

Characterization of the Sweet Pitaya (*Stenocereus thurberi*) Fruit Peel Transcriptome: Analysis of Genes Playing a Role in Cuticle Biosynthesis and Identification of Reference Genes

Heriberto García-Coronado

Centro de Investigación en Alimentación y Desarrollo A.C. Carretera Gustavo Enrique Astiazarán

Miguel-Angel Hernández-Oñate (✉ miguel.hernandez@ciad.mx)

CONACYT-Coordinación de Tecnología de Alimentos de Origen Vegetal

Julio-César Tafolla-Arellano

Universidad Autónoma Agraria Antonio Narro

Alexel-Jesús Burgara-Estrella

Universidad de Sonora

Martín-Ernesto Tiznado-Hernández (✉ tiznado@ciad.mx)

Centro de Investigación en Alimentación y Desarrollo A.C. Carretera Gustavo Enrique Astiazarán

Research Article

Keywords: *Stenocereus thurberi*, RNA-seq, Fruit Peel Transcriptome, Cuticle Biosynthesis Genes, Long Non-coding RNA, Reference Genes

Posted Date: September 15th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3349817/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Cactus (*Cactaceae*) are plants distributed across arid regions of America with ecological and economic value. 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. *Stenocereus thurberi* is a cactus endemic from the Sonoran desert, which produces a fruit named sweet pitaya. Transcripts from *S. thurberi* published in databases are scarce, and no gene expression analysis has been carried out for this species. This study reports for first time the *de novo* assembly and characterization of the sweet pitaya peel transcriptome.

Results

Two hundred forty-three million of 80–150 base pairs reads with at least 25 of quality in the Phred score were used to get the assembly. The transcriptome includes 174,449 transcripts with an N50 value of 2,110 bp and 85.4% of completeness. Out of the total transcripts, 122,234 (70.07%) were classified as coding RNA, and 43,391 were classified as long non-coding RNA. Functional categorization analysis suggests a response to stress and an active cuticle biosynthesis in fruit pitaya peel. The genes elongation factor 1-alpha (StEF1a), α -tubulin (StTUA), and polyubiquitin 3 (StUBQ3) are reliable reference genes for accurate normalization of gene expression analysis in this specie through qRT-PCR. The cuticle biosynthesis transcripts cytochrome p450 family 77 subfamily A (StCYP77A), Gly-Asp-Ser-Leu motif lipase/esterase 1 (StGDLS1), and ATP binding cassette transporter family G member 11 (StABCG11) showed higher expression at the early stages of fruit development and ripening suggesting an active cuticle compound biosynthesis and transport.

Conclusions

This is the first transcriptome developed in the *S. thurberi* specie. Further, housekeeping genes suitable for gene expression analysis by qRT-PCR in this specie are reported for first time. The information generated will help to analyze the molecular mechanism of cuticle biosynthesis and another relevant metabolic pathway in *S. thurberi* and other cactus species. Understanding the role of cuticle in the adaptation to arid environments could help design technologies to ensure fleshy fruit production in the context of the increase in water scarcity for agriculture predicted for the following years.

Background

Columnar cacti are plants of the *Cactaceae* family distributed across arid and semi-arid regions of America, with ecological, economic, and cultural value (1). Some of the traits that make it possible for the columnar cactus to survive in the harsh desert ecosystem are their crassulacean acid metabolism (CAM),

the modification of leaves into spines, photosynthetic and succulent stems, an extended root system, and a thick epidermis covered by a hydrophobic cuticle, which limits water loss in dry and hot conditions (1).

The cuticle is the external layer that covers the non-woody aerial organs of land plants. The careful control of the cuticle biosynthesis pathway could produce drought stress tolerance in relevant crop plants (2). In fleshy fruits, it 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). Despite the relevant role of cuticles in maintaining cacti 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, which produces an ovoid-globose fleshy fruit named sweet pitaya (9). In its mature state, the pulp of sweet pitaya contains around 87% water with a high content of antioxidants and natural pigments such as betalains and phenolic compounds, which have nutraceutical and industrial relevance (9, 10). 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 coding and non-coding transcripts from non-model plants, even if no complete assembled genome is available (11). Long non-coding RNAs (lncRNAs) play regulatory roles in relevant biological processes such as the regulation of drought stress tolerance in plants (12), fruit development, and ripening (13–16). It has been suggested that lncRNAs could be involved in the biosynthesis of cuticle components (17, 18). However, the molecular mechanism by which lncRNAs regulate cuticle biosynthesis in *S. thurberi* fruit has not been elucidated yet. In this study, RNA-seq data were obtained for the *de novo* assembly and characterization of the *S. thurberi* fruit peel transcriptome. Coding and non-coding transcripts were identified, and the tentative function of protein-coding transcripts was predicted by homology searches.

It has been shown that the initial screening of stably expressed transcripts through RNA-seq helps to identify highly stable expressed transcripts by qRT-PCR (19). Because there are no reported experiments in which the gene expression analysis has been carried out for sweet pitaya, we use the RNA-seq data generated in this work along with qRT-PCR data analysis to identify reliable expressed reference genes.

As a first approach, three transcripts tentatively involved in cuticle biosynthesis were identified and used to evaluate the reliability of the best reference genes identified. The understanding of the role of the mechanism of cuticle biosynthesis in the adaptation of cactus fruits to arid environments could help to design new postharvest technologies to assure fleshy fruit production in the context of the increase in mean annual temperatures and water scarcity for agriculture predicted for the following years (1).

Results

Shorts reads and assembly quality

Table 1 shows the different quality variables of the *S. thurberi* fruit peel transcriptome. A total of 288,199,704 reads with 150 base pairs (bp) in length in paired-end mode were sequenced with the Illumina NextSeq500 platform at the Arizona Genetic Core facility of the University of Arizona at Tucson, AZ, USA. 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. BUSCO score showed that 85.4% are completed transcripts, although out of these, 37.2% were found to be duplicated.

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

Metric	Data
Total transcripts	174449
N50	2110
Smallest transcript length (bp)	200
Largest transcript length (bp)	19114
Mean transcript length (bp)	1198.7
GC (%)	41.33
Total assembled bases	209110524
TransRate score	0.05
BUSCO score (%)	C: 85.4% (S:48.2%, D:37.2%) F: 10.7% M: 3.9%

Values were calculated through the TrinityStats function of Trinity (11) and TransRate software (24). Completeness analysis was carried out through BUSCO (25) by aligning the transcriptome to the *Embryophyte* database through BLAST with an E value threshold of 1×10^{-3} . Abbreviations: Complete (C), single (S), duplicated (D), fragmented (F), missing (M).

Homology searches

A summary of the homology search in the main public protein database for the *S. thurberi* transcriptome is shown in Table 2. From these databases, the higher homologous transcripts were found in RefSeq (26) with 93,993 (53.87%). Based on the E value distribution, 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 (27) and RefSeq databases, respectively (**Fig. 1a-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 (**Fig. 1c-d**).

Table 2
Summary of homology search for *S. thurberi* transcripts in different databases.

Category	Number of transcripts (%)
Total transcripts	174449 (100%)
With a hit to RefSeq	93993 (53.87%)
With a hit to SwissProt	70381 (40.34%)
With a hit to nr-NCBI	72162 (41.37%)
With a hit to plantTFDB	45970 (26.35%)
With a hit to iTAK-TR	58704 (33.65%)
With a hit to iTAK-PK	48186 (27.65%)
With a hit to TAIR (<i>Arabidopsis thaliana</i>)	81556 (46.75%)
With a hit to ITAG (<i>Solanum lycopersicum</i>)	85331 (48,91%)
With a hit to any of these databases	98611 (56.52%)
Without a hit to any of these databases	75838 (43.47%)

Homologous sequences were predicted by an alignment through BLASTx (28) to the protein databases listed in the table with an E value threshold of $< 1 \times 10^{-10}$ for the nr-NCBI database and an E value threshold of $< 1 \times 10^{-5}$ for the others.

Figure 1 Homology analysis of assembled transcripts. E value distribution (**a, b**) and identity distribution (**c, d**) of the matches in the Swiss-Prot (**a, c**) and RefSeq (**b, d**) databases. (**a,b**) The number inside the pie chart indicates the number of transcripts recorded using that E value. Alignment by BLASTx with an E value threshold of 1×10^{-5} .

In **Additional file 1**, it is shown the distribution of the top hits found in different species, which showed that the highest number of matches in the Swiss-Prot database were for proteins from *Arabidopsis* (80.2%). For the RefSeq database, the highest number of matches were for proteins from *Beta vulgaris* (62%). In **Fig. 2**, it is shown 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 (*Persea americana*), peach (*Prunus persica*), strawberry (*Fragaria vesca*), orange (*Citrus sinensis*), and grapefruit (*Vitis 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. 2a**).

Figure 2 Venn diagram of the homology search results against model plants databases, commercial fruits, and cacti. The number in the diagram corresponds to the number of transcripts from *S. thurberi* homologous to sequences from that plant specie. **(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 homologous to transcripts or proteins from the cactus dragon fruit (*Hylocereus polyrhizus*), prickly pear cactus (*Opuntia streptacantha*), Mexican giant cardon (*Pachycereus pringlei*), and pitahaya (*Selenicereus undatus*) ranged from 76,238 (43.70%) to 114,933 (65.88%), with 64,009 transcripts homologous to all the four cactus databases (**Fig. 2b**). Further, out of the total of transcripts, 44,040 transcripts (25.25%) showed homology only to sequences from cactus, but not for model plants or commercial fruits (**Fig. 2c**).

A total of 69,622 (39.91%) transcripts showed homology to transcription factors, protein kinases, and other transcriptional regulators in the PlantTFDB and iTAK databases (Table 2). In the PlantTFDB, 45,970 (26.35%) transcripts homologous to transcriptional factors (TF) from 57 families were identified (**Fig. 3**). From these, the most frequent TF families were the basic-helix-loop-helix (bHLH), myeloblastosis-related (MYB-related), NAM, ATAF, and CUC (NAC), ethylene responsive factor (ERF), and the WRKY domain family (WRKY) (**Fig. 3**).

Figure 3 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.

Functional categorization

In **Additional file 2**, it is shown the results of functional annotations by the Blast2GO (29) suite pipeline. 72,162 transcripts showed homology to proteins in the nr-NCBI database (Table 2), and 83,066 transcripts showed to have at least a functional domain in the InterPro database. This data allowed the assignment of Gene Ontology (GO) terms to 68,559 transcripts (39.3% of the total) (See **Additional file 2**).

Figure 4 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 the higher number of transcripts. Further, for MF, organic cyclic compound binding, heterocyclic compound binding, and ion binding were the processes with the higher number of transcripts.

In **Fig. 5**, it is shown the KEGG pathways with the higher number of transcripts involved. However, *S. thurberi* transcripts were classified into 142 KEGG pathways. 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.

Figure 4 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.

Figure 5 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.

Identification of lncRNAs

Out of the total of transcripts, 122,234 (70.07%) were identified as coding transcripts, and 43,391 (24.87%) were classified as lncRNA. In **Fig. 6**, it is shown a comparison of the length (**Fig. 6a**) and expression (**Fig. 6b**) of lncRNA and coding RNA (cRNA). Besides, this graph includes the homology search results among lncRNA and cRNA against the transcriptomes of the cactus *O. streptacantha*, *H. polyrhizus*, *P. pringlei*, and *S. undatus* (**Fig. 6c**).

Both length and expression values were higher in cRNA than in lncRNA. In general, cRNA ranged from 201 to 18,629 bp with a mean length of 1,507.18, whereas lncRNA ranged from 200 to 5198 bp with a mean length of 481.51 (**Fig. 6a**). The higher expression values recorded from cRNA and lncRNA were 12.83 and 9.45 log₂(CPM), respectively (**Fig. 6b**). Besides, the Venn diagram shows that 18,727 lncRNA were homologous to transcripts reported from *H. polyrhizus*, *P. pringlei*, and *S. undatus* (**Fig. 6c**). Furthermore, 111,554 transcripts were found to be homologous between cRNA from *S. thurberi* and the transcriptomes of the different cactus species included. Also, 10,680 transcripts were found to be present in *S. thurberi* only (**Fig. 6c**). The alignment to the cactus transcriptomes found that lncRNA from *S. thurberi* showed lower identity (**Fig. 6d**), higher E values (**Fig. 6e**), and lower coverage (**Fig. 6f**) than the cRNAs.

