizona Genetics Core and Illumina for providing reagents and equipment for library sequencing. The author, Heriberto García-Coronado (CVU 490952), thanks the CONAHCYT (acronym in Spanish) for the Ph.D. scholarship assigned (749341). The author, Heriberto García-Coronado, thanks Dr. Edmundo Domínguez-Rosas for the technical support in bioinformatics for identifying long non-coding RNA.

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary information accompanies this paper at (https://www.maxapress.com/article/doi/10.48130/vegres-0024-0031)

Dates

Received 28 May 2024; Revised 26 July 2024; Accepted 14 August 2024; Published online 4 November 2024

References

1. Hultine KR, Hernández-Hernández T, Williams DG, Albeke SE, Tran N, et al. 2023. Global change impacts on cacti (*Cactaceae*): current threats, challenges and conservation solutions. *Annals of Botany* 132:671–83
2. Liu L, Wang X, Chang C. 2022. Toward a smart skin: harnessing cuticle biosynthesis for crop adaptation to drought, salinity, temperature, and ultraviolet stress. *Frontiers in Plant Science* 13:961829
3. García-Coronado H, Tafolla-Arellano JC, Hernández-Oñate MÁ, Burgara-Estrella AJ, Robles-Parra JM, et al. 2022. Molecular biology, composition and physiological functions of cuticle lipids in fleshy fruits. *Plants* 11:1133
4. Matas AJ, Yeats TH, Buda GJ, Zheng Y, Chatterjee S, et al. 2011. Tissue- and cell-type specific transcriptome profiling of expanding tomato fruit provides insights into metabolic and regulatory specialization and cuticle formation. *The Plant Cell* 23:3893–910
5. Albert Z, Ivanics B, Molnár A, Miskó A, Tóth M, et al. 2013. Candidate genes of cuticle formation show characteristic expression in the fruit skin of apple. *Plant Growth Regulation* 70:71–78
6. Alkio M, Jonas U, Declercq M, Van Nocker S, Knoche M. 2014. Transcriptional dynamics of the developing sweet cherry (*Prunus avium* L.) fruit: sequencing, annotation and expression profiling of exocarp-associated genes. *Horticulture Research* 1:11
7. Tafolla-Arellano JC, Zheng Y, Sun H, Jiao C, Ruiz-May E, et al. 2017. Transcriptome analysis of mango (*Mangifera indica* L.) fruit epidermal peel to identify putative cuticle-associated genes. *Scientific Reports* 7:46163
8. Wu X, Shi X, Bai M, Chen Y, Li X, et al. 2019. Transcriptomic and gas chromatography–mass spectrometry metabolomic profiling analysis of the epidermis provides insights into cuticular wax regulation in developing 'Yuluxiang' pear fruit. *Journal of Agricultural and Food Chemistry* 67:8319–31
9. Xiao F, Mark Goodwin S, Xiao Y, Sun Z, Baker D, et al. 2004. *Arabidopsis CYP86A2* represses *Pseudomonas syringae* type III genes and is required for cuticle development. *The EMBO Journal* 23:2903–13
