

RNA-Seq-Based Transcriptome Profiling of Early Fruit Development in Chieh-qua and Analysis of Related Transcription Factors

Xuan Du

Shanghai Academy of Agricultural Sciences, Horticultural Research Institute

Na Liu

liuna202303@126.com

Shanghai Academy of Agricultural Sciences, Horticultural Research Institute

Panling Lu

Shanghai Academy of Agricultural Sciences, Horticultural Research Institute

Ying Wang

Shanghai Academy of Agricultural Sciences, Horticultural Research Institute

Bo Lu

Shanghai Academy of Agricultural Sciences, Institute of Agricultural Science and Technology Information

Shoubo Tian

Shanghai Academy of Agricultural Sciences, Horticultural Research Institute

Zhaohui Zhang

Shanghai Academy of Agricultural Sciences, Zhuanghang Comprehensive Experiment Station

Research Article

Keywords: Transcriptome, Early fruit development, Chieh-qua, Transcription factors, MYB_superfamily

Posted Date: May 12th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-2853210/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on June 12th, 2024. See the published version at <https://doi.org/10.1038/s41598-024-63871-6>.

Abstract

Background Chieh-qua (*Benincasa hispida* Cogn. var. Chieh-qua How.) early fruit development started post pollination. With the continuous expansion of the fruit, the soluble solid content of the fruit decreased. Although there was transcriptomics study on the early fruit development of cucumber, there was no report on the early fruit development of chieh-qua. In this study, fruit transcriptome of 0-, 3- and 7-days post pollination were compared.

Results 104747 unigenes in the range of 201–14,209bp with a N50 length of 2119bp and 161282 transcripts were assembled from clean reads and comparing with 6 public databases for similarity searching. Principal component analysis separated the fruit ages into three groups. Compared with the 0 dpp (C), there were differences in the expression of 12982 and 6541 genes in the fruit tissue 3 dpp and 7dpp respectively. Compared with 3 dpp (B), there were 14314 differentially expressed genes in the fruit of 7dpp (A). According to the analysis of transcription factors, 213 unigenes in MYB_superfamily was obtained. Among them 94 unigenes of MYB_superfamily differentially expressing in three stages. In the differential expression analysis of pairwise comparison, eight unigenes (Gene_id:

TRINITY_DN32880_c1_g2,

TRINITY_DN35142_c2_g2, TRINITY_DN32454_c11_g6,

TRINITY_DN34105_c2_g7, TRINITY_DN32758_c3_g3,

TRINITY_DN33604_c4_g10, TRINITY_DN34466_c3_g1,

TRINITY_DN35924_c3_g2) are homologous to MYB59, MYB-GT3b, MYB18, MYB4, MYB108, MYB306, MYB340, MYB-bHLH13. And these unigenes are significant differences among the three groups of comparison. Further, MYB59 and MYB18 exhibited higher expression during the fruit pollinated for one week. While, MYB4, MYB-GT3b, MYB108 and MYB306 showed the highest expression levels in the fruits that have been pollinated for three days. Additionally, MYB340 and MYB-bHLH13 have showed higher expressions at the stage of unpollinated.

Conclusion These results indicate that MYB59, MYB-GT3b, MYB18, MYB4, MYB108, MYB306, MYB340, MYB-bHLH13 could play crucial roles in chieh-qua fruit development, defence, and blossom. Overall, this study provides a basis for further investigations of the MYB_superfamily genes of early fruit expansion in chieh-qua.

Introduction

Chieh-qua (*Benincasa hispida* Cogn. var. Chieh-qua How.) is a subspecies of the wax gourd and a vegetable crop in the Cucurbitaceae family, which is widely distributed all over South China and Southeast Asian [1]. Its immature or mature fruit is consumed that can be boiled, fried, or pickled [1]. In the Compendium of Materia Medica, it is still a traditional Chinese medicine. Modern research has found that its juice has hypotensive effects [2].

Chieh-qua has a vigorous annual vine with dioecious flowers [1, 3]. When the female flower is pollinated, it starts the fruit development. Fruit development is generally divided into two stages, i.e., early fruit development and fruit ripening, which include the changes of fruit size, shape, nutritional components and texture. The focus of this study is the expansion and development of early fruits, which includes (i) fruit setting, (ii) fruit growth through cell division, and (iii) fruit growth through cell expansion [4, 5]. Rapid cell division and cell elongation occur in the growth stage and determine the final size and shape of the fruit [4, 6]. The size and shape of fruit is one of the most important characteristics that determine its market value [4]. The shape and size of fruit are regulated by molecular mechanisms that regulate fruit development.

In the process of fruit development, many aspects of general fruit size, shape and further development and changes depending on organ characteristics are determined at an early stage, and this process is controlled by transcription factors. In the process of tomato fruit development, MYB factor FSM1 participated in the negative regulation of fruit size in the early development of tomato fruit [7]. In addition, transcription factors NON-RIPENING (NOR), COLORLESS NON-RIPENING (CNR), and RIPENING INHIBITOR (MADS-RIN) play regulatory roles in tomato fruit ripening [5]. Therefore, the transcription factor MYB may play an important role in the early stage of fruit development.

In recent years, transcriptome sequencing has been applied to the fruit development of Cucurbitaceae crops, such as cucumber, melon and bottle gourd, to study the molecular transcription regulation of the fleshy fruit development [4, 8–10]. However, most of the research focuses on fruit ripening, size and flavor. Only a few studies focused on the early development of fruits, and studied tomatoes and cucumber instead of chieh-qua [8, 11]. In gourd, the difference of transcription level between two different size of fruits was studied [4]. In our opinion, development is a dynamic process, and it is also a differential change before and after an individual's development. Therefore, our research mainly focuses on the difference of transcription level before and after pollination and in the early stage of fruit development. The purpose of our study is to reveal the transcription changes in the early stage of fruit development in chieh-qua, and to find out the regulatory factors related to the early expansion of chieh-qua fruit. It is expected that the information generated by this work will be helpful to the development of the gourd breeding program with improved fruit size.

Results

Chieh-qua development and sampling.

Throughout the growth season of chieh-qua, were sampled at three stages, 7 days post pollination (A), 3 days post pollination (B), before pollination (C) (Fig. 1a). Chieh-qua development is monitored by the measurement of chieh-qua fruit weight and soluble solids content. After pollination, the weight of the fruit increases slowly in the first three days, but it increases greatly in 3–7 days (Fig. 1b). However, the content of soluble solids decreases with the increase of fruit (Fig. 1c). Given our interest in the

transcriptional changes that may be involved in regulating fruit early enlargement, we select unpollinated fruits and fruits developed for 3- and 7-days post pollination for RNA-seq.

Sequencing and de novo transcriptome assembly.