Figure 6 Comparison of coding RNA (cRNA) and lncRNAs from *S. thurberi* transcriptome and among cactus species. (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 ($\log_2(\text{CPM})$) 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, and the higher and lower boxes limits represent the 25th-75th percentiles. **(c)** Venn diagram of the homology search results against the proteins of *Opuntia streptacantha* and transcripts of *Selenicereus undatus*, *Hylocereus polyrhizus*, and *Pachycereus pringlei*. **(d, e, f)** Density plots of the identity **(d)**, E value **(e)**, and coverage **(f)** for the hits recorded from the alignment of the cRNA and lncRNA from *S. thurberi* against transcripts of *S. undatus*, *H. polyrhizus*, and *P. pringlei*. **(e)** The X-axis indicates the negative exponential of the E values ($1e^{-X}$). As the highest negative exponential value recorded was 180, a value of $1e^{-200}$ was assigned to all the E values of 0.0 to represent the lowest E values. Homologous searching was carried out by BLAST alignment using an E value of less than 1×10^{-5} . The box plots, the density plots, and the Venn diagram were drawn by ggplot2 and ggVennDiagram in R Studio.

Identification of tentative reference genes (dup: abstract ?)

Mean transcripts per million of reads (TPM) and coefficient of variation (CV) were calculated for 4,980 not differentially expressed (NDE) transcripts in *S. thurberi* fruit peel transcriptome ($\log_2\text{FC}$ between + 1.0 and - 1.0; $\text{FDR} < 0.05$). Transcripts with a CV value lower than 0.11, corresponding with the percentile five of the CV, and a mean transcript *per* million higher than 1,281.7, corresponding with the percentile 95 of mean TPM, were used as filters to identify stably expressed transcripts. From these, 5 transcripts were selected as tentative reference genes. Besides, three NDE transcripts homologous to previously identified stable expressed reference genes in other species of cactus fruit (32–34) were selected. Homology metrics for the 8 tentative reference genes selected are shown in **Additional file 3**. Primer sequences used to quantify the transcripts by qRT-PCR during pitaya fruit development are shown in **Additional file 4**.

Expression stability of tentative reference genes

In **Additional file 5**, it is shown the melting curves of the eight candidate reference genes in which it is shown the amplification specificity. On the other side, for the eight tentative reference transcripts, cycle threshold (Ct) values obtained by qRT-PCR during sweet pitaya fruit development ranged from 16.85 to 30.26 (**Fig. 7a**). Plastidic ATP/ADP-transporter (StTLC1) showed the highest Ct values with a mean of 27.34. Polyubiquitin 3 (StUBQ3) showed the lowest Ct values in all five sweet pitaya fruit developmental stages (**Fig. 7a**).

Figure 7 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 days after flowering). The black line inside the box indicates the median, and the higher and lower boxes limits represent the 25th -75th percentiles. **(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 V_n/V_{n+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), α -tubulin (StTUA), elongation factor 1- α (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).

NormFinder (22) calculated the stability value for all eight transcripts, ranging from 0.45 to 1.27 (See **Additional file 6**). The lowest stability values were 0.45, 0.51, 0.97, and 0.99, corresponding to the transcripts elongation factor 1- α (StEF1a), α -tubulin (StTUA), plastidic ATP/ADP-transporter (StTLC1), and actin 7 (StACT7), respectively. For BestKeeper (21), the most stable expressed transcripts were polyubiquitin 3 (StUBQ3), α -tubulin (StTUA), and elongation factor 1- α (StEF1a), with values of 0.72, 0.75, and 0.87, respectively. In the case of the delta Ct method (35), the transcripts elongation factor 1- α (StEF1a), α -tubulin (StTUA), and plastidic ATP/ADP-transporter (StTLC1) showed the best stability values (See **Additional file 6**).

According to geNorm (20) analysis, the most stable expressed transcripts were α -tubulin (StTUA), elongation factor 1- α (StEF1a), polyubiquitin 3 (StUBQ3), and Actin 7 (StACT7), with values of 0.74, 0.74, 0.82, 0.96, respectively. All the pairwise variation values (V_n/V_{n+1}) were lower than 0.15, ranging from 0.019 for V_2/V_3 to 0.01 for V_6/V_7 (**Fig. 7c**). The V value of 0.019 recorded for V_2/V_3 indicates that the use of the best two reference genes StTUA and StEF1 is reliable enough for the accurate normalization of qRT-PCR data. No third reference genes are required (20). With the exception of BestKeeper analysis, StEF1a and StTUA were the most stable transcripts for all of the stability analysis methods carried out in this study (See **Additional file 6**). The comprehensive ranking analysis indicates that StEF1, 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. 7b**).

Identification of cuticle biosynthesis-related transcripts

Three cuticle biosynthesis-related transcripts DN17030_c0_g1_i2, DN15394_c0_g1_i1, and 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 family G member 11 transporter (ABCG11/WBC11), respectively, were identified and quantified. The best homology match for StCYP77A (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 (See **Additional file 3**).

TransDecoder, InterPro (36), and TMHMM (37) 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 *Carnegieia gigantean* (See **Additional file 7**).

StGDSL1 (DN15394_c0_g1_i1) alignment showed that it is homologous to a GDSL esterase/lipase from Arabidopsis (Q9LU14) and tomato (Solyc03g121180) (See **Additional file 3**). TransDecoder, InterPro, and SignalP (38) analysis showed that StGDSL 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 (See **Additional file 8**).

In **Fig. 8**, it is shown the bioinformatic analysis carried out in the *in silico* protein predicted StABCG11. The phylogenetic tree constructed shows three clades corresponding to the ABCG13, ABCG12, and ABCG11 protein classes with bootstrap support ranging from 40 to 100% (**Fig. 8a**). 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 (**Fig. 8b**).

The predicted protein sequence of StABCG11 (DN23528_c1_g1_i1) 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 G family. Multiple sequence alignment shows that the Walker A and B motif sequence and the ABC signature (39) are conserved between the ABCG11 transporters from Arabidopsis, tomato, *S. thurberi*, and *C. gigantean* (**Fig. 8c**).

Figure 8 Analysis of the predicted protein StABCG11 from *Stenocereus thurberi*. **(a)** Phylogenetic tree of StABCG11 and related proteins of the classes ABCG11, ABCG12, and ABCG13 from *Arabidopsis thaliana* (At), *Gossypium arboreum* (Ga), *Citrus sinensis* (Cs), *Medicago truncatula* (Mt), *Solanum lycopersicum* (Sl), *Eutrema halophilum* (Eh), *Carnegieia gigantean* (Cg), *Beta vulgaris* (Bv), and *Spinacia oleracea* (So). The database accession number next to the protein name is shown. The scale bar of 0.10 represented a sequence divergence of 10%. The number in the branches is the percentage bootstrap value of 1000 replicates. The highest percentages represent higher significant results. The black square beside the protein name shows AtABCG11, AtABCG12, and AtABCG13 from *A. thaliana*. The red circle and red triangle next to the protein name show StABCG11 from *S. thurberi* and a protein from the closest related species, *C. gigantean*, respectively. Neighbor-joining (NJ) phylogenetic tree constructed by MEGA11 software. **(b)** The predicted transmembrane helices of StABCG11. The probability of membrane insertion (Y-axis) and transmembrane region represented by purple color was determined by TMHMM software. **(c)** Multiple sequence alignment of StABCG11 and its homologous from *A. thaliana* (AT1G17840), *S. lycopersicum* (Solyc03g019760), and *C. gigantean* (KAJ8441854). Amino acids are colored according to the chemistry classification of their side-chain. The darkest blue bars below the protein sequences indicate 100% conservation. Black rectangles show the conserved sequence of the Walker A and B motif

and the ABC signature, named below the rectangles. Black width lines below the sequence show the predicted transmembrane helices of StABCG11. Alignment was carried out by MUSCLE in MEGA11 and drawn by ggmsa in R Studio.

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

According to the results of the expression stability analysis (**Fig. 7**), four normalization strategies were tested to quantify 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 (DN17030_c0_g1_i2), StGDSL1 (DN15394_c0_g1_i1), and ABCG11 (DN23528_c1_g1_i1) by qRT-PCR during sweet pitaya fruit development are shown in **Additional file 4**. Melting curves showing the amplification specificity of StCYP77A, StGDSL1, and ABCG11 are shown in **Additional file 5**.

The three cuticle biosynthesis-related transcripts showed differences in expression during sweet pitaya fruit development (See **Additional file 9**). The same expression pattern was recorded for the three cuticle biosynthesis transcripts when normalization was carried out by StEF1a, StTUA, StUBQ3, or StEF1 + StTUA (**Fig. 9**). A higher expression of StCYP77A and StGDSL1 is shown at the 10 and 20 days after flowering (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. 9**).

Figure 9 Relative expression of cutin biosynthesis-related transcripts StCYP77A, StGDSL1, and StABCG11 during sweet pitaya fruit development. Relative expression was calculated through the $2^{-\Delta\Delta Ct}$ method using elongation factor 1-alpha (StEF1a), α -tubulin (StTUA), polyubiquitin 3 (StUBQ3), or StEF1a + StTUA as normalizing genes at 10, 20, 30, 35, and 40 days after flowering (DAF). The Y-axis and error bars represent the mean of the relative expression $\pm$ 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 family G member 11 (StABCG11).

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 2000 bp, a high percentage of conserved transcripts completely assembled ($> 80\%$), and a high proportion of reads mapping back to the assembled transcripts (40). In the present study, we report the first collection of

174,449 transcripts from *S. thurberi* fruit peel. 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 (41).

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 to predict the sequence of almost all the transcripts expressed in sweet pitaya fruit peel (11). The number of transcripts assembled and the percentage of complete duplicated orthologous (37.2%) found by BUSCO suggest a significant redundancy (25). This redundancy is most likely because RNA isolated from the fruit exocarp for sequencing was obtained from fruits with close stages of development. Furthermore, four libraries were created, generating 21.95 Gb of information, corresponding to 288,199,704 short reads with 150 bp in length.

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

The percentage of transcripts homologous found, E values, and identity distribution (Table 2, **Fig. 1**) were similar to that reported in the *de novo* transcriptome assembly for non-model plants and other cactus fruits (32–34, 42) and further suggests that the transcriptome assembled of *S. thurberi* peel is robust (40). Swiss-Prot is a highly curated protein sequence database with a high level of annotation (27). However, this database has scarce information on protein sequences from cactus species. The *A. thaliana* genome dataset has been rigorously annotated, and it is one of the most comprehensive datasets (43). That explains why, at the level of species distribution found in the Swiss-Prot database, the highest number of matches (80.2%) were for Arabidopsis proteins (See **Additional file 1**). The advantage of these results is because they provide valuable information to make inferences about gene functions for *S. thurberi* through homology searches.

In the RefSeq database, the highest number of matches (62%) were for homologous proteins from *Beta vulgaris* (See **Additional file 1**), which is one of the closer phylogenetic-related species to *S. thurberi* with more information about sequences published in RefSeq. *B. vulgaris* and *S. thurberi* belong to the order *Caryophyllales*, which includes plants characterized for being adapted to survive in harsh environmental conditions and synthesize betalains, a group of antioxidant and nutraceutical compounds only synthesized by plants of this order (42, 44). The transcripts dataset generated in the present study provides helpful genetic information to study the metabolic pathway of betalains biosynthesis in *S. thurberi* fruit.

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 (**Fig. 2c**). This could suggest that exists a significant conservation of sequences among plants of the *Cactaceae* family which most likely are having 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, "glycerophospholipid metabolism", "glycerolipid metabolism", and "cutin, suberine, and wax biosynthesis" suggest an active cuticle biosynthesis in pitaya fruit peel (**Fig. 5**). In agreement with mentioned above, one of the main GO terms assigned for the molecular function (MF) category were "organic cyclic compound binding", "oxidoreductase activity", "transmembrane transporter activity", and "lipid binding" (**Fig. 4**). For the biological processes (BP) category, the critical GO terms for our 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 from the stress.

The most frequent transcription factors (TF) families found in *S. thurberi* transcriptome were NAC, WRKY, bHLH, ERF, and MYB-related (**Fig. 3**), which had been reported that play a function in the tolerance to abiotic stress in plants (45, 46). Although the role of NAC, WRKY, bHLH, ERF, and MYB TF in improving drought tolerance in relevant crop plants has been widely documented (46–48), their contribution to the adaptation of cactus to arid ecosystems has not been clearly elucidated yet and more 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 (49). In tomato fruits, the gene *SIMIXTA-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 (50). Despite the relevant role of cuticles in maintaining cactus physiology in desert environments, experimental evidences showing the role of the different TF inducing cuticle biosynthesis has yet to be reported for cacti fruits.

