Research Article

De Novo Transcriptome Sequencing of the Orange-Fleshed Sweet Potato and Analysis of Differentially Expressed Genes Related to Carotenoid Biosynthesis

Ruijie Li, Hong Zhai, Chen Kang, Degao Liu, Shaozhen He, and Qingchang Liu

Beijing Key Laboratory of Crop Genetic Improvement/Laboratory of Crop Heterosis and Utilization, Ministry of Education, China Agricultural University, Beijing 100193, China

Correspondence should be addressed to Qingchang Liu; liuqc@cau.edu.cn

Sweet potato, *Ipomoea batatas* (L.) Lam., is an important food crop worldwide. The orange-fleshed sweet potato is considered to be an important source of beta-carotene. In this study, the transcriptome profiles of an orange-fleshed sweet potato cultivar "Weiduoli" and its mutant "HVB-3" with high carotenoid content were determined by using the high-throughput sequencing technology. A total of 13,767,387 and 9,837,090 high-quality reads were produced from Weiduoli and HVB-3, respectively. These reads were *de novo* assembled into 58,277 transcripts and 35,909 unigenes with an average length of 596 bp and 533 bp, respectively. In all, 874 differentially expressed genes (DEGs) were obtained between Weiduoli and HVB-3, 401 of which were upregulated and 473 were downregulated in HVB-3 compared to Weiduoli. Of the 697 DEGs annotated, 316 DEGs had GO terms and 62 DEGs were mapped onto 50 pathways. The 22 DEGs and 31 transcription factors involved in carotenoid biosynthesis were identified between Weiduoli and HVB-3. In addition, 1,725 SSR markers were detected. This study provides the genomic resources for discovering the genes involved in carotenoid biosynthesis of sweet potato and other plants.

1. Introduction

Sweet potato, *Ipomoea batatas* (L.) Lam., is an important food crop widely cultivated in the world, especially in the tropics, subtropics, and some temperate zones of the developing countries [1, 2]. This crop is also used to produce alcohol and various antioxidants such as anthocyanin and carotenoids [3, 4]. The storage roots of orange-fleshed sweet potato are rich in beta-carotene, a precursor of vitamin A [5]. High carotenoid content has become one of the most important objectives in sweet potato breeding [6]. Sweet potato is an autohexaploid ($2n = 6x = 90$) and its estimated genome size is 2.4 Gb [7]. The genome data sources for sweet potato are important for gene discovery due to its complex genome.

Carotenoids are widely produced in plants, algae, fungi, and bacteria and provide potent nutritional benefits to humans because their bodies are unable to synthesize carotenoids independently [8, 9]. The necessity of this nutritional component has caused scientists to try to increase the carotenoid content of crops through different approaches. In plants, carotenoids are synthesized through a series of chemical reactions including condensation, dehydrogenation, cyclization, hydroxylation, and epoxidation. To date, a number of genes involved in the carotenoid biosynthesis have been cloned from several plants and their overexpression was found to significantly increase carotenoid levels in canola seeds [10], tomato fruits [11], and rice seeds [12, 13]. Several carotenoid biosynthesis-associated genes have also been isolated from sweet potato [6, 14–17]. However, the molecular mechanisms regulating flux through the pathway are unclear though carotenoid synthesis is well characterized.

Genomic approaches have been used for discovering the important genes involved in plant secondary metabolism pathways. However, the genome of sweet potato is still unavailable. Transcription sequencing is an efficient way for discovering and characterizing novel enzymes and transcription factors from sweet potato. Transcriptome sequencing of sweet potato has provided an important transcriptional data

source for studying storage root formation, flower development, and anthocyanin biosynthesis of this crop [7, 18–22]. Here, we performed *de novo* transcriptome sequencing of the orange-fleshed sweet potato by Illumina paired-end (PE) RNA sequencing technology and analyzed differentially expressed genes related to carotenoid biosynthesis.

2. Materials and Methods

2.1. Plant Materials. The orange-fleshed sweet potato cultivar "Weiduoli" and its high carotenoid mutant "HVB-3" were used in this study. Weiduoli is a commercial cultivar with carotenoid content of 9.02 mg/100 g (FW) and β-carotene content of 7.70 mg/100 g. In HVB-3, the contents of carotenoid and β-carotene are up to 21.42 mg/100 g and 19.95 mg/100 g, respectively. The storage roots of both materials were harvested after 110 days of planting, cleaned with sterilized water, and cut into 5 mm × 5 mm pieces. The collected samples were immediately frozen in liquid nitrogen and stored at −80°C for RNA extraction.

2.2. RNA Extraction. Total RNA from storage roots of Weiduoli and HVB-3 was extracted using the RNAprep Pure Plant Kit (Tiangen Biotech, Beijing, China). To avoid the contamination of genomic DNA, the extracted RNA was treated with DNase I (Takara, Dalian, China) for 4 h at 37°C. The quality of RNA was examined using 1% agarose gel before proceeding. Total RNA was quantified by using a Nanodrop spectrophotometer (Thermo Nanodrop Technologies, Wilmington, DE, USA). Both the A260/280 and A260/230 ratios were checked to ensure the purity of the total RNA and 10 μg RNA was used for Illumina paired-end (PE) sequencing.

2.3. cDNA Library Construction and Illumina Sequencing. Using magnetic beads with oligo (dT), poly-A mRNA was enriched from total RNA to construct a cDNA library for RNA sequencing. The enriched mRNA was broken into short fragments by adding fragment buffer. Using these short fragments, the first-strand of cDNA was synthesized by random hexamer primers. Then, using DNA polymerase I and RNase H (Tiangen Biotech, Beijing, China), the second-strand of cDNA was synthesized. After purification with a QiaQuick PCR extraction kit (Qiagen, Valencia, CA, USA), the cDNA fragments were resolved in elution buffer (EB) for end reparation and the addition of a poly(A) tail. Sequencing adapters were connected to the short fragments. These products were purified by agarose gel electrophoresis and suitable fragments (about 180 bp) were isolated as templates for PCR amplification. The cDNA library was constructed for sequencing by 2 × 100 PE using Illumina HiSeq 2000.

2.4. Raw Sequence Processing and De Novo Assembly. To obtain high quality reads for *de novo* assembly, the raw reads from RNA-seq were cleaned by removing adaptor sequences, empty reads, low quality reads (with ambiguous sequences "*N*"), and reads with more than 10% $Q < 20$ bases ($Q = -10 \times \lg E$). The clean reads from the two libraries were assembled together with the Trinity software [23]. The reads

were assembled into the contigs with the Inchworm program. The minimally overlapping contigs were clustered into sets of connected components by the Chrysalis program, and then the transcripts were constructed by the Butterfly program. The transcripts were clustered by similarity of correct match length beyond the 80% of longer transcript or 90% of shorter transcript using multiple sequence alignment tool— BLAST [24]. Taking the longest transcript as the unigene of each cluster, these unigenes formed into the nonredundant unigene database.