To obtain the chieh-quia fruits development transcriptome expression profile, one non-normalized library is constructed using fruits at different developmental stages from 0 dpp to 7 dpp fruit tissue. Illumina sequencing data from chieh-quia fruits is deposited in the NCBI SRA database under accession number PRJNA970527. In total, 536977732 Illumina PE raw reads are generated (Table 1). After removing adaptor sequences, ambiguous nucleotides and low-quality sequences, there are 533.845180 million clean reads remaining. Assembly of clean reads result in 104747 unigenes in the range of 201–14,209 bp with a N50 length of 2119 bp (Table 2).

Table 1
Summary of sequences analysis.

Sample	Raw reads	Raw bases	Clean reads	Clean bases	Error rate(%)	Q20(%)	Q30(%)	GC content(%)
A1	57582732	8.69G	57255182	8.56G	0.0232	98.86	95.95	44.32
A2	63237646	9.55G	62845702	9.39G	0.0235	98.73	95.56	44.54
A3	57047340	8.61G	56679826	8.47G	0.0231	98.86	95.97	44.6
B1	58379262	8.82G	58016168	8.68G	0.0237	98.66	95.37	44.52
B2	58940310	8.9G	58543660	8.73G	0.0235	98.72	95.61	44.57
B3	61404706	9.27G	61052804	9.12G	0.0235	98.71	95.52	44.1
C1	63277140	9.55G	62964486	9.42G	0.0232	98.84	95.88	44.6
C2	59538590	8.99G	59241530	8.86G	0.0232	98.85	95.91	44.72
C3	57570006	8.69G	57245822	8.55G	0.0233	98.82	95.84	44.71
Note: A1, A2, A3: 7dpp fruit tissue. B1, B2, B3: 3dpp fruit tissue. C1, C2, C3: 0dpp fruit tissue. Q20: The percentage of bases with a Phred value > 20. Q30: The percentage of bases with a Phred value > 30.								

Sequence annotation.

By comparing with 6 public databases for similarity searching, 161282 transcripts and 104047 unigenes are obtained. Analyses show that 44,845 unigenes (41.79%) have significant matches in the Swiss-Prot database, 114,83 (10.96%) in the COG database. Our results also show that 43,773 (41.79%) of non-redundant unigenes demonstrate similarity to the known genes in NR database. In total, there are 56,807 unigenes (54.23%) successfully annotated in at least one of the NR, Swiss-Prot, Pfam, COG, GO and KEGG databases (Table 1).

In addition, all unigenes are subjected to a search against the COG database for functional prediction and classification. 11,483 non-redundant unigenes (Table 2) are subdivided into 24 COG classifications (Fig. 2). Among them, the cluster of ‘Translation, ribosomal structure and biogenesis’ (946) is the largest group, followed by ‘Posttranslational modification, protein turnover, chaperones’ (715), ‘General function prediction only’ (556), ‘Energy production and conversion’ (371) and ‘Signal transduction mechanisms’ (353). Only a few unigenes are assigned to ‘Cell motility’ (12) and ‘Nuclear structure’ (5).

Table 2 BLAST analysis of transcripts and unigenes against public databases.		
	Transcript number(percent)	Unigene number(percent)
NR	90618(0.5619)	43773(0.4179)
Swiss-Prot	82407(0.5109)	44845(0.4281)
Pfam	67718(0.4199)	34580(0.3301)
COG	24772(0.1536)	11483(0.1096)
GO	51119(0.317)	21944(0.2095)
KEGG	53955(0.3345)	31076(0.2967)
Total_anno	104528(0.6481)	56807(0.5423)
Total	161282(1)	104747(1)

According to the international standardized gene functional classification system GO database, 21,944 non-redundant unigenes are classified into three major functional ontologies (biological process, cellular component and molecular function) (Fig. 3). For biological process (BP), dominant subcategories are ‘metabolic process’ (11,676), ‘cellular process’ (10,866) and ‘single-organism process’ (5,599). In the category of cellular component (CC), ‘cell’ (7,756), ‘cell part’ (7,645) and ‘membrane’ (6,345) are highly represented. Among molecular function (MF) terms, ‘binding’ (11,234) and ‘catalytic activity’ (9,934) are most represented. However, within each of the three categories, few genes are assigned to subcategories of ‘rhythmic process’, ‘nucleoid’ and ‘metallochaperone activity’.

KEGG is a large-scale knowledge base for systematic analysis of gene function, linkage of genomic information and functional information. According to KEGG, 3,1076 unigenes (Table 2) are assigned to 6 major metabolic pathways (Fig. 4). The pathways involving the largest number of unique transcripts are ‘translation’ (2712), followed by ‘folding, sorting and degradation’ (1992), whereas ‘drug resistance: antimicrobial’ (2) is the smallest group.

Differential expression analysis of assembled early development of chieh-quua fruits.

To better survey the biological mechanism of early development of chieh-quua fruits after pollination, it is important to identify the DE genes between 3 different developing stages. In order to improve the

accuracy of the measured expression level of further analysis, the data from three biological replicates are merged, and the FPKM (Fragments Per Kilobase per Million) value is calculated based on the merged data set. It's shown that three biological repetitions of sample A and sample C are clustered together, while two biological repetitions of sample B are clustered together in the PCA analysis (Fig. 5). Based on the quantitative results of expression level, the genes with different expression between two groups were analyzed, and the difference analysis software was DESeq2, and the screening threshold was $|\log_2FC| \geq 1$ & $p \text{ adjust} < 0.05$.

Table 3
Statistics of differential gene number

different_group	Total DEGs	up	down
C_vs_B	12982	6035	6947
C_vs_A	6541	2479	4062
B_vs_A	14314	6873	7441

Subsequently, the DEGs of three different fruits in early development stage are analyzed (Table 3, Fig. 6). Compared with the 0 dpp (C), there are differences in the expression of 12982 genes in the fruit tissue 3 dpp(B), among which 6035 genes are up-regulated and 6947 genes are down-regulated (Fig. 6a). Compared with 0 dpp (C), there are 6541 differentially expressed genes, of which 2479 genes are up-regulated and 4062 genes are down-regulated (Fig. 6b). Compared with 3 dpp (B), there are 14314 differentially expressed genes in the fruit of 7dpp (A), of which 6873 genes are up-regulated and 7441 genes are down-regulated (Fig. 6c).