10. Li-Beisson Y, Pollard M, Sauveplane V, Pinot F, Ohlrogge J, et al. 2009. Nanoridges that characterize the surface morphology of flowers require the synthesis of cutin polyester. *Proceedings of the National Academy of Sciences of the United States of America* 106:22008–13
11. Sauveplane V, Kandel S, Kastner PE, Ehlting J, Compagnon V, et al. 2009. *Arabidopsis thaliana* CYP77A4 is the first cytochrome P450 able to catalyze the epoxidation of free fatty acids in plants. *The FEBS Journal* 276:719–35
12. Grausem B, Widemann E, Verdier G, Nosbüsch D, Aubert Y, et al. 2014. CYP77A19 and CYP77A20 characterized from *Solanum tuberosum* oxidize fatty acids *in vitro* and partially restore the wild phenotype in an *Arabidopsis thaliana* cutin mutant. *Plant, Cell & Environment* 37:2102–15
13. Yang L, Shi C, Mu X, Liu C, Shi K, et al. 2015. Cloning and expression of a wild eggplant cytochrome P450 gene, *StoCYP77A2*, involved in plant resistance to *Verticillium dahliae*. *Plant Biotechnology Reports* 9:167–77
14. Ge S, Qin K, Ding S, Yang J, Jiang L, et al. 2022. Gas chromatography–mass spectrometry metabolite analysis combined with transcriptomic and proteomic provide new insights into revealing cuticle formation during pepper development. *Journal of Agricultural and Food Chemistry* 70:12383–97
15. Luo B, Xue XY, Hu WL, Wang LJ, Chen XY. 2007. An ABC transporter gene of *Arabidopsis thaliana*, *AtWBC11*, is involved in cuticle development and prevention of organ fusion. *Plant and Cell Physiology* 48:1790–802
16. Bird D, Beisson F, Brigham A, Shin J, Greer S, et al. 2007. Characterization of Arabidopsis ABCG11/WBC11, an ATP binding cassette (ABC) transporter that is required for cuticular lipid secretion. *The Plant Journal* 52:485–98
17. McFarlane HE, Shin JJH, Bird DA, Samuels AL. 2010. *Arabidopsis* ABCG transporters, which are required for export of diverse cuticular lipids, dimerize in different combinations. *The Plant Cell* 22:3066–75
18. Panikashvili D, Shi JX, Schreiber L, Aharoni A. 2011. The Arabidopsis ABCG13 transporter is required for flower cuticle secretion and patterning of the petal epidermis. *New Phytologist* 190:113–24
19. Chen N, Song B, Tang S, He J, Zhou Y, et al. 2018. Overexpression of the ABC transporter gene *TsABCG11* increases cuticle lipids and abiotic stress tolerance in *Arabidopsis*. *Plant Biotechnology Reports* 12:303–13
20. Yeats TH, Martin LBB, Viart HMF, Isaacson T, He Y, et al. 2012. The identification of cutin synthase: formation of the plant polyester cutin. *Nature Chemical Biology* 8:609–11
21. Girard AL, Mounet F, Lemaire-Chamley M, Gaillard C, Elmorjani K, et al. 2012. Tomato GDSL1 is required for cutin deposition in the fruit cuticle. *The Plant Cell* 24:3119–34
22. Lashbrooke J, Adato A, Lotan O, Alkan N, Tsimbalist T, et al. 2015. The tomato MIXTA-like transcription factor coordinates fruit