2.5. Analysis of Differentially Expressed Genes (DEGs). The expression of unigenes in Weiduoli and HVB-3 was calculated according to the RPKM method (reads per kb per million reads) described by Mortazavi et al. [25]. The IDEG6 software [26] was used to identify DEGs in the two libraries. The results of all statistical tests were corrected for multiple testing with the Benjamini-Hochberg false discovery rate (FDR < 0.01) and an absolute value of log2 ratio >1 was used to determine significant differences in gene expression.

2.6. Functional Annotation and Classification of DEGs. In order to deduce the correct transcription direction and protein sequences coded by DEGs, a BLASTX search was performed against the National Center for Biotechnology Information (NCBI) nonredundant (Nr) protein database (http://www.ncbi.nlm.nih.gov), the Swiss-Prot protein database, (http://www.expasy.ch/sprot), the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (http://www.genome.jp/kegg), Pfam database, and Cluster of Orthologous Groups (COG) database (http://www.ncbi.nlm.nih.gov/COG) with a typical cut-off E value of 10^{-5}. Gene ontology (GO) was applied with the Blast2GO program to obtain annotation of DEGs [27]. The WEGO software was then used to perform GO functional classification of DEGs. DEGs were annotated with corresponding Enzyme Commission (EC) numbers using BLASTX alignments against KEGG with a cut-off E value of 10^{-5}. Gene names were assigned to each DEG based on the best BLAST hit (highest score). Searches were limited to the first 10 significant hits for each query to increase computational speed.

2.7. SSR Detection. SSRs were detected among the unigenes with length >1,000 bp using the software MISA (http://pgrc.ipk-gatersleben.de/misa/). A total of 6 types of SSRs were investigated, including mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats.

3. Results

3.1. Transcriptome Sequencing and De Novo Assembly. Illumina HiSeq 2000 was used to determine the transcriptome profiles of sweet potato. After removing adaptor sequences and unknown or discarding low quality reads, 13,767,387 and 9,837,090 high-quality reads were obtained from Weiduoli and HVB-3, respectively (Table 1). With the Trinity assembly software, the high-quality reads were assembled into 1,557,001 contigs with an average length of 58 bp and N50 length of 58 bp. These contigs were assembled into 58,277 transcripts
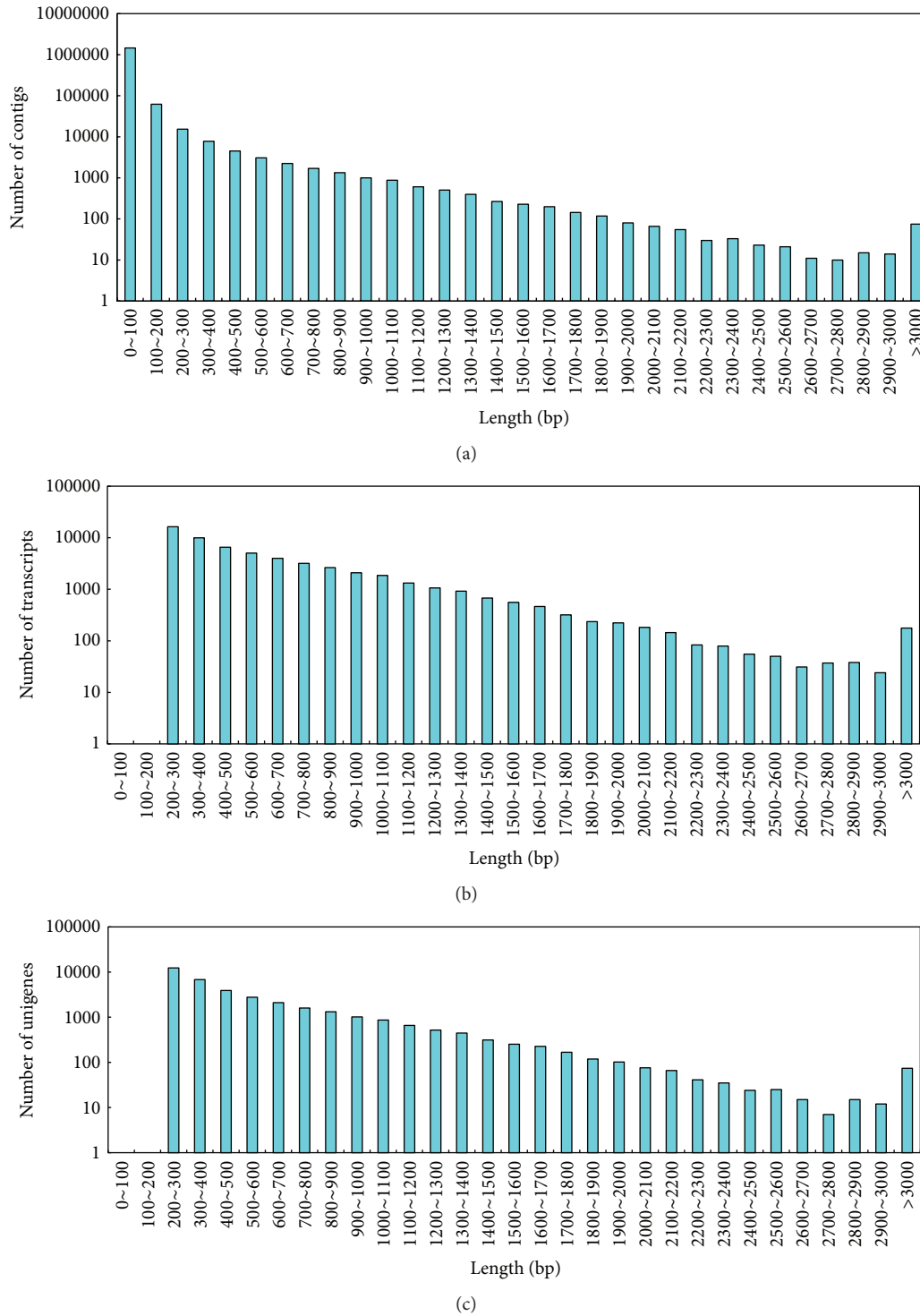
(a)

(b)

(c)

FIGURE 1: Overview of the sweet potato transcriptome assembly. (a) Size distribution of contigs; (b) size distribution of transcripts; and (c) size distribution of unigenes.

with an average length of 596 bp and N50 length of 767 bp. The transcripts were further clustered into 35,909 unigenes with an average length of 533 bp and N50 length of 669 bp (Table 1). The length distributions of contigs, transcripts, and unigenes were shown in Figure 1.

3.2. Identification and Functional Annotation of DEGs. According to the BLASTX results, most of the unigenes had homologous proteins in the Nr protein database. Interestingly, 4,903 (18.71%) and 4,463 (17.03%) unigenes showed significant homology with sequences of *Nicotiana sylvestris*

TABLE 1: Length distribution of assembled contigs, transcripts, and unigenes from Weiduoli and HVB-3.