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

According to the different criteria carried out in this study, out of the 174,449 transcripts found, 43,391 transcripts with a mean length of 481 bp were classified as lncRNAs. Further, this is the first report of lncRNA identification for the specie *S. thurberi*. In fruits, 3,679 lncRNA has been identified from tomato (13), 3,330 from peach (*P. persica*) (14), 3,857 from melon (*Cucumis melo*) (15), 2,505 from hot pepper (*Capsicum annuum*) (16), 3,194 from pomegranate (*Punica granatum*) (18), and 5,884 from strawberry (*F. vesca*) (51). Despite the stringent criteria to classify the lncRNA of sweet pitaya fruit (*S. thurberi*), a higher number of lncRNAs is shown when compared with previous reports. This finding is most likely due to the higher level of redundancy found during the transcriptome analysis and the need for the complete assembled genome of the specie *S. thurberi*.

Previous studies showed that lncRNAs are shorter and have lower expression levels than coding RNA (17, 52–54). In agreement with those findings, both length and expression values of lncRNA from *S. thurberi* were lower than coding RNAs (**Fig. 6**). It has been suggested that lncRNAs could be involved in the biosynthesis of cuticle components in cabbage (17) and pomegranate (18) and that they could be involved in the tolerance to water deficit and cold stress through the regulation of cuticle biosynthesis in wild banana (52). Nevertheless, the molecular mechanism by which lncRNAs may regulate the cuticle biosynthesis in *S. thurberi* fruits has not been clearly elucidated yet.

Interestingly, BLASTn alignment (E-value < 1×10^{-5}) showed that 18,727 lncRNAs from *S. thurberi* were homologous to transcripts reported from the cactus *H. polyrhizus*, *P. pringlei*, and *S. undatus* (**Fig. 6**), which could indicate a possible lncRNA sequence conservation between cactus species. However, despite the efforts that have been carried out to find conserved lncRNAs between plant species, the results show that lncRNA sequences are poorly conserved (51, 52). It has been suggested that this is because the lncRNA function is not carried out by codon usage but by target binding through conserved secondary structures. Although lncRNA may possess short conserved sequences, these are not well identified by BLAST alignment (51). In agreement with that, alignment results showed that the homology to the cactus transcriptomes of lncRNAs from *S. thurberi* is less significant than that for coding RNA (**Fig. 6d, 6e, 6f**).

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 are genes that 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 (55). Nevertheless, their expression could change depending on plant species, developmental stages, and experimental conditions (19). Reliable reference genes for a specific experiment in a given specie must be identified to carry out an accurate qRT-PCR data normalization (55). An initial screening of the transcript expression pattern through RNA-seq improves the identification of reliably expressed transcripts by qRT-PCR (19, 56).

Identification of stable expressed reference transcripts during fruit development has been carried out in blueberry (*Vaccinium bracteatum*) (57), kiwifruit (*Actinidia chinensis*) (58), peach (*P. persica*) (59), apple (*Malus domestica*) (60), orange (*C. sinensis*) (61), and soursop (*Annona muricata*) (62). These studies include the expression stability analysis through geNorm, NormFinder, and BestKeeper algorithms (60–62), some of which are supported in RNAseq data (57, 58). 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*) (56), pear (*Pyrus pyrifolia* and *P. calleryana*) (19), and pepper (*C. annuum*) (63).

For fruits of the *Cactaceae* family, only a few studies identifying reliable reference genes have been reported (32–34). Mainly because gene expression analysis has not been carried out previously for sweet

pitaya (*S. thurberi*), we use the RNA-seq data generated in this work along with geNorm, NormFinder, BestKeeper, and RefFinder algorithms 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. 7b**). All the pairwise variation values (V_n/V_{n+1}) were lower than 0.15 (**Fig. 7c**), 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 in the specie *S. thurberi*.

The genes StEF1a, StTUA, and StUBQ3 are homologous to transcripts found in the cactus specie known as dragon fruit (*Hylocereus monacanthus* and *H. undatus*). Further, they putatively code for proteins involved in essential cellular functions, and they have been reported as reference genes to carry out plant gene expression analysis in the species mentioned (32). It has been reported that EF1a could be used as a reliable reference gene during the analysis of changes in gene expression of dragon fruit (*H. monacanthus* and *H. undatus*) (32), peach (*P. persica*) (59), apple (*M. domestica*) (60), and soursop (*A. muricata*) (62). For watermelon (*C. lanatus*) development, the combination of alpha-tubulin 5 (CITUA5), clathrin adaptor complex subunit (CICAC), and beta-actin (CIAC) was required for reliable gene expression normalization (64).

According to our results of the expression stability analysis (**Fig. 7**), StEF1a, StTUA, and StUBQ3 were selected, and four normalization strategies were designed. The same gene expression pattern was recorded for the three target transcripts evaluated when normalization was carried out by StEF1a, StTUA, StUBQ3, or StEF1 + StTUA (**Fig. 9**). Further, these data indicates that these reference genes are reliable enough to be used in qRT-PCR experiments during fruit development of *S. thurberi* and perhaps in experiments to study gene expression in fruits of other related species.

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 long-chain fatty acids, and the cuticular wax, composed mainly of very long-chain (VLC) fatty acids, and their derivatives VLC alkanes, VLC primary alcohols, VLC ketones, VLC aldehydes, and VLC esters (3). In *Arabidopsis*, cytochrome p450 family 77 subfamily A polypeptide 4 and 6 (CYP77A4 and CYP77A6) catalyze the synthesis of midchain epoxy and hydroxy ω -OH long-chain fatty acids, respectively (65, 66), which are one of the main components of fleshy fruit cuticle (3).

The length of StCYP77A (518 aa) is similar to its homologous (AT5G04660 and P37124) and includes a cytochrome P450 E-class domain. The predicted membrane-spanning region within the StCYP77A structure (See **Additional file 7**) has been previously characterized in *A. thaliana* and *Brassica napus*. It suggests that it could catalyze the oxidation of fatty acids embedded in the endoplasmic reticulum membrane (66, 67).

Phylogenetic analysis showed that StCYP77A was closer to proteins from its phylogenetic-related species *B. vulgaris* (BvCYP772; XP_010694692) and *Carnegieia gigantea* (Cgig2_012892) (See **Additional file 7**). StCYP77A, BvCYP77A2, and Cgig2_012892 were closer to SlCYP77A2 and SmCYP77A2 than to CYP77A4 and CYP77A6 proteins, suggesting that StCYP77A (DN17030_c0_g1_i2) could correspond to a CYP77A2 protein.

Five CYP77A are present in the Arabidopsis genome (CYP77A4-7 and CYP77A9), but their role in cuticle biosynthesis has only been reported for CYP77A4 and CYP77A6 (68). It has been suggested that CYP77A2 from eggplant (*Solanum torvum*) could contribute to the defense against fungal phytopathogen infection by the synthesis of specific compounds (69). 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 (70).

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 AT3G16370 and Solyc03g121180 (See **Additional file 3**). Findings suggest that a GDSL1 protein (CD1) polymerizes midchain oxygenated ω -OH long-chain fatty acids to form the cutin polyester at the extracellular space of tomato fruit peel (71, 72). It has been suggested that the 25-amino acid N-signal peptide found in StGDSL1 (See **Additional file 8**), previously reported in GDSL1 from Arabidopsis, *B. napus*, and tomato, plays a role during the protein exportation to the extracellular space (71, 73).

In agreement with the changes in expression of CYP77A2 (A0A1U8GYB0) and ABCG11 (LOC107862760) reported during pepper fruit development (70), a higher expression of StCYP77A and StGDSL1 is shown at the 10 and 20 DAF of sweet pitaya fruit, with a reduction at 30, 35 and 40 DAF (**Fig. 9**). Also, StABCG11 showed similar behavior, with a higher expression at 10 and 20 DAF than at 30 and 35 DAF (**Fig. 9**), suggesting the active cuticle biosynthesis at the early stages of sweet pitaya fruit development. In agreement with that, GDSL lipase/hydrolase proteins SGN-U583101 and SGN-U579520 from tomato were highly expressed in the early stages and during the expansion stages of tomato fruit development, respectively, with a decrease in later stages (74). It has been suggested that the expression of GDSL proteins increases in expanding organs to accelerate cuticle polymerization during growth (72).

The expression of CYP77A6, CYP77A4, GDSL esterase/lipase (CUS1), and ABCG11 contribute to the high-cutin phenotype of pepper fruits (*Capsicum chinense*) (75). The gene cutin synthase from mango (MiCUS1), a homologous to the gene cutin deficient (CD1) from tomato, shows a higher expression at early stages of fruit development, decreases in intermediate stages, and increases again during ripening (7).

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

The phylogenetic analysis, the functional domains, and the six transmembrane helices found in the StABCG11 predicted protein (**Fig. 8**), which is a putative ABCG plasma membrane transporter of sweet

pitaya, suggest that it could play a role as a cuticle compound transporter in the cellular membrane of sweet pitaya peel (39). Indeed, an increased expression of StABCG11 at 40 DAF was recorded in the present study (**Fig. 9**). Further, this data strongly suggests that it could be playing a relevant role in the transport of cuticle components at the beginning and during later stages of sweet pitaya fruit ripening.

AtABCG11 (AtWBC11) is an extracellular transporter of cuticular wax and cutin compounds from Arabidopsis localized at the plasma membrane (39, 76, 77). Further, it has been reported that the overexpression of ABCG11 induces drought tolerance in plants, tentatively through the increase in the transport of cuticle compounds (78, 79). Also, it had been reported that a high expression of the ABC plasma membrane transporter from mango (MiWBC11) during mango fruit ripening, which is homologous to the AtABCG11 transporter from Arabidopsis just mentioned. MiWBC11 expression correlates with a higher cuticle deposition and suggests an essential role in the transport of cuticle compounds during mango fruit ripening (7).

The expression pattern showed for StABCG11, StCYP77A, and StGDLS1 suggests a role of StABCG11 as a cutin compound transporter in the earlier stages of sweet pitaya fruit development (**Fig. 9**). 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 StGDLS1 in the later stages.

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 (80). PaWCB11 expression correlated with cuticle membrane deposition at the earlier and intermediate stages of sweet cherry fruit development but not at the later (80). The increased expression of StABCG11 found in this study could be due to the increased transport of cuticular wax compounds, such as VLC fatty acids and their derivatives, in the later stages of sweet pitaya development.

Findings suggest that *S/MIXTA-like*, which codes for a MYB transcription factor, reduces water loss in ripe tomato fruits by increasing the expression of CYP77A2, GDLS esterase/lipase, and ABCG11 (50). The gene expression pattern found in this study could be due to the growth rate at the early development stages of sweet pitaya fruit and the high water content increase at the ripening. On the other hand, it has been suggested that cuticular waxes more than cutin compounds contribute to avoid water loss in fruits (81–83) and abiotic stress tolerance in plants (84). Further expression analysis of cuticular wax biosynthesis genes, complemented with metabolomic analysis, can contribute to a more holistic understanding of the role of cuticles in the adaptation of cactus fruits to abiotic stress in the desert ecosystem. Besides, co-expression analysis of the lncRNAs found in this study with coding transcripts involved in cuticle biosynthesis can lead to the identification of lncRNAs regulating genes playing a role in the cuticle biosynthesis of *S. thurberi* specie.

Conclusion

In this study, the transcriptome of the sweet pitaya (*S. thurberi*) fruit peel was assembled for the first time. The reference genes found are a helpful tool for gene expression analysis in sweet pitaya fruit and fruits from other cactus species.

Transcripts tentatively involved in cuticle compound biosynthesis and transport are reported for the first time in sweet pitaya. Results suggest a relevant role of cutin compound biosynthesis and transport at the early and later stages of fruit ripening development. The lncRNAs identified in this study will allow to carry out further studies to elucidate their role in regulating the genes involved in the molecular mechanism of cuticle biosynthesis.

The information generated will help to improve the elucidation of the molecular mechanism of cuticle biosynthesis and other metabolic pathways 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 global temperature increase and the increase in water scarcity for agricultural production predicted for the following years.

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. Peels from fruits were pooled, frozen in liquid nitrogen, and pulverized. Total RNA was isolated from the peels through the Hot Borate method (85). 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 Genetic Core Facility. A total of 288,199,704 short reads with a length of 150 bp were obtained.