epidermis conical cell development and cuticular lipid biosynthesis and assembly. *Plant Physiology* 169:2553–71

23. Castro-Enríquez DD, Montaño-Leyva B, Del Toro-Sánchez CL, Juárez-Onofre JE, Carvajal-Millán E, et al. 2020. Effect of ultrafiltration of Pitaya extract (*Stenocereus thurberi*) on its phytochemical content, antioxidant capacity, and UPLC-DAD-MS profile. *Molecules* 25:281
24. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, et al. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 8:1494–512
25. Patra GK, Gupta D, Rout GR, Panda SK. 2023. Role of long non coding RNA in plants under abiotic and biotic stresses. *Plant Physiology and Biochemistry* 194:96–110
26. Zhu B, Yang Y, Li R, Fu D, Wen L, et al. 2015. RNA sequencing and functional analysis implicate the regulatory role of long non-coding RNAs in tomato fruit ripening. *Journal of Experimental Botany* 66:4483–95
27. Ou L, Liu Z, Zhang Z, Wei G, Zhang Y, et al. 2017. Noncoding and coding transcriptome analysis reveals the regulation roles of long noncoding RNAs in fruit development of hot pepper (*Capsicum annuum* L.). *Plant Growth Regulation* 83:141–56
28. Tian Y, Bai S, Dang Z, Hao J, Zhang J, Hasi A, et al. 2019. Genome-wide identification and characterization of long non-coding RNAs involved in fruit ripening and the climacteric in *Cucumis melo*. *BMC Plant Biology* 19:369
29. Zhou H, Ren F, Wang X, Qiu K, Sheng Y, et al. 2022. Genome-wide identification and characterization of long noncoding RNAs during peach (*Prunus persica*) fruit development and ripening. *Scientific Reports* 12:11044
30. Wan CY, Wilkins TA. 1994. A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (*Gossypium hirsutum* L.). *Analytical Biochemistry* 223:7–12
31. Bairoch A, Apweiler R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research* 28:45–48
32. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, et al. 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* 40:D1202–D1210
33. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* 44:D733–D745
34. Jin J, Tian F, Yang DC, Meng YQ, Kong L, et al. 2017. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research* 45:D1040–D1045
35. Conesa A, Götz S. 2008. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics* 2008:619832
36. Wang Y, Zhao Y, Wu Y, Zhao X, Hao Z, et al. 2022. Transcriptional profiling of long non-coding RNAs regulating fruit cracking in *Punica granatum* L. under bagging. *Frontiers in Plant Science* 13:943547
37. Kalvari I, Nawrocki EP, Argasinska J, Quinones-Olvera N, Finn RD, et al. 2018. Non-coding RNA analysis using the Rfam database. *Current Protocols in Bioinformatics* 62:e51
38. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, et al. 2007. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research* 35:W345–W349
39. Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, et al. 2017. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Research* 45:W12–W16
40. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–40
41. Chen C, Wu J, Hua Q, Tel-Zur N, Xie F, et al. 2019. Identification of reliable reference genes for quantitative real-time PCR normalization in pitaya. *Plant Methods* 15:70
42. Nong Q, Yang Y, Zhang M, Zhang M, Chen J, et al. 2019. RNA-seq-based selection of reference genes for RT-qPCR analysis of pitaya. *FEBS Open Bio* 9:1403–12
43. Zheng Q, Wang X, Qi Y, Ma Y. 2021. Selection and validation of reference genes for qRT-PCR analysis during fruit ripening of red pitaya (*Hylocereus polyrhizus*). *FEBS Open Bio* 11:3142–52
44. González-Agüero M, García-Rojas M, Di Genova A, Correa J, Maass A, et al. 2013. Identification of two putative reference genes from grapevine suitable for gene expression analysis in berry and related tissues derived from RNA-Seq data. *BMC Genomics* 14:878
45. Thornton B, Basu C. 2015. Rapid and simple method of qPCR primer design. In *PCR Primer Design*, ed. Basu C. New York, NY: Humana Press. pp. 173–79. doi:10.1007/978-1-4939-2365-6_13
46. Xie F, Wang J, Zhang B. 2023. RefFinder: a web-based tool for comprehensively analyzing and identifying reference genes. *Functional & Integrative Genomics* 23:125
47. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, et al. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* 3:research0034.1
48. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113
49. Tamura K, Stecher G, Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution* 38:3022–27
50. Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* 25:402–08
51. Silver N, Best S, Jiang J, Thein SL. 2006. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Molecular Biology* 7:33
52. Raghavan V, Kraft L, Mesny F, Rigerte L. 2022. A simple guide to *de novo* transcriptome assembly and annotation. *Briefings in Bioinformatics* 23:bbab563
53. Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, et al. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574:679–85
54. Xi X, Zong Y, Li S, Cao D, Sun X, et al. 2019. Transcriptome analysis clarified genes involved in betalain biosynthesis in the fruit of red pitayas (*Hylocereus costaricensis*). *Molecules* 24:445
55. Erpen L, Devi HS, Grosser JW, Dutt M. 2018. Potential use of the DREB/ERF, MYB, NAC and WRKY transcription factors to improve abiotic and biotic stress in transgenic plants. *Plant Cell, Tissue and Organ Culture (PCTOC)* 132:1–25
56. Hu Y, Chen X, Shen X. 2022. Regulatory network established by transcription factors transmits drought stress signals in plant. *Stress Biology* 2:26
57. Jiang B, Liu R, Fang X, Tong C, Chen H, et al. 2022. Effects of salicylic acid treatment on fruit quality and wax composition of blueberry (*Vaccinium virgatum* Ait). *Food Chemistry* 368:130757
58. Baillo EH, Kimotho RN, Zhang Z, Xu P. 2019. Transcription factors associated with abiotic and biotic stress tolerance and their potential for crops improvement. *Genes* 10:771
59. Zhang JY, Broeckling CD, Sumner LW, Wang ZY. 2007. Heterologous expression of two *Medicago truncatula* putative ERF transcription factor genes, *WXP1* and *WXP2*, in *Arabidopsis* led to increased leaf wax accumulation and improved drought tolerance, but differential response in freezing tolerance. *Plant Molecular Biology* 64:265–78
60. Liu W, Cheng C, Lin Y, XuHan X, Lai Z. 2018. Genome-wide identification and characterization of mRNAs and lncRNAs involved in cold stress in the wild banana (*Musa itinerans*). *PLoS One* 13:e0200002