Length range	Contig	Transcript	Unigene
0–300	1,531,538 (98.36%)	16,317 (28.00%)	12,325 (34.32%)
300–500	12,342 (0.79%)	16,504 (28.32%)	10,710 (29.83%)
500–1000	9,349 (0.60%)	16,921 (29.04%)	8,813 (24.54%)
1000–2000	3,419 (0.22%)	7,635 (13.10%)	3,671 (10.22%)
2000+	353 (0.02%)	900 (1.54%)	390 (1.09%)
Total number	1,557,001	58,277	35,909
Total length	91,371,759	34,741,399	19,150,802
Mean length	58	596	533
N50 length	58	767	669

(a)

Significant
· False
· True

(b)

FIGURE 3: Comparative analysis of gene expression in Weiduoli and HVB-3. (a) A scatter plot of RPKM logarithmic values in libraries of Weiduoli and HVB-3. Each dot represents the RPKM value of a specific gene. The greater deviation from the diagonal slope shows a greater expression level of the gene in the corresponding material. (b) A scatter plot of the ratio of RPKM logarithmic numerical values of genes in Weiduoli and HVB-3. This plot graphically represents genes differentially expressed between Weiduoli and HVB-3. Blue dots represent genes that had significant difference and red dots represent genes where no significant difference was observed between Weiduoli and HVB-3.

Legend:
- Nicotiana sylvestris [4903]
- Nicotiana tomentosiformis [4463]
- Solanum tuberosum [2656]
- Coffea canephora [2022]
- Solanum lycopersicum [1869]
- Sesamum indicum [1829]
- Vitis vinifera [948]
- Erythranthe guttata [538]
- Citrus sinensis [509]
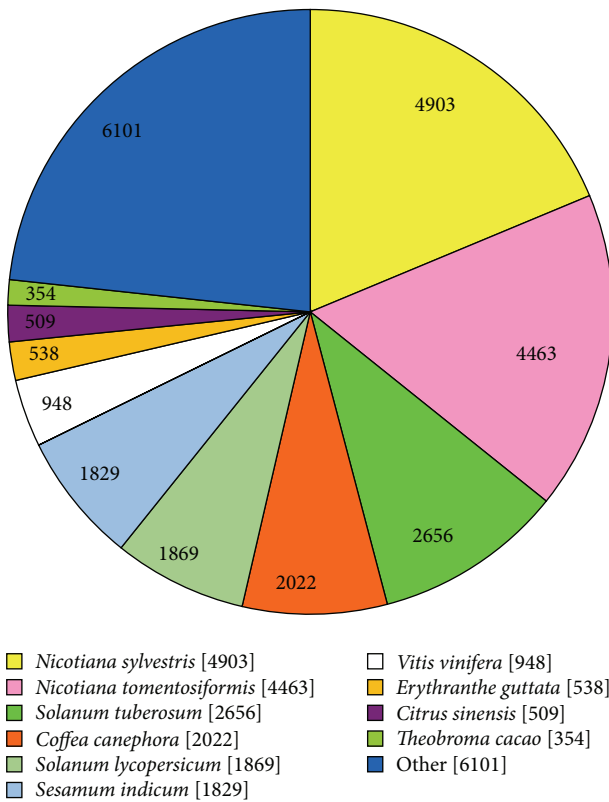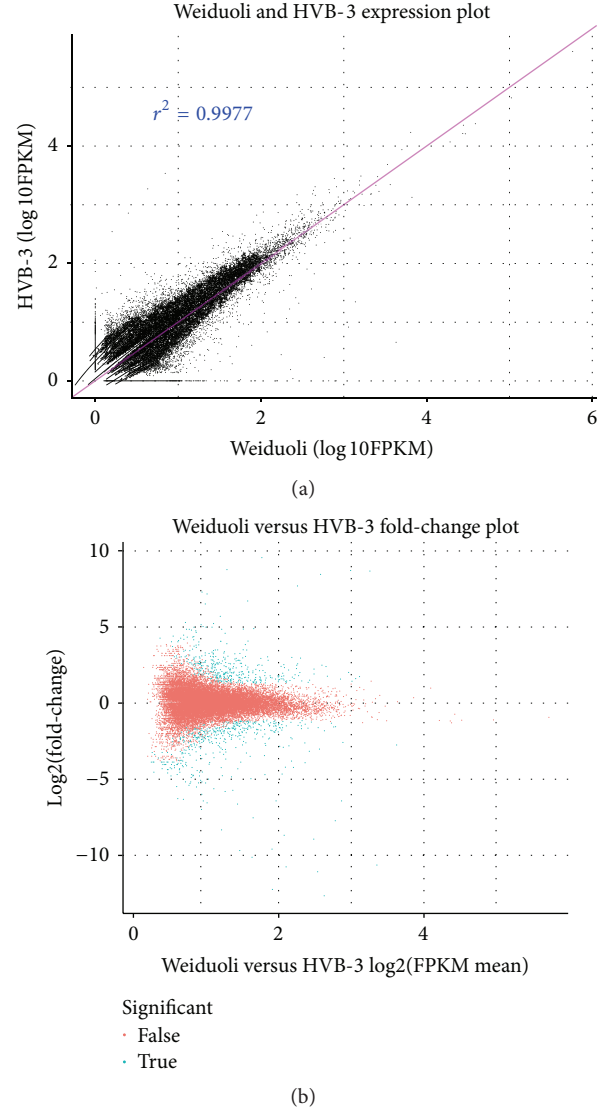- Theobroma cacao [354]
- Other [6101]

FIGURE 2: Species distribution of the top BlastX hits for each unigene in the Nr database.

and *Nicotiana tomentosiformis*, respectively (Figure 2). Furthermore, 14,316 (54.21%) unigenes had significant matches in the Pfam database, and 17,058 (64.59%) unigenes had similarity to proteins in the Swiss-Prot database.

The expression of unigenes was calculated according to the RPKM method. A total of 35,909 unigenes had detectable levels of expression in Weiduoli and HVB-3 (Figure 3(a)). Using the IDEG6 software, a total of 874 genes were found to be differentially expressed between HVB-3 and Weiduoli, and 401 of them were upregulated and 473 were downregulated in HVB-3 compared to Weiduoli (Figure 3(b)). A total of 697 DEGs were annotated against the public databases and

94.12% of them were mapped to the Nr library, suggesting that most of the DEGs can be translated into proteins. The mapping rate of DEGs against the Swiss-Prot protein database was 67.58%. The overall functional annotation is listed in Table 2.