Analysis of transcription factors

TF is a kind of protein which can combine with specific DNA sequence. It can recognize and bind to the cis-acting elements in the upstream regulatory region of genes through specific functional domain, which can activate or hinder the expression of genes. Transcription factor analysis is carried out on the assembled gene/transcript to infer whether the gene/transcript is a transcription factor and its transcription factor family. According to the analysis of transcription factors, the top 20 families of transcription factors are MYB_superfamily, C2HA, C2C2, etc (Fig. 7a). Among them, the number of nuigenes in MYB_ superfamily is the largest, with 213 nuigenes. During the early development of chieh-qua fruit, there are 94 unigenes of MYB_superfamily differentially expressing in three stages, among which there were 34 unigenes differentially expressed between B and C, 56 unigenes differentially expressed between C and A, and 45 genes differentially expressed between B and A, and the number of genes in all three groups of comparisons is four (Fig. 7b). The expression analysis of these differential myb_superfamily unigenes is described in detail in the supplementary materials. RT-qPCR is used to verify this expression. Among the eight selected unigenes of MYB_superfamily, the expression patterns of all unigenes in the early three stages of chieh-qua fruit are consistent with that of transcriptome

sequencing (Fig. 7c,7d). Myb59 and myb18 exhibited higher expression during the fruit pollinated for one week (Fig. 7c,7d). While, myb4, myb-GT3b, myb108 and myb306 showed the highest expression levels in the fruits that have been pollinated for three days (Fig. 7c,7d). Additionally, myb340 and myb-Bhlh13 have showed higher expressions at the stage of unpollinated (Fig. 7c,7d).

Discussion

Transcriptome differences of early fruit development after pollination of chieh-qua

Fruit expansion is influenced by external environment and internal gene expression regulation. Transcriptome sequencing can promote the identification of whole gene expression and systematic regulation mechanism by analyzing the fruits at different developmental stages at transcription level. In this study, in order to screen the genes related to early fruit development of chieh-qua, we sequenced and annotated the transcriptome of three fruit development stages of the same chieh-qua variety at same conditions. The female flower of cucurbitaceae crops bears tender fruit when it blooms, and after pollination, the fruit expands and develops (Fig. 1a), After one week of pollination, the fruit weight increased exponentially compared with that unpollination (Fig. 1b). Transcriptome sequencing of fruits in three different developmental stages produced 104747 single genes. In two different varieties of bottle gourd, only a small number of genes have differences in fruit development at the same development stage [4]. In this study, there are three different early fruit development stages of the same variety of chieh-qua, and the gene expression in these three stages is great different (Table 3).

Identification of DEGs in fruit transcriptome

In order to identify the corresponding genes related to the early swelling and development of chieh-qua fruit, the gene expression profiles of transcriptome range were compared among libraries of three development stages. When the transcription factor MYB is used as the gene set for analysis, the number of deg identified in the comparison of A to C and A to B is significantly higher than that detected in the comparison of B to C (Fig. 7b). These changes are consistent with the morphological and physiological changes of early development of chieh-qua fruits after pollination, with little change between B and C and exponential increase from B to A (Fig. 1a). However, in the whole transcriptome analysis, the number of deg identified in the comparison of A to B and B to C is significantly higher than that detected in the comparison of A to C (Table 3). This indicates that stage B is a key stage of development after pollination, and the ratio of B to C is the difference of gene expression before and after pollination, and the ratio of B to A is the difference of gene expression of fruit expansion after pollination.

Functional annotation of transcriptome

In total, there were 56,807 unigenes (54.23%) successfully annotated in at least one of the NR, Swiss-Prot, Pfam, COG, GO and KEGG databases. In the COG database, the number of unigenes that may have the functions of 'translation, ribosomal structure and biogenesis' and 'post translational modification, protein turn over, chaperones' is in the top three. Unigenes with the largest number of comments in GO database may have the functions of metabolic process, cellular process, cell, cell part, binding, catalytic activity. In KEGG database, the most frequently commented functions of unigenes are translation, folding, sorting and degradation, and carbohydrate metabolism.

Transcription factors play an important regulatory role in plant development

Eight candidate genes homologous to MYB family were obtained from the transcriptome data, and then compared with the genome data of winter melon in NCBI database. It was found that MYB340 and EBOII were homologous genes. EBOII was a transcription factor related to flowering in *Petunia hybrida*, and its transcript accumulation was at the peak during flowering [12]. Similarly, the expression value of this gene in flowering period of chieh-qua was significantly higher than that in two stages of fruit development after pollination (Fig. 7c,7d). In *Arabidopsis thaliana*, MYB59 is a transcription factor related to plant growth and stress response [13]. In chieh-qua, the transcript accumulation of MYB59's homologous gene is the largest in fruit of stage A (Fig. 7c,7d). EaMYB18 and SsMYB18 transgenic plants showed effective tolerance to drought and cold stress [14]. During the fruit development of chieh-qua, the expression of MYB18 homologous gene in the fruit after pollination for one week was significantly higher than that in the other two stages, which also indicated that the stronger stress resistance of old chieh-qua fruits might be related to the expression regulation of transcription factors. In *Arabidopsis thaliana*, Bhlh13 plays a negative regulatory role in the JA signaling pathway [15]. During the fruit expansion of chieh-qua, the expression of its homologous gene is negatively correlated with the expansion. In *Arabidopsis*, Myb-GT3b seemed to be predominantly expressed in floral buds and roots [16]. While in this study, the homolog in chieh-qua showed the highest expression during 3dpp (Fig. 7c,7d). There is no report on the function of MYB306-LIKE. In this study, it was found that the homologous gene expressed the highest in the fruit 3dpp. MYB108 and MYB4 are transcription factors that regulate plant lignification [17]. In rice, MYB108 acts as a transcriptional repressor of lignin biosynthetic genes and its mutation increases the lignin content in cell walls [18]. In *Arabidopsis thaliana*, ARF17 regulates the process of anther endothelial lignification by targeting MYB108 DNA [19]. In *Arabidopsis thaliana*, MYB4 is regulated by MKP-MAPK to inhibit the lignification of leaves [17]. There is also a lignification process between the flesh and the peel during the development of the fruit. After pollination, the homologous gene expression of MYB108 and MYB4 in the early stage of fruit development is the highest, which indicates that this stage is the key stage of fruit expansion and development, and may be the node stage of fruit appearance and texture change.

Conclusion

In the early development of chieh-qua fruit, there are differences in gene expression at different stages before and after pollination and during fruit expansion, which means that fruit expansion starts after pollination, and transcription factors play an important role in the expansion process.

These results indicate that MYB59, MYB-GT3b, MYB18, MYB4, MYB108, MYB306, MYB340, MYB-bHLH13 could play crucial roles in chieh-qua fruit development, defence, and blossom. Overall, this study provides a basis for further investigations of the MYB_superfamily genes of early fruit expansion in chieh-qua.

Materials and Methods

Chiec-qua plant and it's fruits early development measurements

The chiec-qua "Verdant No. 1" used in this research were originally obtained from Zhongziku APP (<http://www.zhongziku.cc/>) commercially. The seeds of chieh-qua were sown on March 10th, and planted in Zhuanghang Experimental Station of Shanghai Academy of Agricultural Sciences on April 18th. The fruits were selected 0,3,7 days post simultaneous pollination, and the weight of the fruit and the content of soluble solids were measured. Then fruits were harvested and washed under running tap water followed by sterile water, the fruits were sampled with a knife and wrapped in foil paper and put into liquid nitrogen, and stored in a refrigerator at -86°C for subsequent analysis. Each sample had three biological replicates, and each repeat included five individual plants

RNA extraction, library construction, and sequencing.