De novo transcriptome assembly and quality analysis

FastQC software was used for short reads quality analysis. Short reads with poor quality were trimmed or eliminated by Trimmomatic (86) 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 25 quality score on the Phred scale were used to carry out the *de novo* assembly by Trinity (11) 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. With the goal of removing redundancy, transcripts with equal or more than 90% of identity were merged through CD-hit (87). Alignment and quantification in terms of Transcripts per Million (TPM) were carried

out through Bowtie (88) and RSEM (89), respectively. Transcripts showing a low expression ($\text{TPM} < 0.01$) 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 (24). The transcriptome completeness was determined through a BLASTn (28) alignment ($E \text{ value} < 1 \times 10^{-3}$) by BUSCO against the database of conserved orthologous genes from *Embryophyte* (25).

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 BLASTx algorithm with an E value threshold of 1×10^{-5} (26, 27, 43, 90). An additional alignment was carried out to the protein databases of commercial fruits (*P. Americana*, *P. persica*, *F. Ananassa*, *C. cinensis*, and *V. vinifera*) to proteins of the cactus *O. streptacantha* (BioProject: PRJNA320545), and the transcriptomes of the cactus *H. polyrhizus* (PRJNA490886), *P. pringlei* (PRJNA413949), and *S. undatus* (PRJNA515117). The open reading frame (ORF) of the transcripts was predicted by TransDecoder, considering a minimal ORF length of 75 amino acids (aa) (11). The search for protein domains was carried out by InterPro (36) and the Pfam (91) databases. Functional categorization was carried out by Blast2GO (29) based on GO terms and KEGG metabolic pathways.

Identification of long non-coding transcripts

LncRNAs were identified based on previous studies (12–18). Transcripts without homology (E value threshold of 1×10^{-5}) to any protein from Swiss-Prot, RefSeq, nr-NCBI, PlantTFDB, iTAK, TAIR, ITAG, *P. Americana*, *P. persica*, *F. Ananassa*, *C. cinensis*, *V. vinifera*, and *O. Streptacantha* databases, without a predicted ORF longer than 75 aa, and without protein domains in the InterPro and Pfam databases were selected to identify tentative lncRNAs.

Transcripts coding for signal peptide or transmembrane helices were identified by SignalP (38) and TMHMM (37), 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 (92) and discarded. The remaining transcripts were analyzed by CPC (30) and CPC2 (31) 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 lncRNAs.

Identification of tentative reference genes

Aligning short reads, quantifying transcripts based on TPM, and differential expression analysis were done through Bowtie, RSEM, and edgeR software, respectively (93). Transcripts with $\log_2$ Fold Change

(log₂FC) between + 1.0 and – 1.0 and with a False Discovery Rate (FDR) lower than 0.05 were taken as not differentially expressed (NDE). The NDE transcripts were aligned by BLASTn (E value < 1x10⁻⁵) to 43 constitutive genes previously reported from the cactus *H. polyrhizus*, *S. monacanthus*, and *S. undatus* (32–34) to identify possible homologous reference genes in *S. thurberi*. Based on previous studies (56), the percentile 5 value of the coefficient of variation (CV) and the percentile 95 value of the mean TPM were used to identify stably expressed transcripts.

Identification of reliable reference genes.

Open flowers were tagged, and fruits with 10, 20, 30, 35, and 40 Days After Flowering (DAF) were collected. The fruit harvesting was carried out as described above. One biological replicate consists of peels from three fruits collected from the same plant. Three biological replicates from three different plants were analyzed for each developmental stage. RNA extraction, quantification, RNA purity, and RNA integrity analysis were determined 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 (<http://www.idtdna.com>) and following the method proposed (94). 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 during 2 min at 95 °C, followed by 40 cycles of denaturation step at 95°C during 15 s, annealing and extension steps during 30 s at 60°C.

The Cycle threshold (Ct) values were analyzed through the algorithms BestKeeper, geNorm, NormFinder, and the delta Ct method (20–22, 35). RefFinder (23) was used to integrate the stability results and to find the most stable expressed transcripts in pitaya peel fruit during development. The pairwise variation value (Vn/Vn + 1) was calculated through the geNorm algorithm in R Studio software (20).

Quantification of cuticle biosynthesis-related transcripts

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 family G member 11 (ABCG11) were identified according to the functional annotation described above. The aa sequence coded from the transcripts was predicted by TransDecoder. 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 (95). MEGA11 constructed a neighbor-joining (NJ) phylogenetic tree with a bootstrap of 1000 replications (96).

Fruit sampling, primer design, RNA extraction, cDNA synthesis, and transcript quantification were performed as described above to evaluate reliable reference genes. Relative expression was calculated according to the 2^{-ΔΔCt} method (97). The sample corresponding to 10 DAF was used as the calibrator.

The transcripts StEF1a, StTUA, StUBQ3, and StEF1a + StTUA5 were used as normalizer genes to evaluate their reliability.

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 stats packages in R Studio.

Abbreviations

CV: Coefficient of variation. TPM: Transcripts per Million of reads. GC: guanine-cytosine. BUSCO: Benchmarking Universal Single-Copy Orthologs. BLAST: Basic Local Alignment Search Tool. NDE: Not differentially expressed. FDR: False Discovery Rate. Ct: Cycle threshold. qRT-PCR: quantitative reverse transcriptase-polymerase chain reaction.

Declarations

Acknowledgments

We appreciate the University of Arizona Genetic Core and Illumina for providing reagents and equipment for libraries sequencing. The author, Heriberto García Coronado, thanks the National Council of Humanities, Science and Technology (CONAHCyT, acronym in Spanish) for the Ph.D. scholarship assigned. The author, Heriberto García Coronado, thanks to Dr. Edmundo Domínguez-Rosas for the technical support in bioinformatics for the identification of long non-coding RNAs.

Authors' contributions

JT, MT, JB, MH, and HG designed the research. JT, MT, JB, and MH supervised the research. JT, MT, MH, and HG collected the plant materials. JT and MT participated in libraries preparation for sequencing. HG carried out the experiments and data analysis. MH and HG analyzed the transcriptome data by bioinformatics and discussed experiments. HG and MT wrote the paper. All authors read and approved the final manuscript.

Funding

This work was done with the financial support of the National Council of Science Humanities and Technology (CONAHCYT, an acronym in Spanish) of Mexico.

Availability of data and materials

The datasets generated and analyzed during the current study are not publicly available but are available from the corresponding author upon request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Hultine KR, Hernández-Hernández T, Williams DG, Albeke SE, Tran N, Puente R, et al. Global change impacts on cacti (*Cactaceae*): current threats, challenges and conservation solutions. *Ann Bot.* 2023. <https://doi.org/10.1093/aob/mcad040>.
2. Liu L, Wang X, Chang C. Toward a smart skin: Harnessing cuticle biosynthesis for crop adaptation to drought, salinity, temperature, and ultraviolet stress. *Front Plant Sci.* 2022. <https://doi.org/10.3389/fpls.2022.961829>.
3. García-Coronado H, Tafolla-Arellano JC, Hernández-Oñate MÁ, Burgara-Estrella AJ, Robles-Parra JM, Tiznado-Hernández ME. Molecular Biology, Composition and Physiological Functions of Cuticle Lipids in Fleshy Fruits. *Plants.* 2022. <https://doi.org/10.3390/plants11091133>.
4. Matas AJ, Yeats TH, Buda GJ, Zheng Y, Chatterjee S, Tohge T, et al. Tissue- and cell-type specific transcriptome profiling of expanding tomato fruit provides insights into metabolic and regulatory specialization and cuticle formation. *Plant Cell.* 2011. <https://doi.org/10.1105/tpc.111.091173>.
5. Albert Z, Ivanics B, Molnár A, Miskó A, Tóth M, Papp I. Candidate genes of cuticle formation show characteristic expression in the fruit skin of apple. *Plant Growth Regul.* 2013. <https://doi.org/10.1007/s10725-012-9779-y>.
6. Alkio M, Jonas U, Declercq M, Van Nocker S, Knoche M. Transcriptional dynamics of the developing sweet cherry (*Prunus avium* L.) fruit: sequencing, annotation and expression profiling of exocarp-associated genes. *Hortic Res.* 2014. <https://doi.org/10.1038/hortres.2014.11>.
7. Tafolla-Arellano JC, Zheng Y, Sun H, Jiao C, Ruiz-May E, Hernández-Oñate MA, et al. Transcriptome Analysis of Mango (*Mangifera indica* L.) Fruit Epidermal Peel to Identify Putative Cuticle-Associated Genes. *Sci Rep.* 2017. <https://doi.org/10.1038/srep46163>.
8. Wu X, Shi X, Bai M, Chen Y, Li X, Qi K, et al. Transcriptomic and Gas Chromatography-Mass Spectrometry Metabolomic Profiling Analysis of the Epidermis Provides Insights into Cuticular Wax

- Regulation in Developing Yuluxiang Pear Fruit. J Agric Food Chem. 2019. <https://doi.org/10.1021/acs.jafc.9b01899>.
9. Castro-Enríquez DD, Montañó-Leyva B, Del Toro-Sánchez CL, Juárez-Onofre JE, Carvajal-Millán E, López-Ahumada GA, et al. Effect of ultrafiltration of Pitaya extract (*Stenocereus thurberi*) on its phytochemical content, antioxidant capacity, and UPLC-DAD-MS profile. Molecules. 2020. <https://doi.org/10.3390/molecules25020281>.
 10. García-Cruz L, Valle-Guadarrama S, Guerra-Ramírez D, Martínez-Damián MT, Zuleta-Prada H. Cultivation, quality attributes, postharvest behavior, bioactive compounds, and uses of *Stenocereus*: A review. Sci Hortic. 2022. <https://doi.org/10.1016/j.scienta.2022.111336>.
 11. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013. <https://doi.org/10.1038/nprot.2013.084>.
 12. Patra GK, Gupta D, Rout GR, Panda SK. Role of long non coding RNA in plants under abiotic and biotic stresses. Plant Physiol Biochem. 2023. <https://doi.org/10.1016/j.plaphy.2022.10.030>.
 13. Zhu B, Yang Y, Li R, Fu D, Wen L, Luo Y, et al. RNA sequencing and functional analysis implicate the regulatory role of long non-coding RNAs in tomato fruit ripening. J Exp Bot. 2015. <https://doi.org/10.1093/jxb/erv203>.
 14. Zhou H, Ren F, Wang X, Qiu K, Sheng Y, Xie Q, et al. Genome-wide identification and characterization of long noncoding RNAs during peach (*Prunus persica*) fruit development and ripening. Sci Rep. 2022. <https://doi.org/10.1038/s41598-022-15330-3>.
 15. Tian Y, Bai S, Dang Z, Hao J, Zhang J, Hasi A. Genome-wide identification and characterization of long non-coding RNAs involved in fruit ripening and the climacteric in *Cucumis melo*. BMC Plant Biol. 2019. <https://doi.org/10.1186/s12870-019-1942-4>.
 16. Ou L, Liu Z, Zhang Z, Wei G, Zhang Y, Kang L, et al. Noncoding and coding transcriptome analysis reveals the regulation roles of long noncoding RNAs in fruit development of hot pepper (*Capsicum annuum* L). Plant Growth Regul. 2017. <https://doi.org/10.1007/s10725-017-0290-3>.
 17. Zhu X, Tai X, Ren Y, Chen J, Bo T. Genome-wide analysis of coding and long non-coding RNAs involved in cuticular wax biosynthesis in cabbage (*Brassica oleracea* L. var. *capitata*). Int J Mol Sci. 2019. <https://doi.org/10.3390/ijms20112820>.
 18. Wang Y, Zhao Y, Wu Y, Zhao X, Hao Z, Luo H, et al. Transcriptional profiling of long non-coding RNAs regulating fruit cracking in *Punica granatum* L. under bagging. Front Plant Sci. 2022. <https://doi.org/10.3389/fpls.2022.943547>.
 19. Wang Y, Dai M, Cai D, Shi Z. Screening for quantitative real-time PCR reference genes with high stable expression using the mRNA-sequencing data for pear. Tree Genet Genomes. 2019. <https://doi.org/10.1007/s11295-019-1361-6>.
 20. Smith-Unna R, Boursnell C, Patro R, Hibberd JM, Kelly S, TransRate. Reference-free quality assessment of de novo transcriptome assemblies. Genome Res. 2016. 10.1101/gr.196469.115.

21. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015. <https://doi.org/10.1093/bioinformatics/btv351>.
22. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016. <https://doi.org/10.1093/nar/gkv1189>.
23. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res*. 2000. <https://doi.org/10.1093/nar/28.1.45>.
24. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
25. Conesa A, Götz S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics*. 2008. 10.1155/2008/619832.
26. Chen C, Wu J, Hua Q, Tel-Zur N, Xie F, Zhang Z, et al. Identification of reliable reference genes for quantitative real-time PCR normalization in pitaya. *Plant Methods*. 2019. <https://doi.org/10.1186/s13007-019-0455-3>.
27. Nong Q, Yang Y, Zhang M, Zhang M, Chen J, Jian S, et al. RNA-seq-based selection of reference genes for RT-qPCR analysis of pitaya. *FEBS Open Bio*. 2019. <https://doi.org/10.1002/2211-5463.12678>.
28. Zheng Q, Wang X, Qi Y, Ma Y. Selection and validation of reference genes for qRT-PCR analysis during fruit ripening of red pitaya (*Hylocereus polyrhizus*). *FEBS Open Bio*. 2021. <https://doi.org/10.1002/2211-5463.13053>.
29. Andersen CL, Jensen JL, Ørntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res*. 2004. <https://doi.org/10.1158/0008-5472.CAN-04-0496>.
30. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations. *Biotechnol Lett*. 2004. <https://doi.org/10.1023/B:BILE.0000019559.84305.47>.
31. Silver N, Best S, Jiang J, Thein SL. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Mol Biol*. 2006. <https://doi.org/10.1186/1471-2199-7-33>.
32. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*. 2002. <https://doi.org/10.1186/gb-2002-3-7-research0034>.
33. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro: The integrative protein signature database. *Nucleic Acids Res*. 2009. <https://doi.org/10.1093/nar/gkn785>. 37 Suppl. D211–D215.

34. Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol.* 2001. <https://doi.org/10.1006/jmbi.2000.4315>.
35. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol.* 2022. <https://doi.org/10.1038/s41587-021-01156-3>.
36. Luo B, Xue XY, Hu WL, Wang LJ, Chen XY. An ABC transporter gene of *Arabidopsis thaliana*, *AtWBC11*, is involved in cuticle development and prevention of organ fusion. *Plant Cell Physiol.* 2007. <https://doi.org/10.1093/pcp/pcm152>.
37. Raghavan V, Kraft L, Mesny F, Rigerte L. A simple guide to de novo transcriptome assembly and annotation. *Brief Bioinform.* 2022. <https://doi.org/10.1093/bib/bbab563>.
38. Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, et al. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature.* 2019. <https://doi.org/10.1038/s41586-019-1693-2>.
39. Xi X, Zong Y, Li S, Cao D, Sun X, Liu B. Transcriptome analysis clarified genes involved in betalain biosynthesis in the fruit of red pitayas (*Hylocereus costaricensis*). *Molecules.* 2019. <https://doi.org/10.3390/molecules24030445>.
40. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, et al. The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Res.* 2012. <https://doi.org/10.1093/nar/gkr1090>.
41. Brockington SF, Yang Y, Gandia-Herrero F, Covshoff S, Hibberd JM, Sage RF, et al. Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in *Caryophyllales*. *New Phytol.* 2015. <https://doi.org/10.1111/nph.13441>.
42. Erpen L, Devi HS, Grosser JW, Dutt M. Potential use of the DREB/ERF, MYB, NAC and WRKY transcription factors to improve abiotic and biotic stress in transgenic plants. *Plant Cell Tiss Organ Cult.* 2018. <https://doi.org/10.1007/s11240-017-1320-6>.
43. Hu Y, Chen X, Shen X. Regulatory network established by transcription factors transmits drought stress signals in plant. *Stress Biology.* 2022. <https://doi.org/10.1007/s44154-022-00048-z>.
44. Liang Y, Ma F, Li B, Guo C, Hu T, Zhang M, et al. A bHLH transcription factor, SlbHLH96, promotes drought tolerance in tomato. *Hortic Res.* 2022. <https://doi.org/10.1093/hr/uhac198>.
45. Baillo EH, Kimotho RN, Zhang Z, Xu P. Transcription factors associated with abiotic and biotic stress tolerance and their potential for crops improvement. *Genes.* 2019. <https://doi.org/10.3390/genes10100771>.
46. Zhang JY, Broeckling CD, Sumner LW, Wang ZY. 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 Mol Biol.* 2007; <https://doi.org/10.1007/s11103-007-9150-2>.

47. Lashbrooke J, Adato A, Lotan O, Alkan N, Tsimbalist T, Rechav K, et al. The Tomato MIXTA-Like Transcription Factor Coordinates Fruit Epidermis Conical Cell Development and Cuticular Lipid Biosynthesis and Assembly. *Plant Physiol.* 2015. <https://doi.org/10.1104/pp.15.01145>.
48. Kang C, Liu Z. Global identification and analysis of long non-coding RNAs in diploid strawberry *Fragaria vesca* during flower and fruit development. *BMC Genom.* 2015. <https://doi.org/10.1186/s12864-015-2014-2>.
49. Liu W, Cheng C, Lin Y, XuHan X, Lai Z. Genome-wide identification and characterization of mRNAs and lncRNAs involved in cold stress in the wild banana (*Musa itinerans*). *PLoS ONE.* 2018. <https://doi.org/10.1371/journal.pone.0200002>.
50. Chen K, Huang Y, Liu C, Liang Y, Li M. Transcriptome Profile Analysis of Arabidopsis Reveals the Drought Stress-Induced Long Non-coding RNAs Associated With Photosynthesis, Chlorophyll Synthesis, Fatty Acid Synthesis and Degradation. *Front Plant Sci.* 2021. <https://doi.org/10.3389/fpls.2021.643182>.
51. Corona-Gomez JA, Coss-Navarrete EL, Garcia-Lopez IJ, Pérez-Patiño JA, Selene L. F-V. Transcriptome-guided annotation and functional classification of long non-coding RNAs in *Arabidopsis thaliana*. *Sci Rep.* 2022. <https://doi.org/10.1038/s41598-022-18254-0>.
52. Lim PK, Zheng X, Goh JC, Mutwil M. Exploiting plant transcriptomic databases: Resources, tools, and approaches. *Plant Commun.* 2022. <https://doi.org/10.1016/j.xplc.2022.100323>.
53. González-Agüero M, García-Rojas M, Di Genova A, Correa J, Maass A, Orellana A, et al. Identification of two putative reference genes from grapevine suitable for gene expression analysis in berry and related tissues derived from RNA-Seq data. *BMC Genom.* 2013. <https://doi.org/10.1186/1471-2164-14-878>.
54. He F, Gui L, Zhang Y, Zhu B, Zhang X, Shen M, et al. Validation of reference genes for gene expression analysis in fruit development of *Vaccinium bracteatum* Thunb. using quantitative real-time PCR. *Sci Rep.* 2022. <https://doi.org/10.1038/s41598-022-20864-7>.
55. Liu J, Huang S, Niu X, Chen D, Chen Q, Tian L, et al. Genome-wide identification and validation of new reference genes for transcript normalization in developmental and post-harvested fruits of *Actinidia chinensis*. *Gene.* 2018. <https://doi.org/10.1016/j.gene.2017.12.012>.
56. Kou X, Zhang L, Yang S, Li G, Ye J. Selection and validation of reference genes for quantitative RT-PCR analysis in peach fruit under different experimental conditions. *Sci Hortic.* 2017. <https://doi.org/10.1016/j.scienta.2017.07.004>.
57. Zhu L, Yang C, You Y, Liang W, Wang N, MA F, et al. Validation of reference genes for qRT-PCR analysis in peel and flesh of six apple cultivars (*Malus domestica*) at diverse stages of fruit development. *Sci Hortic.* 2019. <https://doi.org/10.1016/j.scienta.2018.09.033>.
58. Wu J, Su S, Fu L, Zhang Y, Chai L, Yi H. Selection of reliable reference genes for gene expression studies using quantitative real-time PCR in navel orange fruit development and pummelo floral organs. *Sci Hortic.* 2014. <https://doi.org/10.1016/j.scienta.2014.06.040>.

59. Berumen-Varela G, Palomino-Hermosillo YA, Bautista-Rosales PU, Peña-Sandoval GR, López-Gómez GG, Balois-Morales R. Identification of reference genes for quantitative real-time PCR in different developmental stages and under refrigeration conditions in soursop fruits (*Annona muricata* L). *Sci Hortic*. 2020. <https://doi.org/10.1016/j.scienta.2019.108893>.
60. Cheng Y, Pang X, Wan H, Ahammed GJ, Yu J, Yao Z, et al. Identification of optimal reference genes for normalization of qPCR analysis during pepper fruit development. *Front Plant Sci*. 2017. <https://doi.org/10.3389/fpls.2017.01128>.
61. Kong Q, Yuan J, Gao L, Zhao L, Cheng F, Huang Y, et al. Evaluation of appropriate reference genes for gene expression normalization during watermelon fruit development. *PLoS ONE*. 2015. <https://doi.org/10.1371/journal.pone.0130865>.
62. Li-Beisson Y, Pollard M, Sauveplane V, Pinot F, Ohlrogge J, Beisson F. Nanoridges that characterize the surface morphology of flowers require the synthesis of cutin polyester. *Proc Natl Acad Sci U S A*. 2009. <https://doi.org/10.1073/pnas.0909090106>.
63. Sauveplane V, Kandel S, Kastner PE, Ehrling J, Compagnon V, Werck-Reichhart D, et al. *Arabidopsis thaliana* CYP77A4 is the first cytochrome P450 able to catalyze the epoxidation of free fatty acids in plants. *FEBS J*. 2009. <https://doi.org/10.1111/j.1742-4658.2008.06819.x>.
64. McCartney AW, Dyer JM, Dhanoa PK, Kim PK, Andrews DW, McNew JA, et al. Membrane-bound fatty acid desaturases are inserted co-translationally into the ER and contain different ER retrieval motifs at their carboxy termini. *Plant J*. 2004. <https://doi.org/10.1111/j.1365-313X.2004.01949.x>.
65. Pineau E, Sauveplane V, Grienemberger E, Bassard JE, Beisson F, Pinot F. CYP77B1 a fatty acid epoxidase specific to flowering plants. *Plant Sci*. 2021. <https://doi.org/10.1016/j.plantsci.2021.110905>.
66. Yang L, Shi C, Mu X, Liu C, Shi K, Zhu W, et al. Cloning and expression of a wild eggplant cytochrome P450 gene, StoCYP77A2, involved in plant resistance to *Verticillium dahliae*. *Plant Biotechnol Rep*. 2015. <https://doi.org/10.1007/s11816-015-0355-6>.
67. Ge S, Qin K, Ding S, Yang J, Jiang L, Qin Y, et al. Gas Chromatography-Mass Spectrometry Metabolite Analysis Combined with Transcriptomic and Proteomic Provide New Insights into Revealing Cuticle Formation during Pepper Development. *J Agric Food Chem*. 2022. <https://doi.org/10.1021/acs.jafc.2c04522>.
68. Girard AL, Mounet F, Lemaire-Chamley M, Gaillard C, Elmorjani K, Vivancos J, et al. Tomato GDGL1 is required for cutin deposition in the fruit cuticle. *Plant Cell*. 2012. <https://doi.org/10.1105/tpc.112.101055>.
69. Yeats TH, Martin LBB, Viart HMF, Isaacson T, He Y, Zhao L, et al. The identification of cutin synthase: Formation of the plant polyester cutin. *Nat Chem Biol*. 2012. <https://doi.org/10.1038/nchembio.960>.
70. Ding LN, Guo XJ, Li M, Fu ZL, Yan SZ, Zhu KM, et al. Improving seed germination and oil contents by regulating the GDGL transcriptional level in *Brassica napus*. *Plant Cell Rep*. 2019. <http://dx.doi.org/10.1007/s00299-018-2365-7>.