A total of 14,136 unigenes were classified into three categories, cellular component, biological progress, and molecular function, through GO analysis (Figure 4). In all, 316 DEGs were classified into three categories, 149 with cellular component, 235 with biological progress, and 245 with
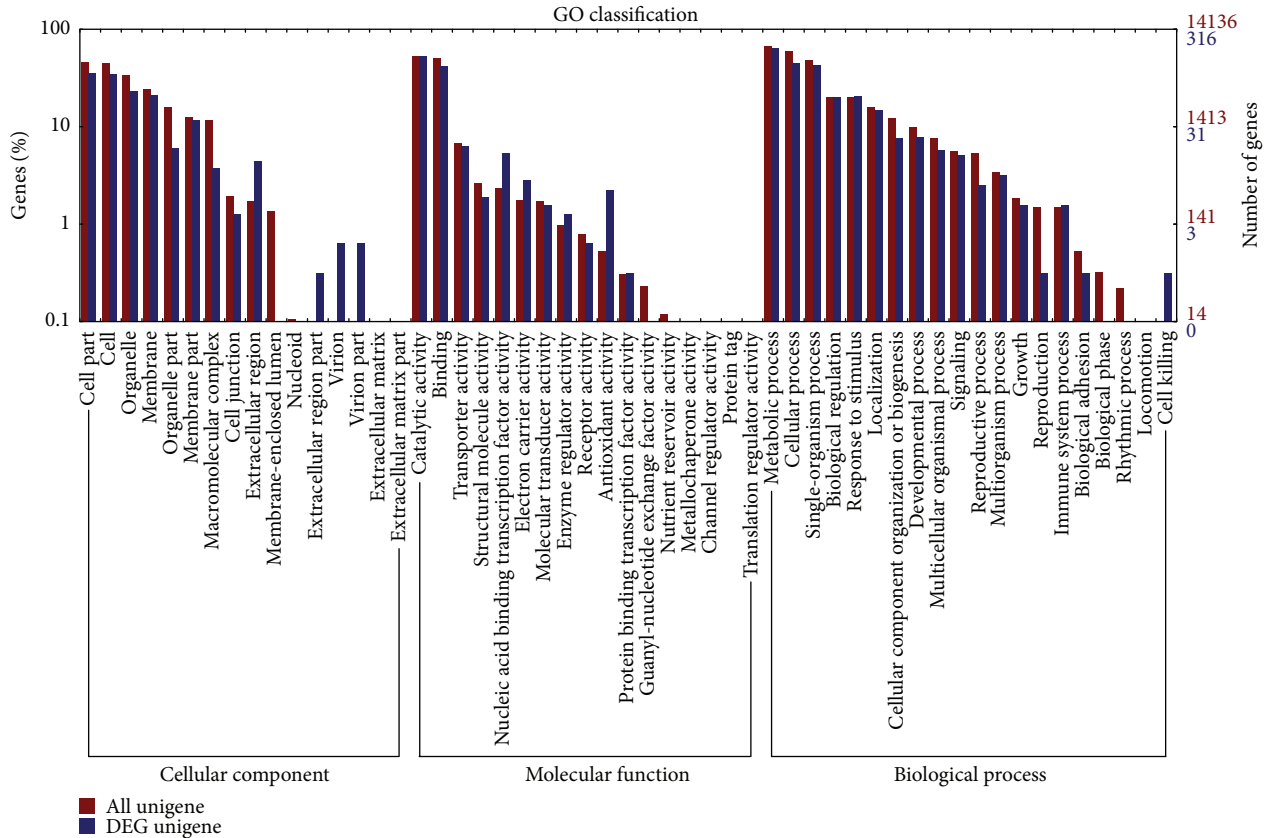
FIGURE 4: GO classification of unigenes in transcriptomes of Weiduoli and HVB-3. The red bars represent all the unigenes and the blue bars represent the DEGs.

TABLE 2: Functional annotation of DEGs between Weiduoli and HVB-3.

Annotation database	Annotation number
COG annotation	167
GO annotation	316
KEGG annotation	83
KOG annotation	290
Pfam annotation	477
Swiss-Prot annotation	471
Nr annotation	656
Total	697

molecular function, through GO analysis (Figure 4). The GO analysis revealed that most of the DEGs were involved in catalytic activity and metabolic process. Compared with the COG database, 240 DEGs were subdivided into 22 COG classifications, including secondary metabolite biosynthesis, transport, and catabolism, signal transduction mechanisms, replication, recombination, and repair, amino acid transport and metabolism, inorganic ion transport and metabolism, carbohydrate transport and metabolism, energy production and conversion, transcription, and lipid transport and metabolism (Figure 5).

Carotenoid biosynthesis which belongs to the secondary metabolisms is a dynamic and complex process catalyzed by a series of enzymes. Functional category analysis revealed that the DEGs were involved in a number of important pathways, including metabolite biosynthesis and signal transduction mechanisms (Figure 6), similar to the results of GO and COG analyses. According to the KEGG pathway enrichment results, 62 DEGs were assigned to the 50 pathways. The most noticeable pathways were terpenoid backbone biosynthesis and fatty acid metabolism. As shown in Table 3, 22 DEGs were found to be directly or indirectly involved in carotenoid biosynthesis. These 22 DEGs encoded geranylgeranyl pyrophosphate synthase (GGPS), geranylgeranyl diphosphate reductase (GGPR), dehydrodolichyl diphosphate synthase (DHDDS), alcohol dehydrogenases homologous, aldehyde dehydrogenase, alcohol dehydrogenase, long chain acyl-CoA synthetase, and 15 cytochrome P450, respectively (Table 3). Interestingly, several important transcription factors, including NAC, MYB, AP2/ERF, Zifc fingers, WRKY, bZIP, and ARF, were found to be significantly upregulated in HVB-3 compared to Weiduoli (Table 3).

3.3. SSR Markers. The MISA was used to search for SSRs. A total of 1,725 potential cDNA-derived SSRs (cSSRs) were identified from 4,061 unigenes. Most of them were mononucleotide repeats (1,005), followed by trinucleotide repeats

TABLE 3: Differentially expressed genes and transcription factors related to carotenoid biosynthesis between Weiduoli and HVB-3.