RNA was extracted using the RNAPrep Pure polysaccharide polyphenol plant total RNA extraction kit (DP441, Tiangen Biotechnology, Beijing, China) following the protocol they made. And the quantity and quality of RNA were determined using a NanoPhotometer® NP80 (IMPLEN, CA, USA). The RNA integrity number (RIN) was measured by Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). All extracted RNA samples reached the total amount $\geq 1\mu\text{g}$, the concentration $\geq 35\text{ng}/\mu\text{L}$, $\text{OD}_{260}/280 \geq 1.8$ and $\text{OD}_{260}/230 \geq 1.0$. At first, the mRNA can be separated from the total RNA by A-T base pairing of magnetic beads with Oligo(dT) and ployA. Then, fragmentation buffer was added to randomly fragment the mRNA, and a small fragment of about 300bp was separated by magnetic bead screening. Under the action of reverse transcriptase, six-base random primers were added, and mRNA was used as template to reverse synthesize one-strand cDNA, followed by two-strand synthesis to form a sTable double-stranded structure. End Repair Mix was added to make up the sticky end of double-stranded cDNA into a flat end, and then an "A" base was added to the 3' end to connect the Y-shaped linker. Polymerase chain reaction (PCR) and agarose gel separation were to enrich and recover target fragments. The target fragments were quantified by TBS380 (Picogreen) and mixed on the computer according to the data ratio. Then, bridge PCR amplification was carried out on cBot to generate clusters, which were subjected

to Illumina Novaseq 6000 for sequencing. Illumina platform converts the sequenced image signals into text signals by CASAVA Base Calling, and stores them in fastq format as raw data.

Sequence reads de novo assembly, raw data statistics and quality control

Software fastx_toolkit_0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/) was used to count the base distribution and quality fluctuation in each cycle of all sequencing reads, which can visually reflect the library construction quality and sequencing quality of sequencing samples from a macro perspective, and analyze the base quality, base error rate and base distribution of each sample.

The original sequencing data will contain sequencing linker sequences, low-quality reading segments, sequences with high N (N represents uncertain base information) rate and sequences with too short length, which will seriously affect the quality of subsequent analysis. Therefore, the original sequencing data will be quality controlled before analysis, so as to obtain high-quality clean data to ensure the accuracy of subsequent analysis results. Software fastp was used to remove connector sequence from reads, reads without inserted fragment due to connector self-connection, the base with low quality (Q-score < 20) at the end of the sequence (3' end), reads with N ratio exceeding 10%, reads with length less than 30bp.

All clean data was assembled by Trinity(<https://github.com/trinityrnaseq/trinityrnaseq>). TransRate and CD-HIT were used to optimize and filter the obtained initial assembly sequence. BUSCO (Benchmarking Universal Single-Copy Orthologs, <http://busco.ezlab.org>) was used for evaluating the assembly integrity of the transcriptome. Make statistics on the length of assembled transcripts, so as to visually obtain the length distribution of all transcripts. Comparing the clean reads of each sample with the reference sequence assembled by Trinity to obtain the mapping result of each sample, which is also the basis for subsequent quantification of genes and transcripts of each sample.

The assembled unigenes/transcripts annotation, expression analysis and TFs identification and expression analysis

The assembled transcriptome sequences are compared with six databases (NR, Swiss-Prot, Pfam, COG, GO and KEGG databases), and the annotation information in each database was obtained, and the annotation situation in each database was counted. Quantitative analysis of the expression level of unigenes/transcripts was carried out by software RSEM, so as to analyze the differential expression of unigenes/transcripts among different samples, and revealed the regulation mechanism of genes by combining the sequence function information. Differentially expressed genes (DEGs) were analyzed by DESeq2, and the P-value threshold in multiple tests was determined by the false discovery rate (FDR) (Benjamini and Hochberg, 1995). Based on $p\text{-adjust} < 0.05$ & $|\log_2FC| \geq 1$, genes/transcripts with different expression between groups were obtained by Transcripts Per Million reads (TPM).

By analyzing the domain information contained in transcripts, gene/transcripts can be predicted for transcription factors and family analysis, and then compared with the databases Animal TFDB (<http://bioinfo.life.hust.edu.cn/AnimalTFDB/>) or Plant TFDB (<http://planttfdb.cbi.pku.edu.cn/>) by using HMMER analysis method to obtain homologous transcription factor information. The expression difference analysis method of TFs is the same as the analysis method of differential gene/transcripts mentioned above.

Quantitative real-time PCR validation of expression of MYB_superfamily

Total RNAs were isolated by using the RNAPrep Pure polysaccharide polyphenol plant total RNA extraction kit (DP441, Tiangen Biotechnology, Beijing, China) according to the manufacturer's instruction. cDNA was synthesized from 3 µg of each total RNA by using Hifair® II 1st Strand cDNA Synthesis Kit (Yisheng Biotechnology, Shanghai, China), according to the manufacturer's instructions. qRT-PCR is performed using a QuantStudio 5 (ABI, USA) in combination with the UNICON® Power qPCR SYBR Master Mix (Yisheng Biotechnology, Shanghai, China). The primers to detect the Mybs transcripts are listed in File S1 in Supporting Information. The program of qRT-PCR is described as follows: an initial 95°C enzyme activation step for 30 s, followed by denaturation for 10 s at 95°C, annealing/extension 30 s at 60°C for 40 cycles. The data were collected using the QuantStudio 5 Manager software and processed in Microsoft Excel. The transcript levels of target genes were normalized firstly to the transcript levels of the constitutively expressed *Bhactin* gene, and then to *myb* gene transcripts in A (negative control) according to the $2^{-\Delta\Delta C_t}$ method [20].

Declarations

Acknowledgements

Not applicable

Authors' contributions

XD, NL, PL, YW, BL and ST performed the plant cultivation, transcriptome experiment. XD and ZZ performed the bioinformation analysis and was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Funding

This project was funded by the Program of Hu Nongke Tuizi (2022) NO. 1-20.

Availability of data and materials

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Source of plants statement

The chieh-qua varietie used in the study were complied with relevant institutional, national, and international guidelines and legislation. And the study also complies with the IUCN Policy Statement on Research Involving Species at Risk of Extinction and the Convention on the Trade in Endangered Species of Wild Fauna and Flora for collection of seed specimens used in the study.