71. Yeats TH, Howe KJ, Matas AJ, Buda GJ, Thannhauser TW, Rose JKC. Mining the surface proteome of tomato (*Solanum lycopersicum*) fruit for proteins associated with cuticle biogenesis. J Exp Bot. 2010. <https://doi.org/10.1093/jxb/erq194>.
72. Natarajan P, Akinmoju TA, Nimmakayala P, Lopez-Ortiz C, Garcia-Lozano M, Thompson BJ, et al. Integrated metabolomic and transcriptomic analysis to characterize cutin biosynthesis between low- and high-cutin genotypes of *Capsicum chinense* jacq. Int J Mol Sci. 2020. <https://doi.org/10.3390/ijms21041397>.
73. Bird D, Beisson F, Brigham A, Shin J, Greer S, Jetter R, et al. Characterization of *Arabidopsis* ABCG11/WBC11, an ATP binding cassette (ABC) transporter that is required for cuticular lipid secretion. Plant J. 2007. <https://doi.org/10.1111/j.1365-313X.2007.03252.x>.
74. Panikashvili D, Savaldi-Goldstein S, Mandel T, Yifhar T, Franke RB, Höfer R, et al. The arabidopsis *DESPERADO/AtWBC11* transporter is required for cutin and wax secretion. Plant Physiol. 2007. <https://doi.org/10.1104/pp.107.105676>.
75. Chen N, Song B, Tang S, He J, Zhou Y, Feng J, et al. Overexpression of the ABC transporter gene *TsABCG11* increases cuticle lipids and abiotic stress tolerance in *Arabidopsis*. Plant Biotechnol Rep. 2018. <http://dx.doi.org/10.1007/s11816-018-0495-6>.
76. Liu L, Bao A, Li H, Bai W, Liu H, Tian Y, et al. Overexpression of ZxABCG11 from *Zygophyllum xanthoxylum* enhances tolerance to drought and heat in alfalfa by increasing cuticular wax deposition. Crop J. 2023. <https://doi.org/10.1016/j.cj.2022.11.007>.
77. Alkio M, Jonas U, Sprink T, Van Nocker S, Knoche M. Identification of putative candidate genes involved in cuticle formation in *Prunus avium* (sweet cherry) fruit. Ann Bot. 2012. <https://doi.org/10.1093/aob/mcs087>.
78. Vogg G, Fischer S, Leide J, Emmanuel E, Jetter R, Levy AA, et al. Tomato fruit cuticular waxes and their effects on transpiration barrier properties: Functional characterization of a mutant deficient in a very-long-chain fatty acid β -ketoacyl-CoA synthase. J Exp Bot. 2004. <https://doi.org/10.1093/jxb/erh149>.
79. Leide J, Hildebrandt U, Reussing K, Riederer M, Vogg G. The developmental pattern of tomato fruit wax accumulation and its impact on cuticular transpiration barrier properties: Effects of a deficiency in a β -ketoacyl-coenzyme A synthase (LeCER6). Plant Physiol. 2007. <https://doi.org/10.1104/pp.107.099481>.
80. Parsons EP, Popopvsky S, Lohrey GT, Lü S, Alkalai-Tuvia S, Perzelan Y, et al. Fruit cuticle lipid composition and fruit post-harvest water loss in an advanced backcross generation of pepper (*Capsicum* sp). Physiol Plant. 2012. <https://doi.org/10.1111/j.1399-3054.2012.01592.x>.
81. Aragón W, Formey D, Aviles-Baltazar NY, Torres M, Serrano M. *Arabidopsis thaliana* Cuticle Composition Contributes to Differential Defense Response to *Botrytis cinerea*. Front Plant Sci. 2021. <https://doi.org/10.3389/fpls.2021.738949>.
82. Wan CY, Wilkins TA. A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (*Gossypium hirsutum* L). Anal Biochem. 1994.

- <https://doi.org/10.1006/abio.1994.1538>.
83. Bolger AM, Lohse M, Usadel B, Trimmomatic. A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014. <https://doi.org/10.1093/bioinformatics/btu170>.
 84. Li W, Godzik A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006. <https://doi.org/10.1093/bioinformatics/btl158>.
 85. Langmead B. Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinforma*. 2010;32. 11.7.1–11.7.14.
 86. Li B, Dewey CN. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform*. 2011. <https://doi.org/10.1186/1471-2105-12-323>.
 87. Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, et al. PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res*. 2017. <https://doi.org/10.1093/nar/gkw982>.
 88. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. *Nucleic Acids Res*. 2021. <https://doi.org/10.1093/nar/gkaa913>.
 89. Kalvari I, Nawrocki EP, Argasinska J, Quinones-Olvera N, Finn RD, Bateman A, et al. Non-Coding RNA Analysis Using the Rfam Database. *Curr Protoc Bioinforma*. 2018. <https://doi.org/10.1002/cpbi.51>.
 90. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, et al. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res*. 2007;35:2:345–9. <https://doi.org/10.1093/nar/gkm391>.
 91. Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, Wei L, et al. CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res*. 2017. <https://doi.org/10.1093/nar/gkx428>.
 92. Robinson MD, McCarthy DJ, Smyth GK, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2009. <https://doi.org/10.1093/bioinformatics/btp616>.
 93. Thornton B, Basu C. Rapid and simple method of qPCR primer design. In: Basu C, editor. *PCR Primer Design. Methods in Molecular Biology*. New York: Humana Press; 2015. pp. 173–9.
 94. Xie F, Wang J, Zhang B. RefFinder: a web-based tool for comprehensively analyzing and identifying reference genes. *Funct Integr Genomics*. 2023. <https://doi.org/10.1007/s10142-023-01055-7>.
 95. Edgar RC. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform*. 2004. <https://doi.org/10.1186/1471-2105-5-113>.
 96. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol*. 2021. <https://doi.org/10.1093/molbev/msab120>.
 97. Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods*. 2001. <https://doi.org/10.1006/meth.2001.1262>.

Figures

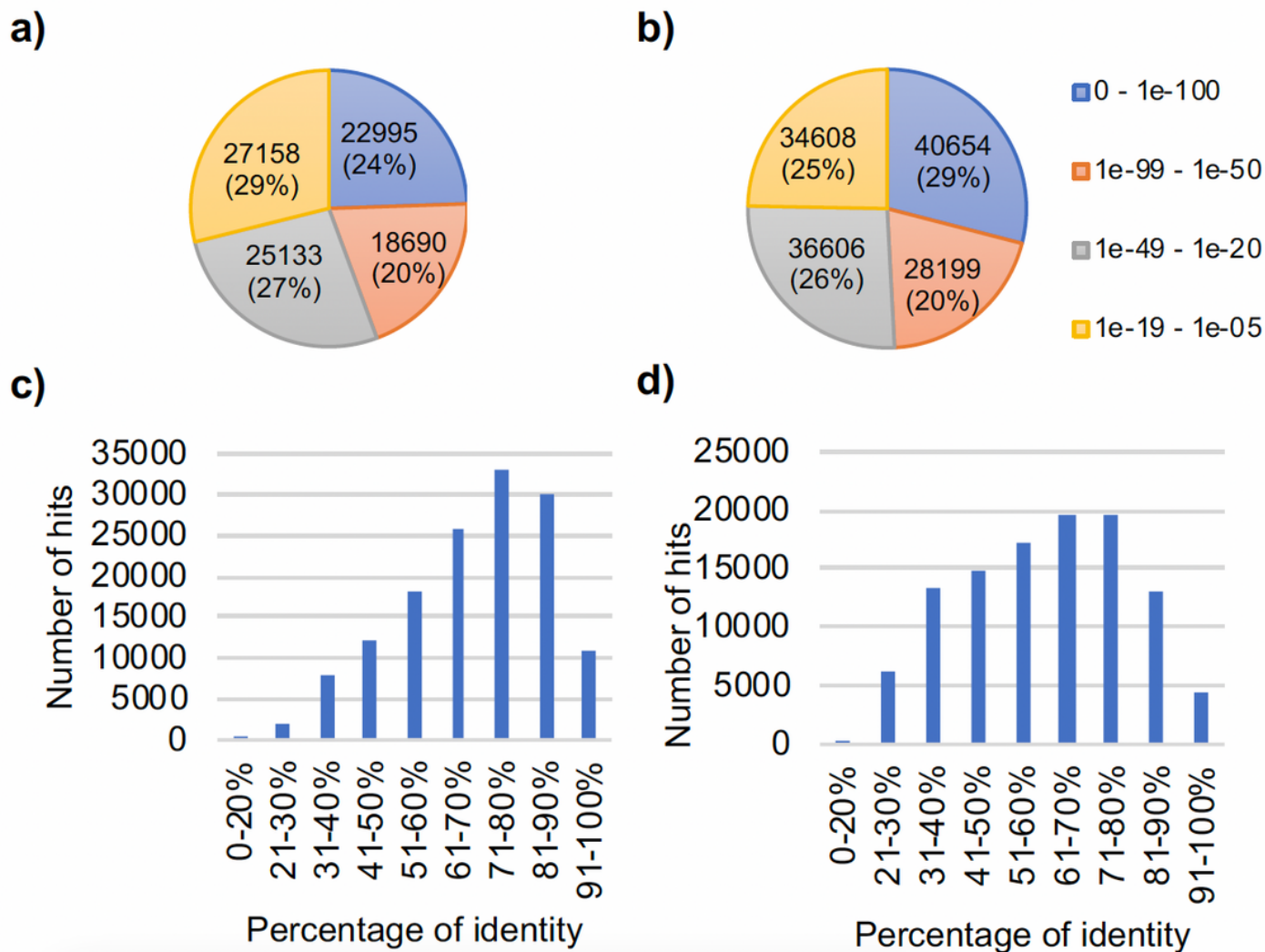


Figure 1

Homology analysis of assembled transcripts. E value distribution (**a, b**) and identity distribution (**c, d**) of the matches in the Swiss-Prot (**a, c**) and RefSeq (**b, d**) databases. (**a,b**) The number inside the pie chart indicates the number of transcripts recorded using that E value. Alignment by BLASTx with an E value threshold of 1×10^{-5} .

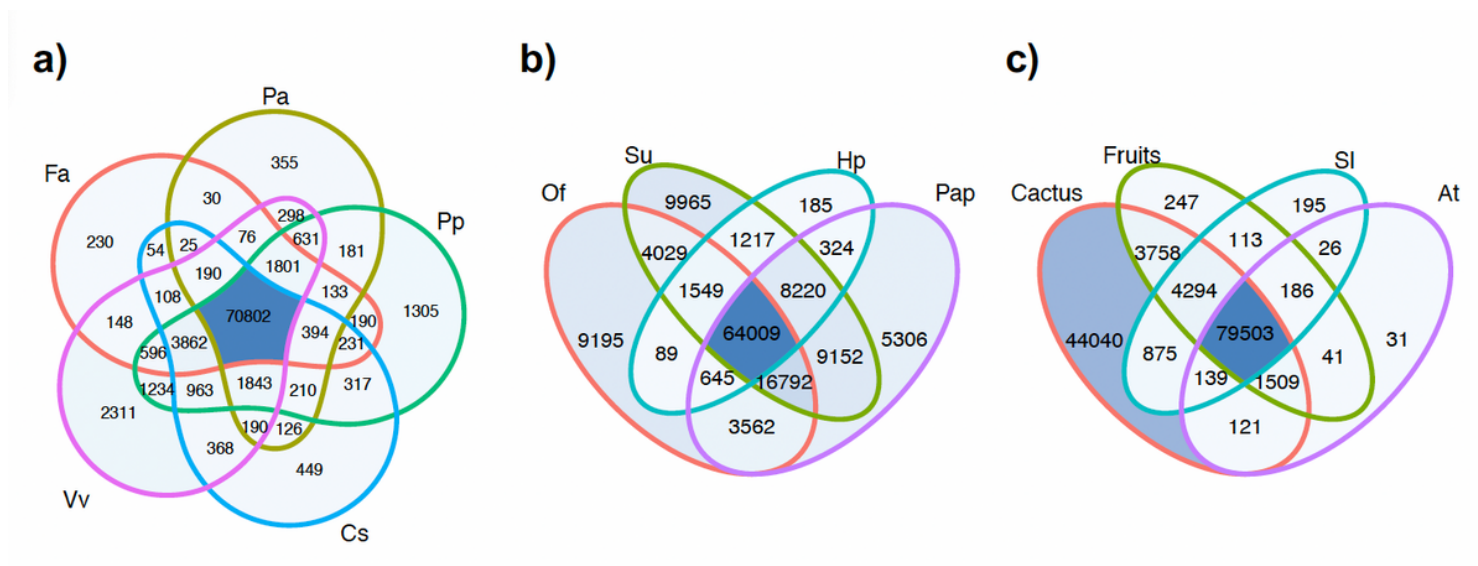


Figure 2

Venn diagram of the homology search results against model plants databases, commercial fruits, and cacti. The number in the diagram corresponds to the number of transcripts from *S. thurberi* homologous to sequences from that plant specie. **(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* (SI), *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.



Figure 3

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.