Gene ID	Log2 fold-change	FDR	Blast annotation
Genes			
c27179.graph_c0	2.28	0	Geranylgeranyl pyrophosphate synthase
c35889.graph_c0	−2.19	$6.71E − 08$	Geranylgeranyl diphosphate reductase
c41187.graph_c4	2.29	$3.36E − 11$	Dehydrodolichyl diphosphate synthase
c21698.graph_c0	−3.96	$1.68E − 05$	Alcohol dehydrogenases homologous
c30364.graph_c1	2.08	0	Aldehyde dehydrogenase
c34785.graph_c0	−2.46	$9.95E − 05$	Alcohol dehydrogenase
c37165.graph_c0	−2.81	$3.18E − 05$	Long chain acyl-CoA synthetase
c26028.graph_c0	−1.58	$7.58E − 05$	Cytochrome P450 82A2
c29728.graph_c0	2.96	0.01	Cytochrome P450 82A4
c29775.graph_c1	−2.20	0	Cytochrome P450 86B1-like
c33165.graph_c0	−2.03	0	Cytochrome p450 CYP82D47-like
c34761.graph_c0	1.77	0	Cytochrome P450 89A2
c34982.graph_c1	−2.66	$2.32E − 07$	Cytochrome P450 82C4
c36936.graph_c0	−2.50	$1.63E − 06$	Cytochrome p450 86B1-like
c38370.graph_c0	2.11	0	Cytochrome P450 CYP72A219
c38487.graph_c0	−1.67	0.01	Cytochrome P450 82C4
c40387.graph_c0	2.38	$7.46E − 08$	Cytochrome P450 83B1
c40784.graph_c0	−1.51	0	Cytochrome P450 CYP736A12
c41229.graph_c0	−2.70	$7.30E − 13$	Cytochrome P450 78A5
c41281.graph_c1	1.22	0.01	Cytochrome P450 76A2
c41491.graph_c0	1.32	0.01	Cytochrome P450 CYP72A219-like
c42321.graph_c0	3.40	0	Cytochrome P450 71A1
Transcription factors			
c39444.graph_c0	2.10	$5.87E − 11$	NAC domain-containing protein
c39928.graph_c0	2.14	$1.04E − 10$	NAC domain-containing protein
c41510.graph_c0	1.55	0	NAC domain-containing protein
c4748.graph_c0	3.66	$6.33E − 06$	MYB-like transcription factor
c27997.graph_c0	2.18	0	Ethylene-responsive transcription factor
c32755.graph_c0	3.98	$3.64E − 11$	Ethylene-responsive transcription factor
c32755.graph_c1	3.20	$5.42E − 09$	Ethylene-responsive transcription factor
c34264.graph_c0	1.92	0	Ethylene-responsive transcription factor
c34616.graph_c0	1.27	0.01	Ethylene-responsive transcription factor
c35745.graph_c1	2.64	$1.63E − 06$	Ethylene-responsive transcription factor
c36135.graph_c0	1.32	0	Ethylene-responsive transcription factor
c37886.graph_c1	1.41	0	Ethylene-responsive transcription factor
c39393.graph_c0	1.42	0	Ethylene-responsive transcription factor
c40060.graph_c0	1.45	0	Ethylene-responsive transcription factor
c40767.graph_c1	1.48	$8.56E − 05$	Ethylene-responsive transcription factor
c41013.graph_c0	1.59	$4.20E − 05$	Ethylene-responsive transcription factor
c29615.graph_c0	1.91	0	Zinc finger protein
c30636.graph_c0	1.56	0.01	Zinc finger protein
c33159.graph_c0	1.63	0.01	Zinc finger protein
c33843.graph_c0	1.91	$2.63E − 05$	Zinc finger protein
c34456.graph_c0	1.43	0	Zinc finger protein
c35956.graph_c0	3.37	0	Zinc finger protein
c36278.graph_c0	1.83	0	Zinc finger protein
c36428.graph_c1	1.78	$9.34E − 06$	Zinc finger protein
c33226.graph_c1	2.30	$9.80E − 05$	WRKY transcription factor
c34181.graph_c0	1.59	0.01	WRKY transcription factor
c39751.graph_c0	1.36	0.01	WRKY transcription factor
c35456.graph_c0	1.71	0	bZIP transcription factor
c39175.graph_c0	1.55	0	bZIP transcription factor
c41663.graph_c1	1.95	$1.82E − 08$	bZIP transcription factor
c19855.graph_c0	2.66	0	Auxin response factor

COG function classification of consensus sequence

A: RNA processing and modification
B: chromatin structure and dynamics
C: energy production and conversion
D: cell cycle control, cell division, and chromosome partitioning
E: amino acid transport and metabolism
F: nucleotide transport and metabolism
G: carbohydrate transport and metabolism
H: coenzyme transport and metabolism
I: lipid transport and metabolism
J: translation, ribosomal structure, and biogenesis
K: transcription
L: replication, recombination, and repair
M: cell wall/membrane/envelope biogenesis
N: cell motility
O: posttranslational modification, protein turnover, and chaperones
P: inorganic ion transport and metabolism
Q: secondary metabolites biosynthesis, transport, and catabolism
R: general function prediction only
S: function unknown
T: signal transduction mechanisms
U: intracellular trafficking, secretion, and vesicular transport
V: defense mechanisms
W: extracellular structures
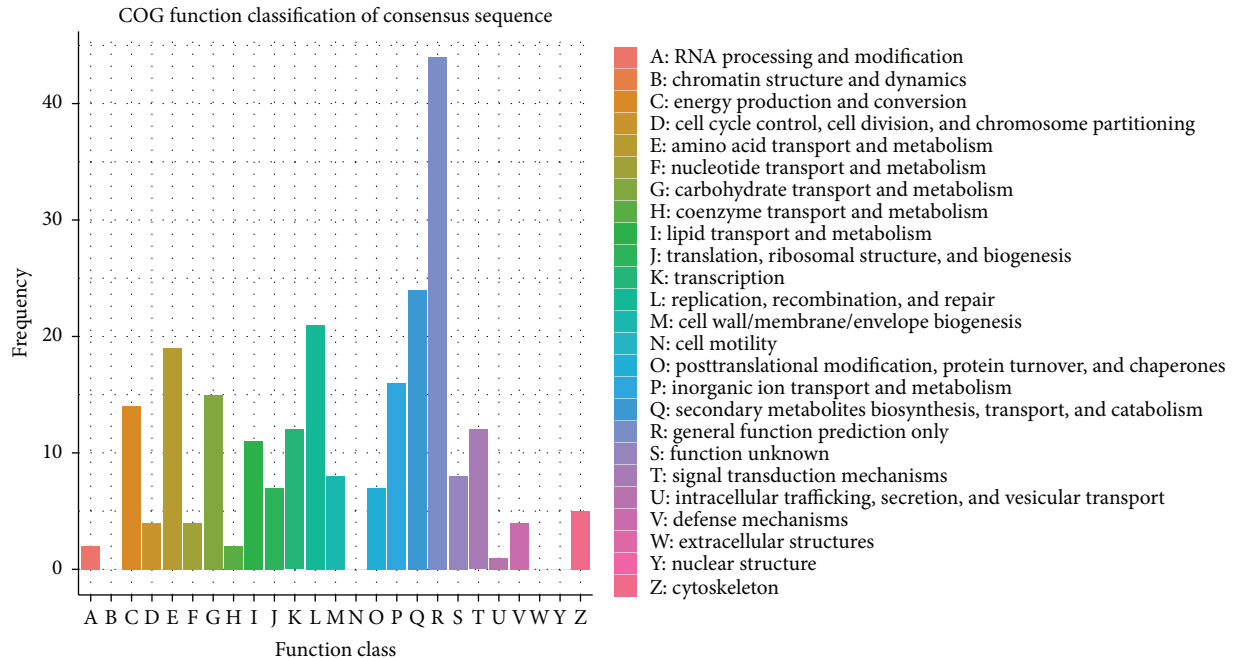Y: nuclear structure
Z: cytoskeleton

Figure 5: COG-based functional classification of DEGs between Weiduoli and HVB-3.