References

1. Mao Y, Jiang B, Peng Q, Liu W, Lin Y, Xie D, He X, Li S. Cloning and characterization of WRKY gene homologs in Chieh-qua (*Benincasa hispida* Cogn. var. Chieh-qua How) and their expression in response to fusaric acid treatment. *3 Biotech*. 2017; 7(1):86.
2. Nakashima M, Shigekuni Y, Obi T, Shiraishi M, Miyamoto A, Yamasaki H, Etoh T, Iwai S. Nitric oxide-dependent hypotensive effects of wax gourd juice. *J Ethnopharmacol*. 2011; 138(2):404-407.
3. Cantwell M, Nie X, Zong RJ, Yamaguchi M. Asian vegeTables: Selected fruit and leafy types. *Progress in new crops*. 1996:488-495.
4. Zhang H, Tan J, Zhang M, Huang S, Chen X. Comparative Transcriptomic Analysis of Two Bottle Gourd Accessions Differing in Fruit Size. *Genes (Basel)*. 2020; 11(4).
5. Karlova R, Chapman N, David K, Angenent GC, Seymour GB, de Maagd RA. Transcriptional control of fleshy fruit development and ripening. *J Exp Bot*. 2014; 65(16):4527-4541.
6. Renaudin JP, Deluche C, Cheniclet C, Chevalier C, Frangne N. Cell layer-specific patterns of cell division and cell expansion during fruit set and fruit growth in tomato pericarp. *J Exp Bot*. 2017; 68(7):1613-1623.
7. Machemer K, Shaiman O, Salts Y, Shabtai S, Sobolev I, Belausov E, Grotewold E, Barg R. Interplay of MYB factors in differential cell expansion, and consequences for tomato fruit development. *Plant J*. 2011; 68(2):337-350.

8. Kaori A, Kevin MC, Rebecca G. Transcriptome analyses of early cucumber fruit growth identifies distinct gene modules associated with phases of development. *BMC Genomics* 2012; 13:518.
9. Xu X, Pan J, He M, Tian H, Qi X, Xu Q, Chen X. Transcriptome profiling reveals key genes related to astringency during cucumber fruit development. *3 Biotech.* 2019; 9(11):390.
10. Shin AY, Kim YM, Koo N, Lee SM, Nahm S, Kwon SY. Transcriptome analysis of the oriental melon (*Cucumis melo* L. var. *makuwa*) during fruit development. *PeerJ.* 2017; 5:e2834.
11. Zhang S, Xu M, Qiu Z, Wang K, Du Y, Gu L, Cui X. Spatiotemporal transcriptome provides insights into early fruit development of tomato (*Solanum lycopersicum*). *Sci Rep.* 2016; 6:23173.
12. Colquhoun TA, Schwieterman ML, Wedde AE, Schimmel BC, Marciniak DM, Verdonk JC, Kim JY, Oh Y, Galis I, Baldwin IT *et al.* EOBII controls flower opening by functioning as a general transcriptomic switch. *Plant Physiol.* 2011; 156(2):974-984.
13. Fasani E, DalCorso G, Costa A, Zenoni S, Furini A. The *Arabidopsis thaliana* transcription factor MYB59 regulates calcium signalling during plant growth and stress response. *Plant Mol Biol.* 2019; 99(6):517-534.
14. Prashant R, Shingote,, Prashant G, Kavar,, Madhuri C, Pagariya,, Abhijeet B, Muley,, K. H, Babu. Isolation and functional validation of stress tolerant EaMYB18 gene and its comparative physio-biochemical analysis with transgenic tobacco plants overexpressing SoMYB18 and SsMYB18. *3 Biotech.* 2020; 10:225.
15. Huang H, Hua G, Bei L, Meng F, Jiaojiao W, Cuili W, Haixia T, Lanxiang W, Chengyuan X, Dewei W *et al.* bHLH13 Regulates Jasmonate-Mediated Defense Responses and Growth. *Evolutionary Bioinformatics.* 2018; 14:1-8.
16. Ayadi M, Delaporte V, Li YF, Zhou DX. Analysis of GT-3a identifies a distinct subgroup of trihelix DNA-binding transcription factors in *Arabidopsis*. *FEBS Lett.* 2004; 562(1-3):147-154.
17. Hui L, Muyang W, Ying C, Kinya N, Shugang H, Jinshan G, Xiawei Z, Yue W, Jiyun L, Qun L *et al.* An MKP-MAPK protein phosphorylation cascade controls vascular immunity in plants. *Science Advances.* 2022; 8:eabg8723.
18. Miyamoto T, Takada R, Tobimatsu Y, Takeda Y, Suzuki S, Yamamura M, Osakabe K, Osakabe Y, Sakamoto M, Umezawa T. OsMYB108 loss-of-function enriches p-coumaroylated and triclin lignin units in rice cell walls. *Plant J.* 2019; 98(6):975-987.
19. Xu XF, Wang B, Feng YF, Xue JS, Qian XX, Liu SQ, Zhou J, Yu YH, Yang NY, Xu P *et al.* AUXIN RESPONSE FACTOR17 Directly Regulates MYB108 for Anther Dehiscence. *Plant Physiol.* 2019; 181(2):645-655.
20. Du X, Wang S, Gao F, Zhang L, Zhao JH, Guo HS, Hua C. Expression of pathogenesis-related genes in cotton roots in response to *Verticillium dahliae* PAMP molecules. *Sci China Life Sci.* 2017; 60(8):852-860.

Figures

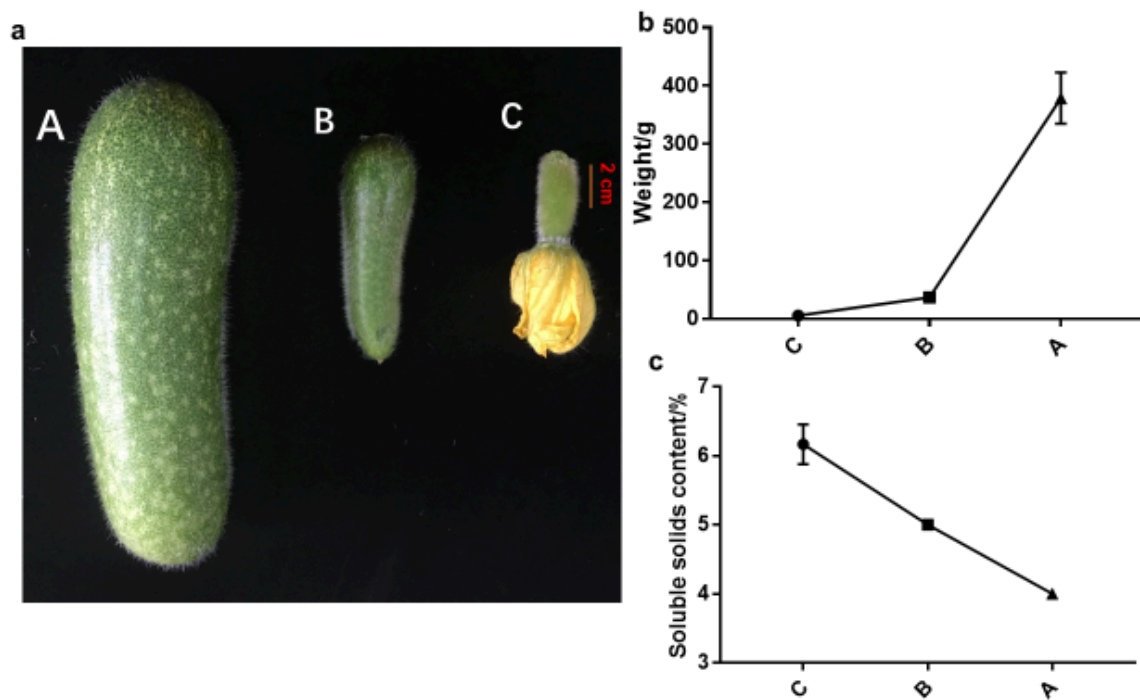


Figure 1

Early fruit development after pollination. Changes in appearance (a), weight (b), soluble solids content (c) during chieh-quia fruit early development, including 7dpp (A), 3dpp(B), 0dpp(C).