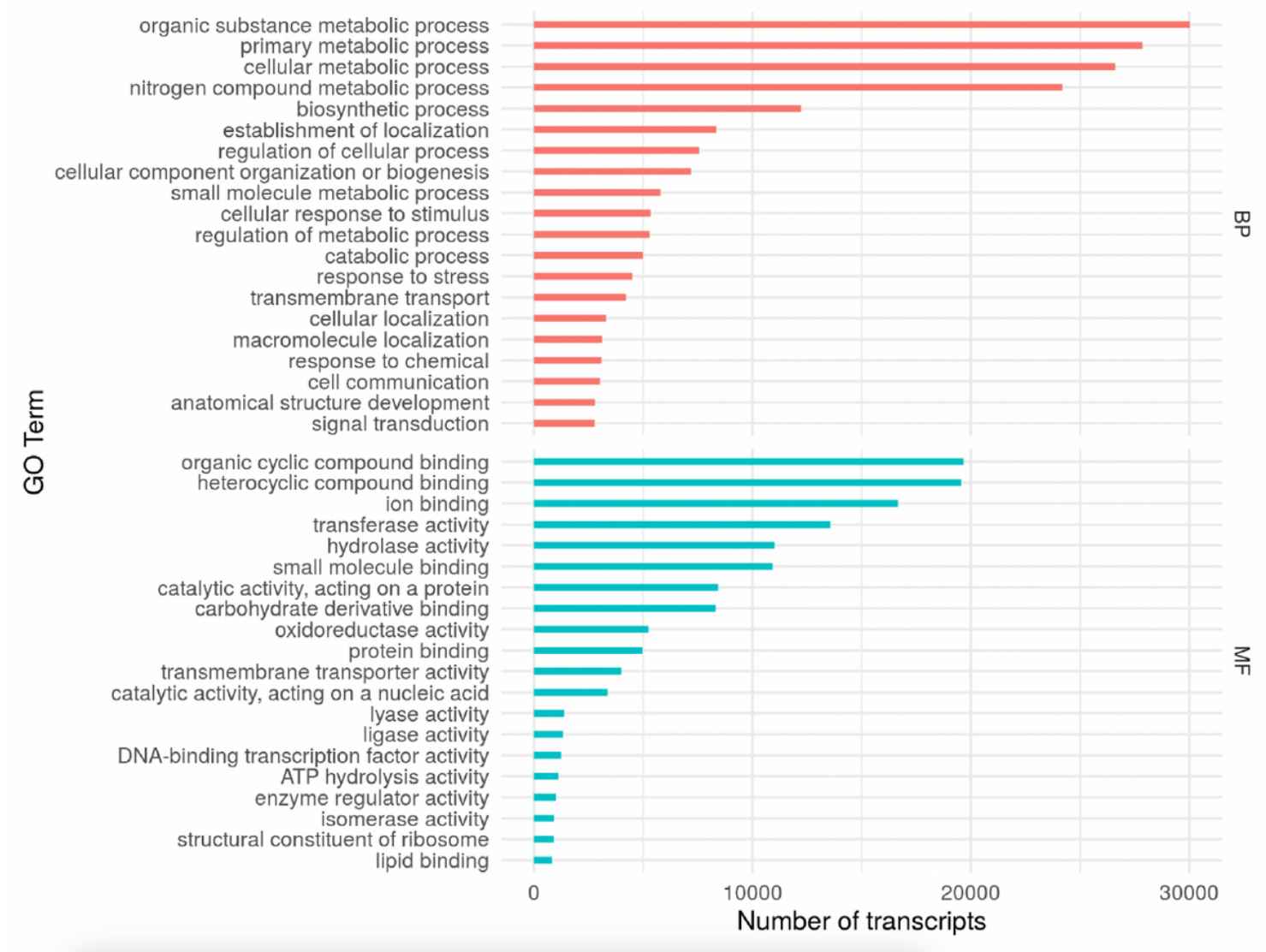


Figure 4

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.

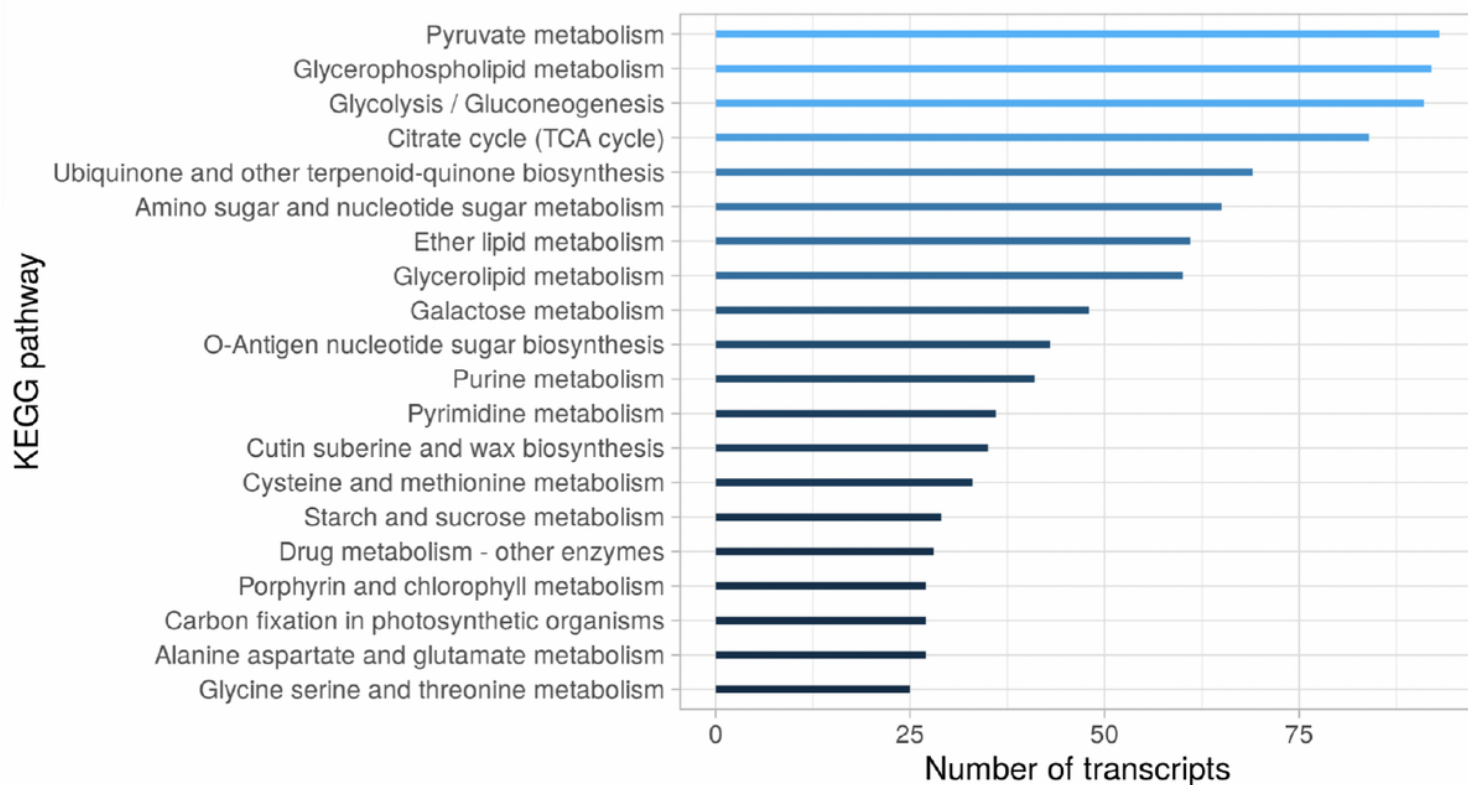


Figure 5

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.

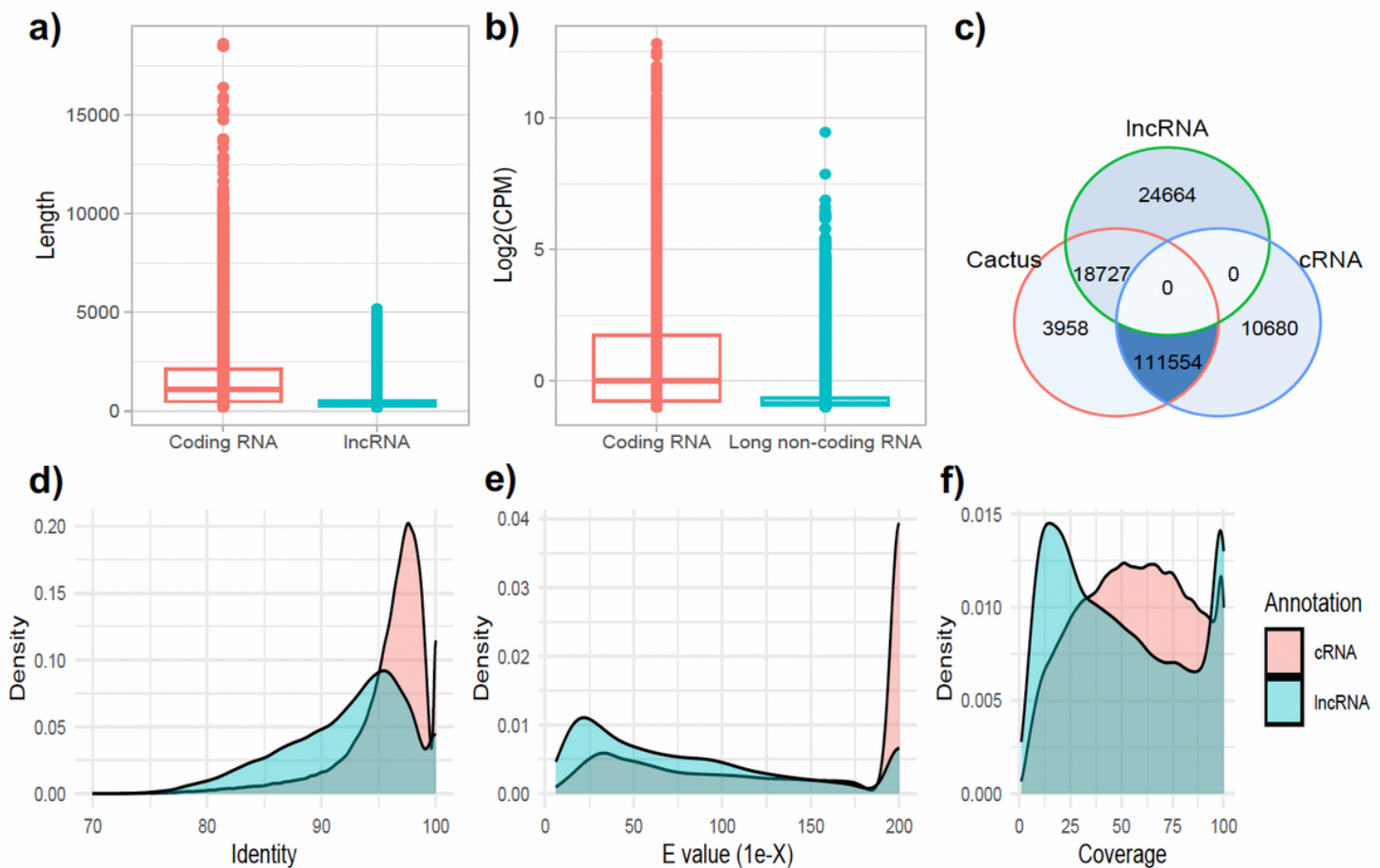


Figure 6

Comparison of coding RNA (cRNA) and lncRNAs from *S. thurberi* transcriptome and among cactus species. **(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 ($\log_2(\text{CPM})$) 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, and the higher and lower boxes limits represent the 25th-75th percentiles. **(c)** Venn diagram of the homology search results against the proteins of *Opuntia streptacantha* and transcripts of *Selenicereus undatus*, *Hylocereus polyrhizus*, and *Pachycereus pringlei*. **(d, e, f)** Density plots of the identity **(d)**, E value **(e)**, and coverage **(f)** for the hits recorded from the alignment of the cRNA and lncRNA from *S. thurberi* against transcripts of *S. undatus*, *H. polyrhizus*, and *P. pringlei*. **(e)** The X-axis indicates the negative exponential of the E values ($1e^{-X}$). As the highest negative exponential value recorded was 180, a value of $1e^{-200}$ was assigned to all the E values of 0.0 to represent the lowest E values. Homologous searching was carried out by BLAST alignment using an E value of less than 1×10^{-5} . The box plots, the density plots, and the Venn diagram were drawn by ggplot2 and ggVennDiagram in R Studio.