(388), dinucleotide repeats (298), and tetranucleotide repeats (26), with only a small portion of pentanucleotide (4) and hexanucleotide repeats (4). There were 323 sequences containing more than 1 cSSR and 125 cSSRs present in compound formation.

4. Discussion

In nonmodel plants, it is difficult to identify the candidate genes involved in complex biosynthetic pathways due to the limited availability of genomic data [28, 29]. With high-throughput transcriptome sequencing technology, this limitation has been overcome, as it can generate large amounts of data on genome wide transcription [30]. Several sweet potato transcriptomes have been sequenced, which provide an important data source for storage root formation, flower development, and anthocyanin biosynthesis of this crop [7, 18–22].

Carotenoids are widely distributed pigments in plants and play an important role as light-harvesting pigments in most photosynthetic organisms [31, 32]. In many photosynthetic and nonphotosynthetic organisms, the carotenoid biosynthesis pathway has been well studied and a series of genes involved in this pathway have been cloned and characterized. However, the molecular mechanisms regulating carotenoid synthesis have not been well understood.

The storage roots of orange-fleshed sweet potato typically have a high carotenoid content [5]. In the present study, the transcriptomes of orange-fleshed sweet potato cultivar "Weiduoli" and its high carotenoid mutant "HVB-3" were sequenced on the Illumina HiSeq 2000 sequencing platform,

and 13,767,387 and 9,837,090 high-quality reads were produced from Weiduoli and HVB-3, respectively. A total of 35,909 unigenes were harvested from Weiduoli and HVB-3 (Table 1). There were 874 DEGs between HVB-3 and Weiduoli, 401 of which were upregulated and 473 were downregulated in HVB-3 compared to Weiduoli (Figure 3(b)).

The present results showed that the 22 DEGs related to carotenoid biosynthesis existed between Weiduoli and HVB-3 (Table 3). GGPS, GGPR, and DHDDS are involved in terpenoid backbone biosynthesis. GGPS is the key enzyme of carotenoid biosynthesis. Transgenic kiwifruit plants expressing GGPS exhibited the increased β-carotene content [33]. GGPR converts geranylgeranyl diphosphate (GGPP), the precursor for carotenoid biosynthesis, to phytyl diphosphate in the tocopherol and chlorophyll biosynthetic pathways. DHDDS is involved in the biosynthesis of isoprenoids, which are the precursors of carotenoid biosynthesis. Four genes encoding alcohol dehydrogenases homologous, aldehyde dehydrogenase, alcohol dehydrogenase, and long chain acyl-CoA synthetase, respectively, are involved in fatty acid metabolism. The biosynthesis of carotenoids and fatty acids requires a common precursor from pyruvate [34]. The 15 DEGs were found to encode the cytochrome P450 family (Table 3). *P450CYP707A* encoding ABA 8′-hydroxylases and *LUT1* encoding cytochrome P450-type monooxygenase (CYP97C1) have been proved to regulate carotenoid biosynthesis in *Arabidopsis* [35, 36]. Thus, it is thought that these DEGs may play important roles in carotenoid biosynthesis of sweet potato.

In the present study, several important transcription factors, including NAC, MYB, AP2/ERF, Zifc fingers, WRKY, bZIP, and ARF, were significantly upregulated in HVB-3 compared to Weiduoli (Table 3). These transcription factors may
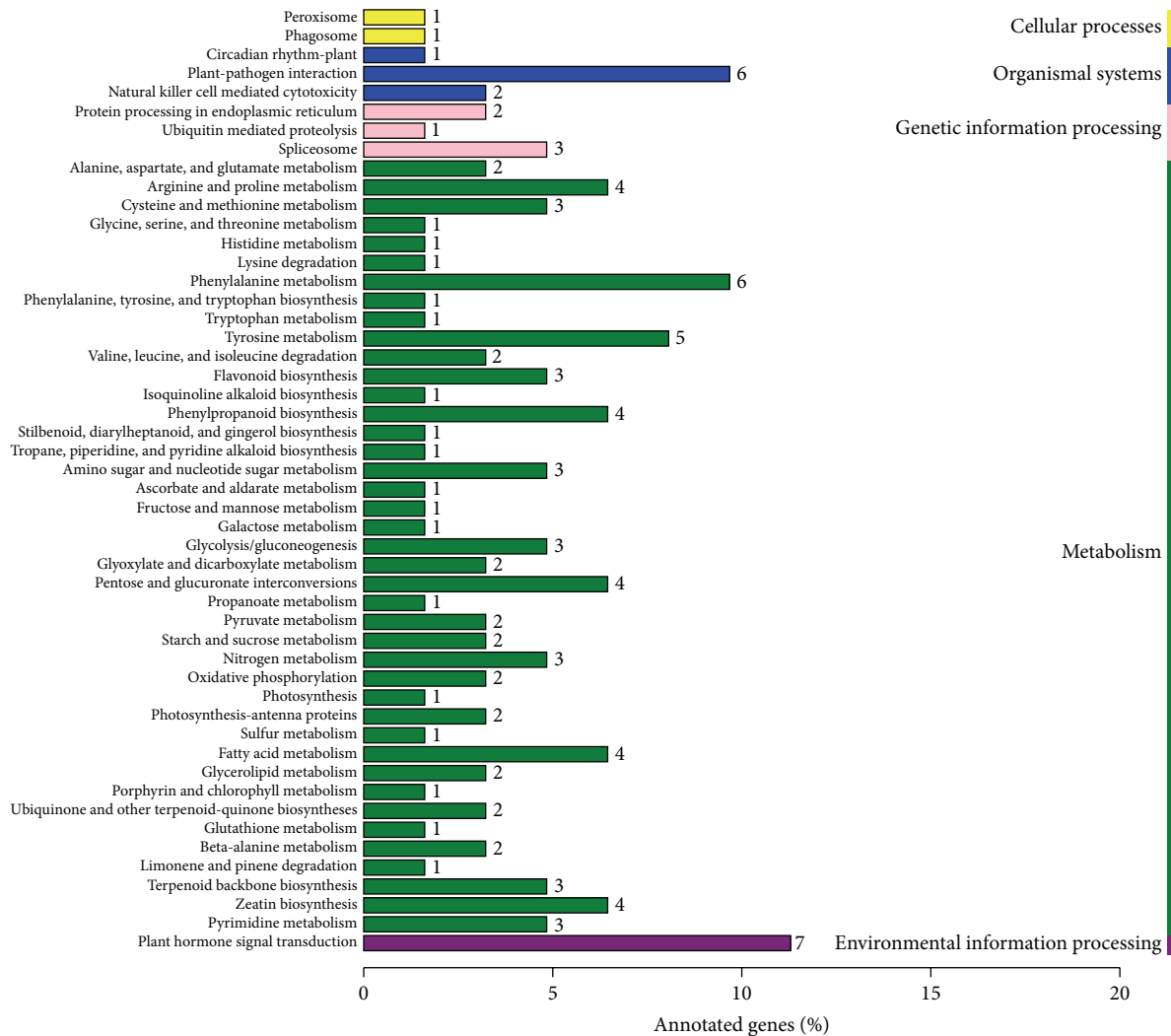
FIGURE 6: KEGG-based functional classification of DEGs between Weiduoli and HVB-3. Numbers beside each bar represent the actual number of DEGs classified in that descriptive term.

regulate carotenoid biosynthesis of sweet potato. In plants, transcription factors of different families have been shown to regulate secondary metabolism pathways [37]. NAC proteins are one of the largest families of plant-specific transcription factors [38]. In *Solanum lycopersicum*, a NAC transcription factor (SlNAC4) was shown to function as a positive regulator of carotenoid accumulation [39]. MYB transcription factors were found to participate in a wide range of biological processes [40, 41]. Overexpression of the *Vitis vinifera* MYB5b in tomato resulted in an increased content of β-carotene [42].