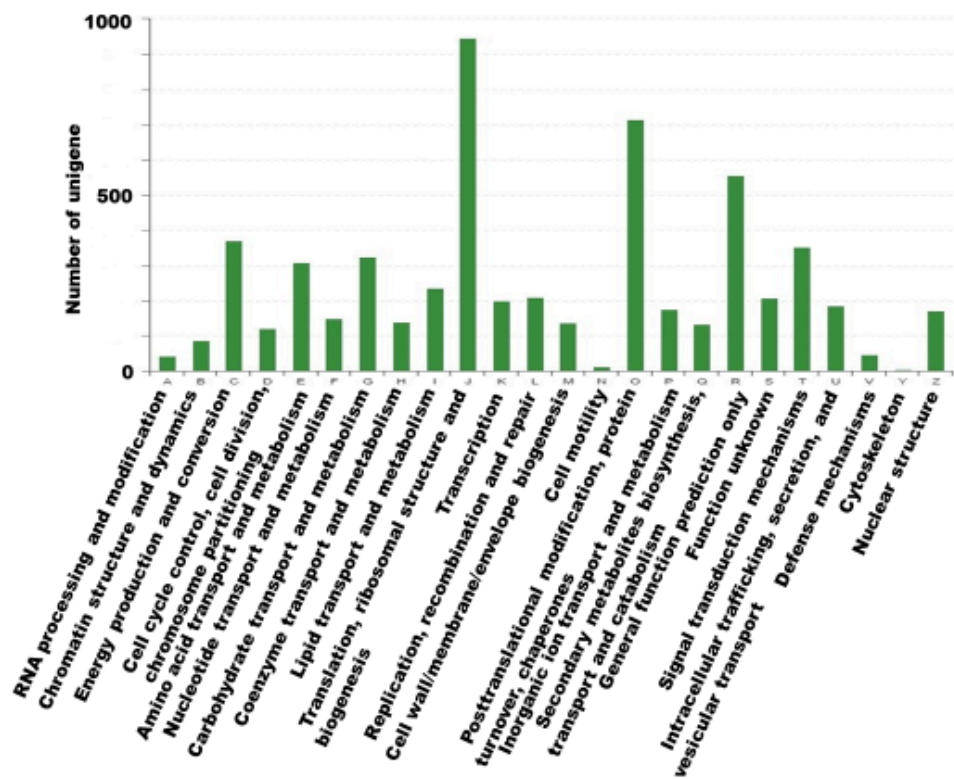


Figure 2

COG annotation of putative proteins.

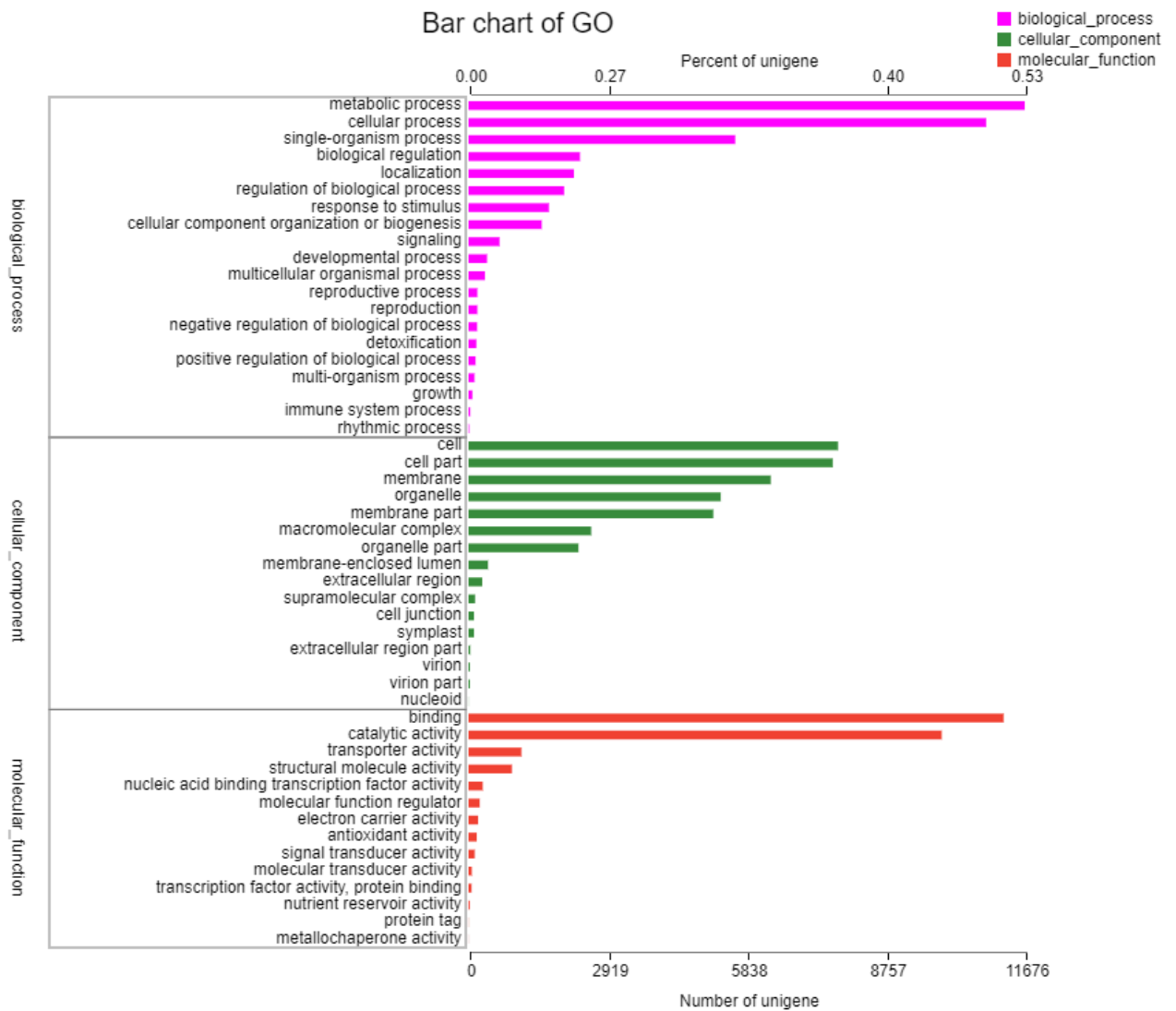


Figure 3

GO categorization of non-redundant unigenes. Each annotated sequence was assigned at least one GO term.