61. Zhu X, Tai X, Ren Y, Chen J, Bo T. 2019. Genome-wide analysis of coding and long non-coding RNAs involved in cuticular wax biosynthesis in cabbage (*Brassica oleracea* L. var. *capitata*). *International Journal of Molecular Sciences* 20:2820
62. Corona-Gomez JA, Coss-Navarrete EL, Garcia-Lopez IJ, Klapproth C, Pérez-Patiño JA, et al. 2022. Transcriptome-guided annotation and functional classification of long non-coding RNAs in *Arabidopsis thaliana*. *Scientific Reports* 12:14063
63. Lim PK, Zheng X, Goh JC, Mutwil M. 2022. Exploiting plant transcriptomic databases: resources, tools, and approaches. *Plant Communications* 3:100323
64. Wang Y, Dai M, Cai D, Shi Z. 2019. Screening for quantitative real-time PCR reference genes with high stable expression using the mRNA-sequencing data for pear. *Tree Genetics & Genomes* 15:54
65. He F, Gui L, Zhang Y, Zhu B, Zhang X, et al. 2022. Validation of reference genes for gene expression analysis in fruit development of *Vaccinium bracteatum* Thunb. using quantitative real-time PCR. *Scientific Reports* 12:16946
66. Liu J, Huang S, Niu X, Chen D, Chen Q, et al. 2018. Genome-wide identification and validation of new reference genes for transcript normalization in developmental and post-harvested fruits of *Actinidia chinensis*. *Gene* 645:1–6
67. Kou X, Zhang L, Yang S, Li G, Ye J. 2017. Selection and validation of reference genes for quantitative RT-PCR analysis in peach fruit under different experimental conditions. *Scientia Horticulturae* 225:195–203
68. Zhu L, Yang C, You Y, Liang W, Wang N, et al. 2019. Validation of reference genes for qRT-PCR analysis in peel and flesh of six apple cultivars (*Malus domestica*) at diverse stages of fruit development. *Scientia Horticulturae* 244:165–71
69. Berumen-Varela G, Palomino-Hermosillo YA, Bautista-Rosales PU, Peña-Sandoval GR, López-Gúzman GG, et al. 2020. Identification of reference genes for quantitative real-time PCR in different developmental stages and under refrigeration conditions in soursop fruits (*Annona muricata* L.). *Scientia Horticulturae* 260:108893
70. Cheng Y, Pang X, Wan H, Ahammed GJ, Yu J, et al. 2017. Identification of optimal reference genes for normalization of qPCR analysis during pepper fruit development. *Frontiers in Plant Science* 8:1128
71. McCartney AW, Dyer JM, Dhanoa PK, Kim PK, Andrews DW, et al. 2004. Membrane-bound fatty acid desaturases are inserted co-translationally into the ER and contain different ER retrieval motifs at their carboxy termini. *The Plant Journal* 37:156–73
72. Pineau E, Sauveplane V, Grienenberger E, Bassard JE, Beisson F, et al. 2021. CYP77B1 a fatty acid epoxygenase specific to flowering plants. *Plant Science* 307:110905
73. Ding LN, Guo XJ, Li M, Fu ZL, Yan SZ, et al. 2019. Improving seed germination and oil contents by regulating the *GDSL* transcriptional level in *Brassica napus*. *Plant Cell Reports* 38:243–53
74. Yeats TH, Howe KJ, Matas AJ, Buda GJ, Thannhauser TW, et al. 2010. Mining the surface proteome of tomato (*Solanum lycopersicum*) fruit for proteins associated with cuticle biogenesis. *Journal of Experimental Botany* 61:3759–71
75. Panikashvili D, Savaldi-Goldstein S, Mandel T, Yifhar T, Franke RB, et al. 2007. The Arabidopsis *DESPERADO/AtWBC11* transporter is required for cutin and wax secretion. *Plant Physiology* 145:1345–60
76. Alkio M, Jonas U, Sprink T, van Nocker S, Knoche M. 2012. Identification of putative candidate genes involved in cuticle formation in *Prunus avium* (sweet cherry) fruit. *Annals of Botany* 110:101–12