To date, the genome of sweet potato is still unavailable; therefore much of the research on sweet potato breeding and genetic linkage maps is based on molecular markers [43]. SSRs are widely distributed in both noncoding and transcribed sequences. As transferable markers, SSRs are a useful source for genome analysis due to their abundance, functionality, high polymorphism, and excellent reproducibility [44]. In the present study, a total of 1,725 potential cSSRs, including

mono-, di-, tri-, tetra-, penta-, and hexa-SSR, were identified from 4,061 unigenes.

5. Conclusion

A total of 35,909 unigenes were identified from the orange-fleshed sweet potato using the high-throughput sequencing technology and most of them are protein-coding genes. There were 874 DEGs between Weiduoli and HVB-3. The 22 DEGs were found to directly or indirectly participate in carotenoid biosyntheses. The 31 important transcription factors were significantly upregulated in HVB-3 compared to Weiduoli. These DEGs and transcription factors may be involved in carotenoid biosynthesis of sweet potato.

Conflict of Interests

The authors declare no conflict of interests.

Acknowledgments

References

[1] Y. O. Ahn, S. H. Kim, C. Y. Kim, J.-S. Lee, S.-S. Kwak, and H.-S. Lee, "Exogenous sucrose utilization and starch biosynthesis among sweetpotato cultivars," *Carbohydrate Research*, vol. 345, no. 1, pp. 55–60, 2010.

[2] G. Padmaja, "Uses and nutritional data of sweetpotato," *The Sweetpotato*, pp. 189–234, 2009.

[3] T. Oki, M. Masuda, S. Furuta, Y. Nishida, N. Terahara, and I. Suda, "Involvement of anthocyanins and other phenolic compounds in radical-scavenging activity of purple-fleshed sweet potato cultivars," *Journal of Food Science*, vol. 67, no. 5, pp. 1752–1756, 2002.

[4] N. Zang, H. Zhai, S. Gao, W. Chen, S. Z. He, and Q. Liu, "Efficient production of transgenic plants using the bar gene for herbicide resistance in sweetpotato," *Scientia Horticulturae*, vol. 122, no. 4, pp. 649–653, 2009.

[5] M. M. Manifesto, S. M. Costa Tártara, C. M. Arizio, M. A. Alvarez, and N. R. Hompanera, "Analysis of the morphological attributes of a sweetpotato collection," *Annals of Applied Biology*, vol. 157, no. 2, pp. 273–281, 2010.

[6] L. Yu, H. Zhai, W. Chen, S.-Z. He, and Q.-C. Liu, "Cloning and functional analysis of lycopene ε-cyclase (*IbLCYe*) gene from sweetpotato, *Ipomoea batatas* (L.) Lam.," *Journal of Integrative Agriculture*, vol. 12, no. 5, pp. 773–780, 2013.

[7] X. Tao, Y.-H. Gu, H.-Y. Wang et al., "Digital gene expression analysis based on integrated *De Novo* transcriptome assembly of sweet potato [*Ipomoea batatas* (L.) Lam.]," *PLoS ONE*, vol. 7, no. 4, Article ID e36234, 2012.

[8] C. I. Cazzonelli and B. J. Pogson, "Source to sink: regulation of carotenoid biosynthesis in plants," *Trends in Plant Science*, vol. 15, no. 5, pp. 266–274, 2010.

[9] M. E. Auldridge, D. R. McCarty, and H. J. Klee, "Plant carotenoid cleavage oxygenases and their apocarotenoid products," *Current Opinion in Plant Biology*, vol. 9, no. 3, pp. 315–321, 2006.

[10] C. K. Shewmaker, J. A. Sheehy, M. Daley, S. Colburn, and D. Y. Ke, "Seed-specific overexpression of phytoene synthase: increase in carotenoids and other metabolic effects," *Plant Journal*, vol. 20, no. 4, pp. 401–412, 1999.

[11] S. Römer, P. D. Fraser, J. W. Kiano et al., "Elevation of the provitamin A content of transgenic tomato plants," *Nature Biotechnology*, vol. 18, no. 6, pp. 666–669, 2000.

[12] J. A. Paine, C. A. Shipton, S. Chaggar et al., "Improving the nutritional value of Golden Rice through increased pro-vitamin A content," *Nature Biotechnology*, vol. 23, no. 4, pp. 482–487, 2005.

[13] T. C. H. Tran, S. Al-Babili, P. Schaub, I. Potrykus, and P. Beyer, "Approaches to improve *CrtI* expression in rice endosperm for increasing the β-carotene content in golden rice," *OmonRice*, no. 13, pp. 1–11, 2005.

[14] S. H. Kim, Y. O. Ahn, M.-J. Ahn, H.-S. Lee, and S.-S. Kwak, "Down-regulation of β-carotene hydroxylase increases β-carotene and total carotenoids enhancing salt stress tolerance in transgenic cultured cells of sweetpotato," *Phytochemistry*, vol. 74, pp. 69–78, 2012.

[15] S. H. Kim, Y.-H. Kim, Y. O. Ahn et al., "Downregulation of the lycopene ε-cyclase gene increases carotenoid synthesis via the β-branch-specific pathway and enhances salt-stress tolerance in sweetpotato transgenic calli," *Physiologia Plantarum*, vol. 147, no. 4, pp. 432–442, 2013.

[16] S. H. Kim, J. C. Jeong, S. Park et al., "Down-regulation of sweetpotato lycopene β-cyclase gene enhances tolerance to abiotic stress in transgenic calli," *Molecular Biology Reports*, vol. 41, no. 12, pp. 8137–8148, 2014.

[17] S.-C. Park, S. H. Kim, S. Park et al., "Enhanced accumulation of carotenoids in sweetpotato plants overexpressing *IbOr-Ins* gene in purple-fleshed sweetpotato cultivar," *Plant Physiology and Biochemistry*, vol. 86, pp. 82–90, 2015.