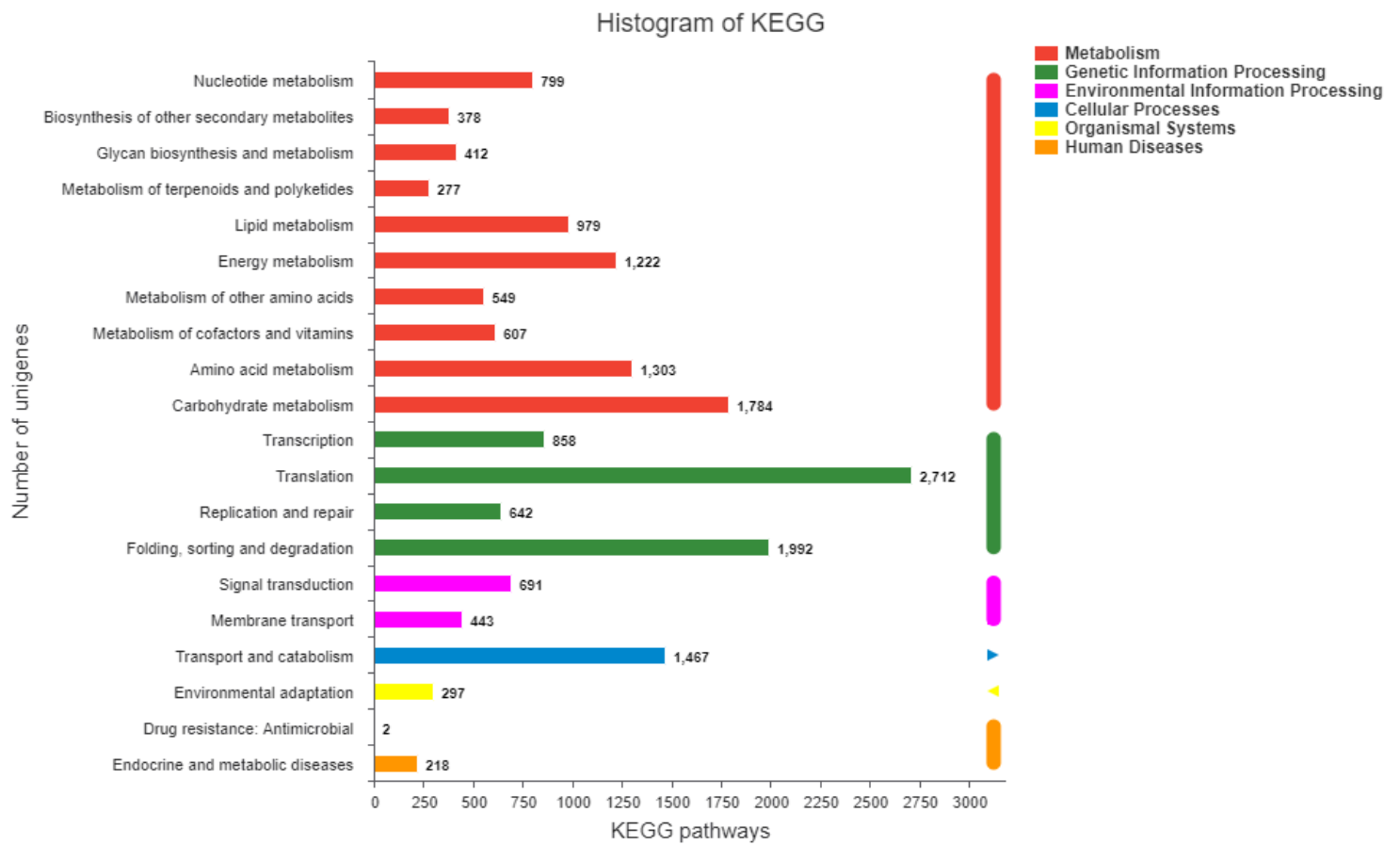


Figure 4

KEGG annotation of putative proteins.