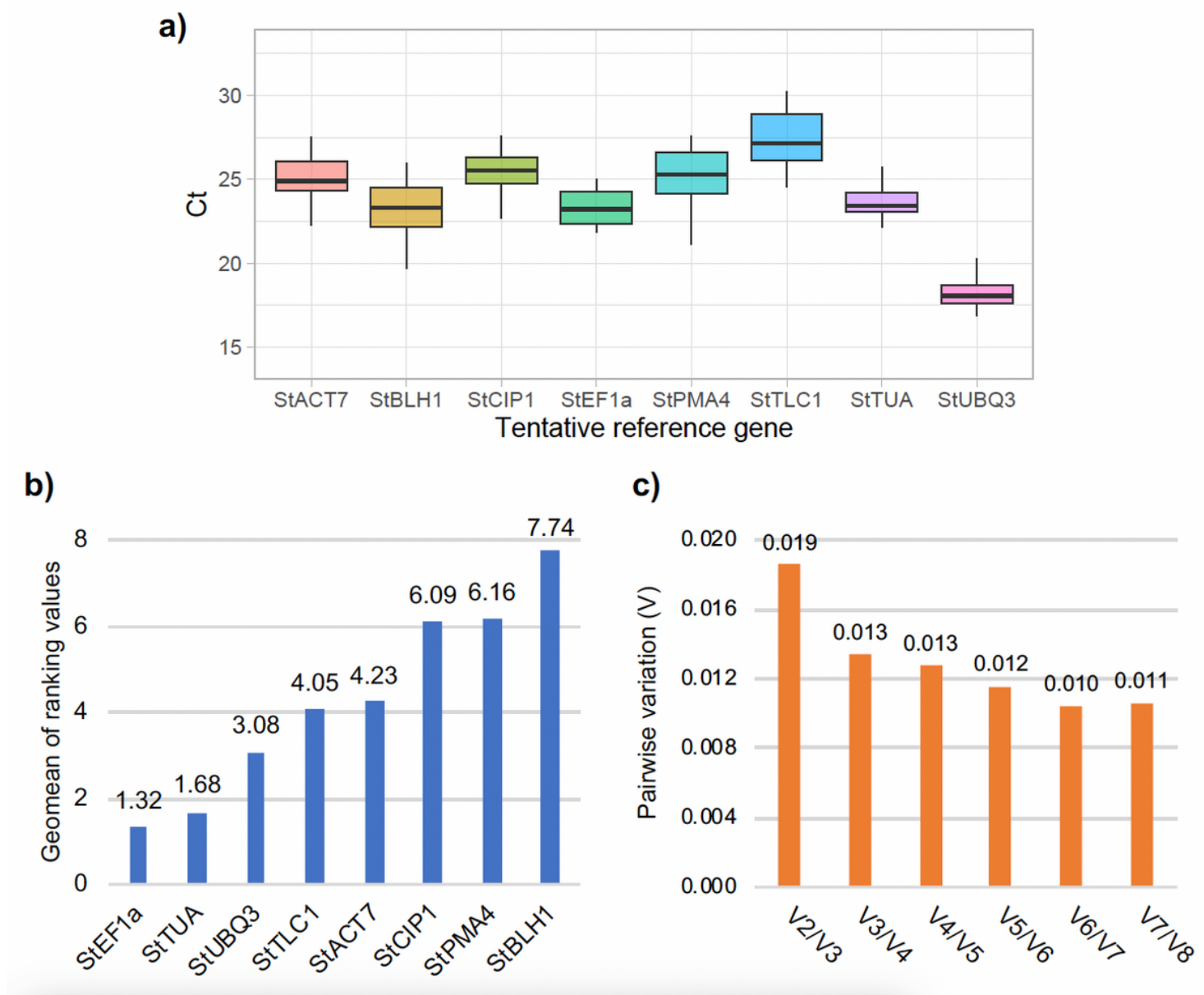


Figure 7

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 days after flowering). The black line inside the box indicates the median, and the higher and lower boxes limits represent the 25th-75th percentiles. **(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 V_n/V_{n+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), α -tubulin (StTUA), elongation factor 1- α (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).

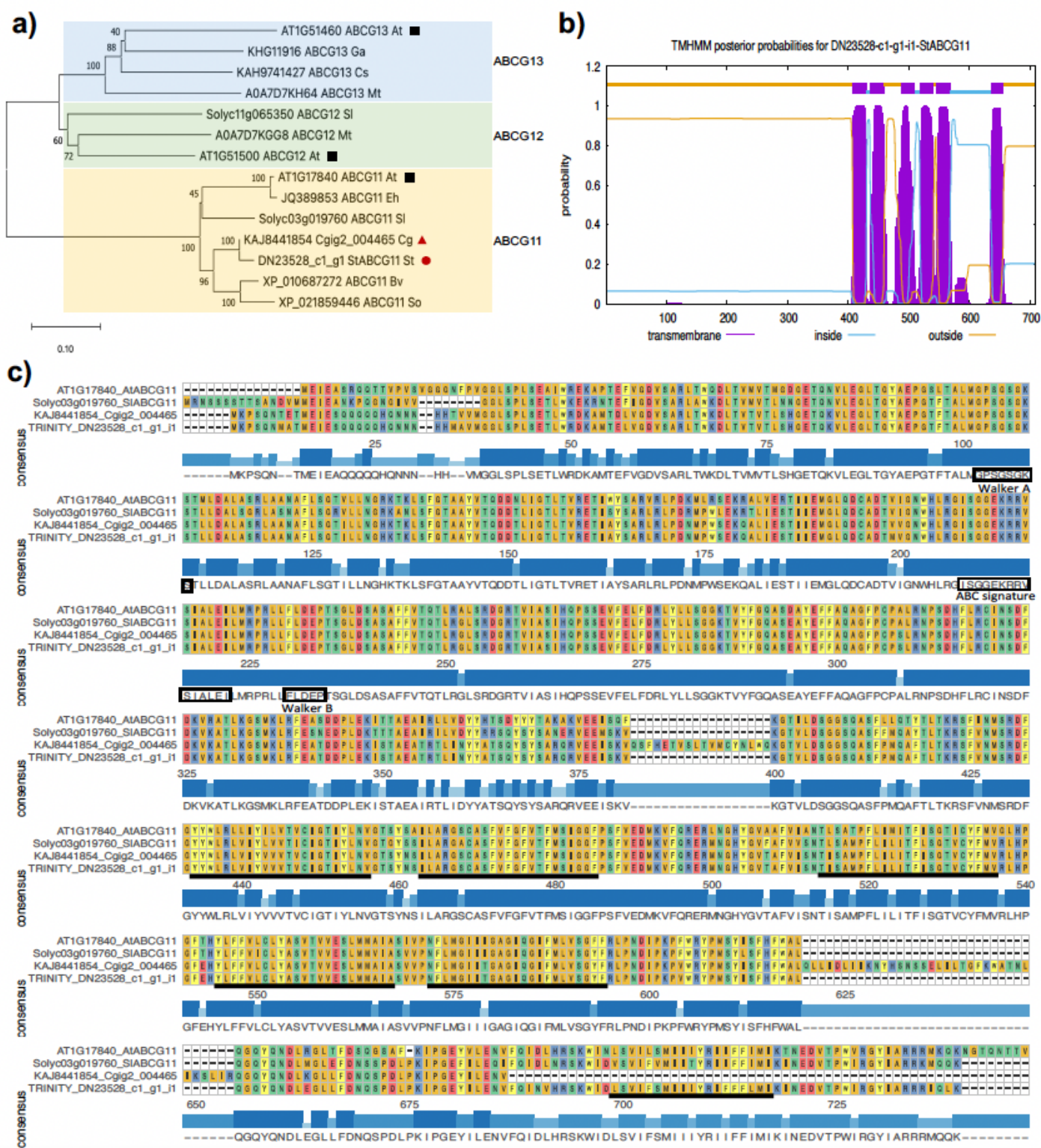


Figure 8

Analysis of the predicted protein StABCG11 from *Stenocereus thurberi*. (a) Phylogenetic tree of StABCG11 and related proteins of the classes ABCG11, ABCG12, and ABCG13 from *Arabidopsis thaliana* (At),

Gossypium arboreum (Ga), *Citrus sinensis* (Cs), *Medicago truncatula* (Mt), *Solanum lycopersicum* (Sl), *Eutrema halophilum* (Eh), *Carnegiea gigantea* (Cg), *Beta vulgaris* (Bv), and *Spinacia oleracea* (So). The database accession number next to the protein name is shown. The scale bar of 0.10 represented a sequence divergence of 10%. The number in the branches is the percentage bootstrap value of 1000 replicates. The highest percentages represent higher significant results. The black square beside the protein name shows AtABCG11, AtABCG12, and AtABCG13 from *A. thaliana*. The red circle and red triangle next to the protein name show StABCG11 from *S. thurberi* and a protein from the closest related species, *C. gigantea*, respectively. Neighbor-joining (NJ) phylogenetic tree constructed by MEGA11 software. **(b)** The predicted transmembrane helices of StABCG11. The probability of membrane insertion (Y-axis) and transmembrane region represented by purple color was determined by TMHMM software. **(c)** Multiple sequence alignment of StABCG11 and its homologous from *A. thaliana* (AT1G17840), *S. lycopersicum* (Soly03g019760), and *C. gigantea* (KAJ8441854). Amino acids are colored according to the chemistry classification of their side-chain. The darkest blue bars below the protein sequences indicate 100% conservation. Black rectangles show the conserved sequence of the Walker A and B motif and the ABC signature, named below the rectangles. Black width lines below the sequence show the predicted transmembrane helices of StABCG11. Alignment was carried out by MUSCLE in MEGA11 and drawn by ggmsa in R Studio.

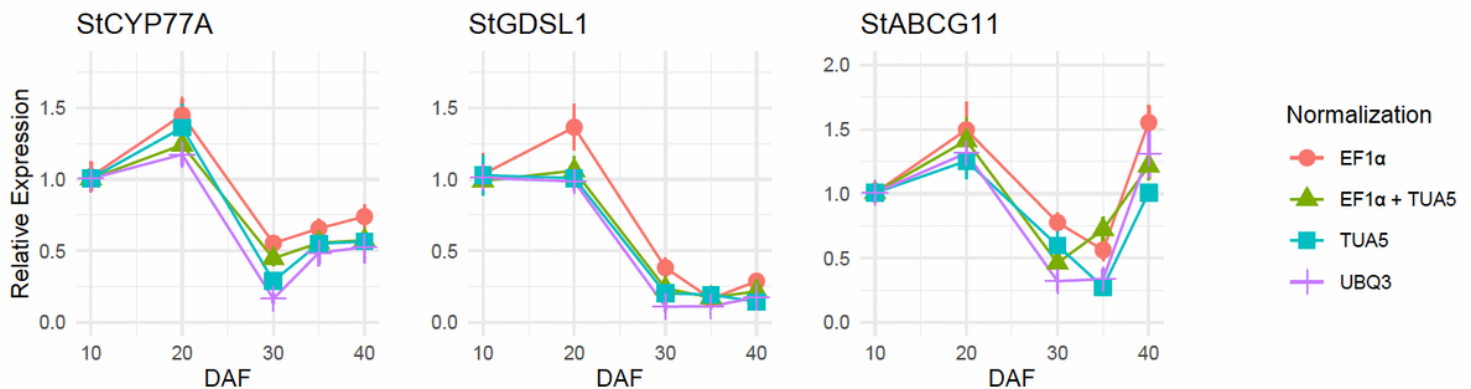


Figure 9

Relative expression of cutin biosynthesis-related transcripts StCYP77A, StGDSL1, and StABCG11 during sweet pitaya fruit development. Relative expression was calculated through the 2^{-DDCt} method using elongation factor 1-alpha (StEF1α), α-tubulin (StTUA), polyubiquitin 3 (StUBQ3), or StEF1α+StTUA as normalizing genes at 10, 20, 30, 35, and 40 days after flowering (DAF). The Y-axis and error bars represent the mean of the relative expression $\pm$ 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 family G member 11 (StABCG11).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)
- [Additionalfile2.docx](#)
- [Additionalfile3.docx](#)
- [Additionalfile4.docx](#)
- [Additionalfile5.docx](#)
- [Additionalfile6.docx](#)
- [Additionalfile7.docx](#)
- [Additionalfile8.docx](#)
- [Additionalfile9.docx](#)