[18] R. Schafleitner, L. R. Tincopa, O. Palomino et al., "A sweetpotato gene index established by de novo assembly of pyrosequencing and Sanger sequences and mining for gene-based microsatellite markers," *BMC Genomics*, vol. 11, no. 1, article 604, 2010.

[19] Z. Y. Wang, B. P. Fang, J. Y. Chen et al., "De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (*Ipomoea batatas*)," *BMC Genomics*, vol. 11, article 726, 2010.

[20] F. L. Xie, C. E. Burklew, Y. F. Yang et al., "De novo sequencing and a comprehensive analysis of purple sweet potato (*Impomoea batatas* L.) transcriptome," *Planta*, vol. 236, no. 1, pp. 101–113, 2012.

[21] N. Firon, D. LaBonte, A. Villordon et al., "Transcriptional profiling of sweetpotato (*Ipomoea batatas*) roots indicates down-regulation of lignin biosynthesis and up-regulation of starch biosynthesis at an early stage of storage root formation," *BMC Genomics*, vol. 14, no. 1, article 460, pp. 1471–2164, 2013.

[22] X. Tao, Y.-H. Gu, Y.-S. Jiang, Y.-Z. Zhang, and H.-Y. Wang, "Transcriptome analysis to identify putative floral-specific genes and flowering regulatory-related genes of sweet potato," *Bioscience, Biotechnology and Biochemistry*, vol. 77, no. 11, pp. 2169–2174, 2013.

[23] M. G. Grabherr, B. J. Haas, M. Yassour et al., "Full-length transcriptome assembly from RNA-Seq data without a reference genome," *Nature Biotechnology*, vol. 29, no. 7, pp. U644–U652, 2011.

[24] W. J. Kent, "BLAT—the BLAST-like alignment tool," *Genome Research*, vol. 12, no. 4, pp. 656–664, 2002.

[25] A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, and B. Wold, "Mapping and quantifying mammalian transcriptomes by RNA-Seq," *Nature Methods*, vol. 5, no. 7, pp. 621–628, 2008.

[26] C. Romualdi, S. Bortoluzzi, F. D'Alessi, and G. A. Danieli, "IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments," *Physiological Genomics*, vol. 12, no. 2, pp. 159–162, 2003.

[27] A. Conesa, S. Götz, J. M. García-Gómez, J. Terol, M. Talón, and M. Robles, "Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research," *Bioinformatics*, vol. 21, no. 18, pp. 3674–3676, 2005.

[28] Q. Tang, X. J. Ma, C. M. Mo et al., "An efficient approach to finding *Siraitia grosvenorii* triterpene biosynthetic genes by RNA-seq and digital gene expression analysis," *BMC Genomics*, vol. 12, article 343, 2011.

[29] K. K. Myung, B.-S. Lee, J.-G. In, H. Sun, J.-H. Yoon, and D.-C. Yang, "Comparative analysis of expressed sequence tags (ESTs) of ginseng leaf," *Plant Cell Reports*, vol. 25, no. 6, pp. 599–606, 2006.

[30] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics," *Nature Reviews Genetics*, vol. 10, no. 1, pp. 57–63, 2009.

[31] H. A. Frank and R. J. Cogdell, "Carotenoids in photosynthesis," *Photochemistry and Photobiology*, vol. 63, no. 3, pp. 257–264, 1996.

[32] G. E. Bartley and P. A. Scolnik, "Plant carotenoids: pigments for photoprotection, visual attraction, and human health," *Plant Cell*, vol. 7, no. 7, pp. 1027–1038, 1995.

[33] M. Kim, S.-C. Kim, K. J. Song et al., "Transformation of carotenoid biosynthetic genes using a micro-cross section method in kiwifruit (*Actinidia deliciosa* cv. Hayward)," *Plant Cell Reports*, vol. 29, no. 12, pp. 1339–1349, 2010.

[34] J. Schwender, M. Seemann, H. K. Lichtenthaler, and M. Rohmer, "Biosynthesis of isoprenoids (carotenoids, sterols, prenyl side-chains of chlorophylls and plastoquinone) via a novel pyruvate/glyceraldehyde 3-phosphate non-mevalonate pathway in the green alga *Scenedesmus obliquus*," *Biochemical Journal*, vol. 316, no. 1, pp. 73–80, 1996.

[35] T. Kushiro, M. Okamoto, K. Nakabayashi et al., "The *Arabidopsis* cytochrome P450 CYP707A encodes ABA $8'$-hydroxylases: key enzymes in ABA catabolism," *The EMBO Journal*, vol. 23, no. 7, pp. 1647–1656, 2004.

[36] L. Tian, V. Musetti, J. Kim, M. Magallanes-Lundback, and D. DellaPenna, "The *Arabidopsis* LUT1 locus encodes a member of the cytochrome P450 family that is required for carotenoid ε-ring hydroxylation activity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 1, pp. 402–407, 2004.

[37] N. De Geyter, A. Gholami, S. Goormachtig, and A. Goossens, "Transcriptional machineries in jasmonate-elicited plant secondary metabolism," *Trends in Plant Science*, vol. 17, no. 6, pp. 349–359, 2012.

[38] S. Puranik, P. P. Sahu, P. S. Srivastava, and M. Prasad, "NAC proteins: regulation and role in stress tolerance," *Trends in Plant Science*, vol. 17, no. 6, pp. 369–381, 2012.

[39] M. K. Zhu, G. P. Chen, S. Zhou et al., "A new tomato NAC (NAM/ATAF1/2/CUC2) transcription factor, SlNAC4, functions as a positive regulator of fruit ripening and carotenoid accumulation," *Plant and Cell Physiology*, vol. 55, no. 1, pp. 119–135, 2014.

[40] A. M. Takos, F. W. Jaffé, S. R. Jacob, J. Bogs, S. P. Robinson, and A. R. Walker, "Light-induced expression of a *MYB* gene regulates anthocyanin biosynthesis in red apples," *Plant Physiology*, vol. 142, no. 3, pp. 1216–1232, 2006.

[41] R. V. Espley, R. P. Hellens, J. Putterill, D. E. Stevenson, S. Kutty-Amma, and A. C. Allan, "Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10," *Plant Journal*, vol. 49, no. 3, pp. 414–427, 2007.

[42] A. Mahjoub, M. Hernould, J. Joubès et al., "Overexpression of a grapevine R2R3-MYB factor in tomato affects vegetative development, flower morphology and flavonoid and terpenoid metabolism," *Plant Physiology and Biochemistry*, vol. 47, no. 7, pp. 551–561, 2009.

[43] M. I. Buteler, R. L. Jarret, and D. R. LaBonte, "Sequence characterization of microsatellites in diploid and polyploid *Ipomoea*," *Theoretical and Applied Genetics*, vol. 99, no. 1-2, pp. 123–132, 1999.

[44] W. Powell, M. Morgante, C. Andre et al., "The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis," *Molecular Breeding*, vol. 2, no. 3, pp. 225–238, 1996.