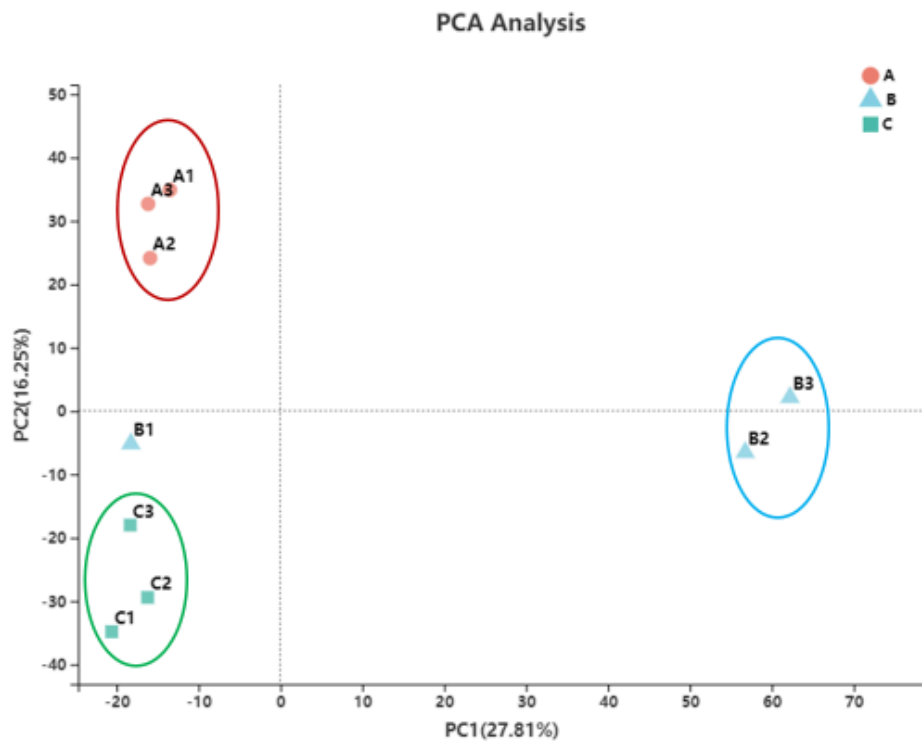


Figure 5

PCA analysis between samples

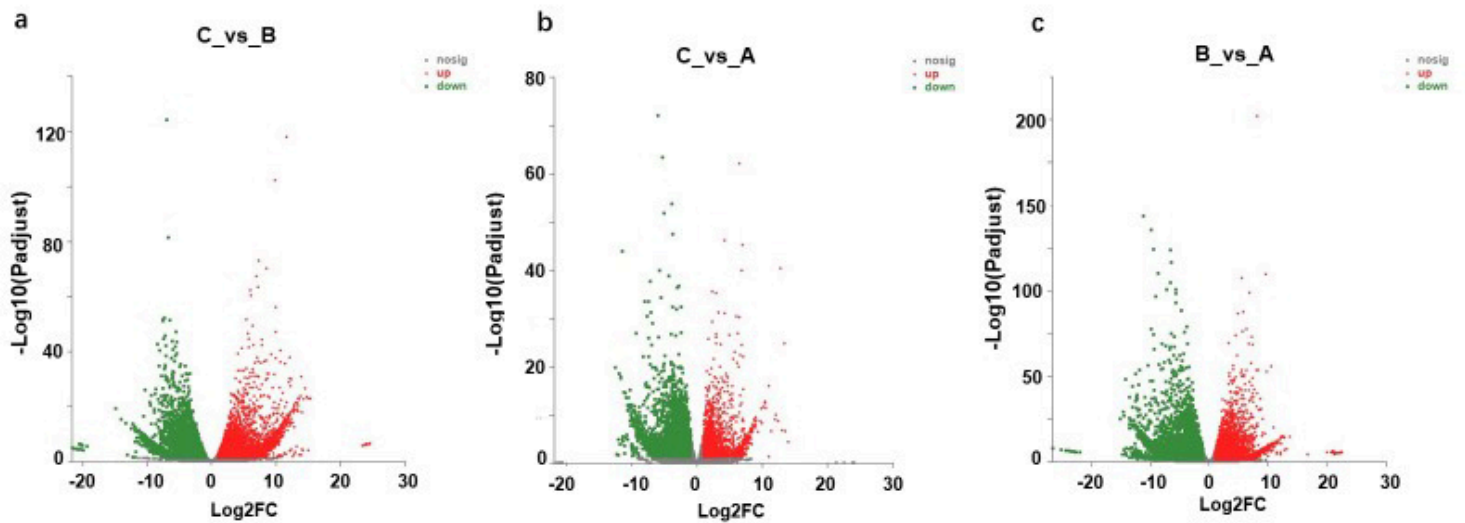


Figure 6

Volcano plot displaying differential expressed genes between two different developing stages. The abscissa is the multiple change value of gene expression difference between two samples, that is, the value obtained by dividing the expression number of treated samples by the expression number of control samples, and the ordinate is the statistical test value of gene expression difference, that is, P

value. The higher the P value is, the more significant the expression difference is, and the values of horizontal and vertical coordinates are all logarithmic. Each dot in the graph represents a specific gene, the red dot represents a significantly up-regulated gene, the green dot represents a significantly down-regulated gene, and the black dot represents a non-significant difference gene

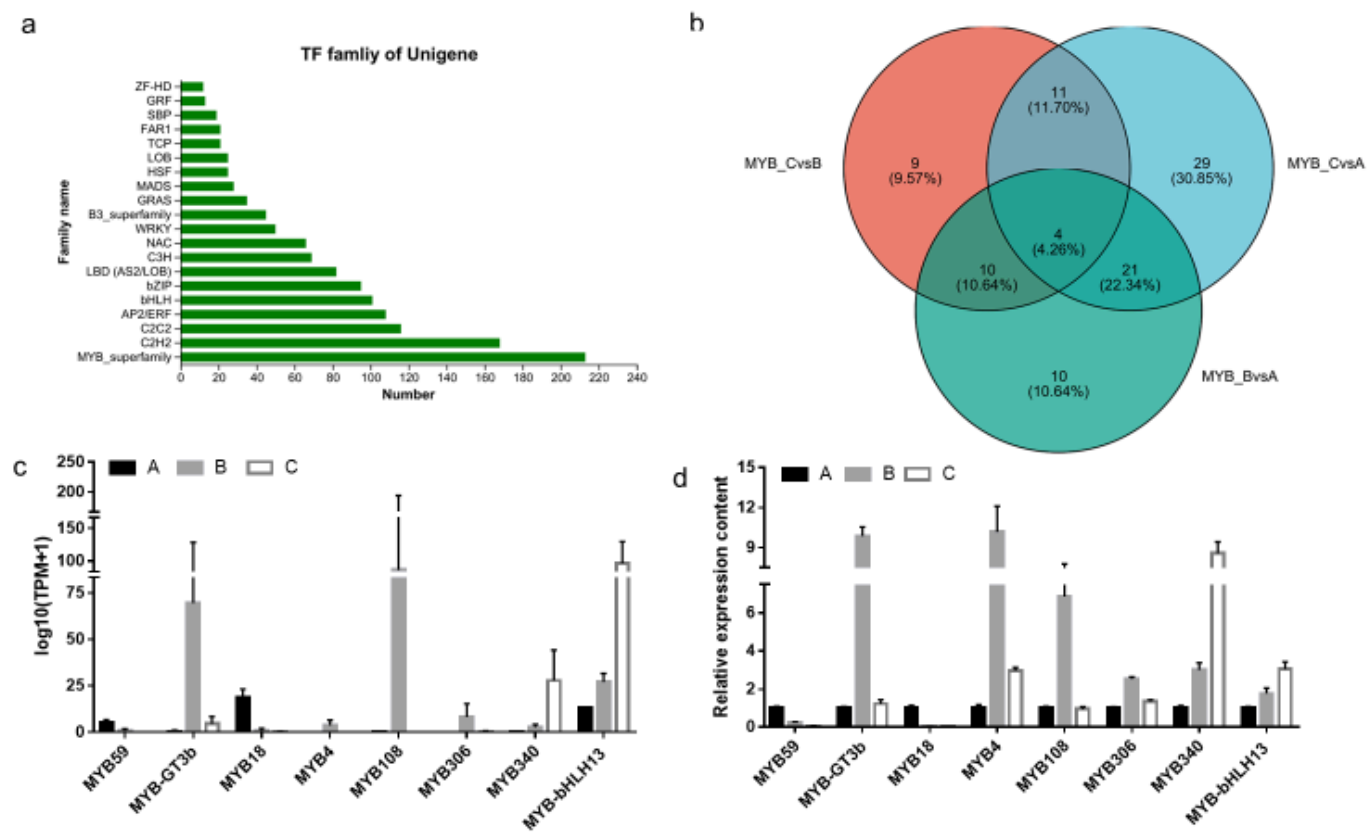


Figure 7

a: The top 20 families of transcription factors. b: Venn graph of gene sets, circles of different colors represent unigene of a set of genes, the value of represents the number of unigene shared or unique among gene sets. c: Expression analysis in this sequencing. d: The relative expression of RT-qPCR.

Supplementary Files

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

- [supplementarymaterials.docx](#)