

Research Report

Open Access

Transcriptome Profiling of *Lisianthus (Eustoma grandiflorum)* under Drought Stress

An Xia¹ ✉, Zhu Qiang¹, Lou Xuping², Li Lufeng², Chen Jie³, Liu Tingting¹, Li Wenlue¹, Luo Xiaohong¹, Zhu Guanlin¹, Yu Lijun¹
1 Zhejiang Xiaoshan Institute of Cotton & Bast Fiber Crops Research, Zhejiang Institute of Landscape Plants and Flowers, Hangzhou, 311251, China
2 Hangzhou Xiaoshan District Agricultural Science and Technology Research Institute, Hangzhou, 311202, China
3 Huazhong Agricultural University, Wuhan, 430070, China

✉ Corresponding author email: anxia@zaas.ac.cn

Molecular Plant Breeding, 2020, Vol.11, No.31 doi: [10.5376/mpb.2020.11.0031](https://doi.org/10.5376/mpb.2020.11.0031)

Received: 27 Oct., 2020

Accepted: 18 Nov., 2020

Published: 26 Dec., 2020

Copyright © 2020 An et al., This article was first published in Molecular Plant Breeding in Chinese, and here was authorized to translate and publish the paper in English under the terms of Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:

An X., Zhu Q., Lou X.P., Li L.F., Chen J., Liu T.T., Li W.L., Luo X.H., Zhu G.L., and Yu L.J., 2020, Transcriptome profiling of *Lisianthus (Eustoma grandiflorum)* under drought stress, Molecular Plant Breeding, 11(31): 1-7 (doi: [10.5376/mpb.2020.11.0031](https://doi.org/10.5376/mpb.2020.11.0031))

Abstract *Lisianthus (Eustoma grandiflorum)* is an important ornamental plant native to North America. So far, the number of basic molecular research reports on *Lisianthus* is relatively small, especially the research related to the molecular mechanism of response to drought stress. In this experiment, transcriptome second-generation sequencing technology was used to study *Lisianthus* seedlings under drought stress. The results showed that the transcriptome sequencing data had good quality control results. In order to obtain genetic information on this basis, the sequences were spliced using software such as Trinity and Corset, a total of 102,014 non-redundant genes were obtained, including 2,929 coding genes which were found to belong to 79 transcription factor families after analyzing. Annotating all genes, *Zhongguo coffee* was found to be the species with the highest comparison similarity. Finally, simple sequence repeat (SSR) analysis was performed on the gene sequences, and all SSR information contained in non-redundant genes and transcription factor coding genes had been obtained. The results of this study can provide candidate molecular resources for subsequent studies on *Lisianthus*'s response to drought stress in the future.

Keywords *Lisianthus*; Drought stress; Transcriptome

Lisianthus (Eustoma grandiflorum), also known as steppe gentian, is native to North America and is a perennial herb of the Gentianaceae family. *Lisianthus* is an important ornamental plant, its floral organs are beautiful in shape and its vase life is long, it has become one of the more and more important fresh cut flowers in the world. *Lisianthus* is very sensitive to the cultivation environment, and the cultivation facilities and technology in China are relatively insufficient. Studying the relevant molecular mechanisms during the growth and development process of *Lisianthus* is helpful to provide theoretical support for the development of new cultivation techniques for *Lisianthus*.

In the process of response to drought stress, the differential expression of some functional genes and regulatory genes in plant is involved, forming a complex signal regulation network, which affects a series of physiological and biochemical reactions in plants. Drought can affect the elongation of *Lisianthus* flower stems, but so far, there are few research reports on how *Lisianthus* responds to drought stress. In this study, a total of 6.43 Gb of high-quality sequencing data was obtained through transcriptome sequencing of *Lisianthus* plants under drought stress treatment. 102,014 genes were obtained by splicing the sequencing results. The obtained genes were annotated, and they were found to be far apart from other species with good research foundations. Therefore, this study can not only provide a large number of molecular resources for subsequent related research on *Lisianthus*, but also provide a reference for high-throughput sequencing studies on other species of Gentianaceae.

1 Results and Analysis

1.1 Acquisition, processing and splicing of transcriptome data

First, strict quality control was performed on the transcriptome sequencing data in this study by mainly removing linkers and low-quality sequence information, and a total of 6.43 Gb of high-quality sequencing data (Clean reads)

was obtained. Then, the high-quality sequencing data obtained in this study were spliced using Trinity software (Grabherr et al., 2011), and a total of 132,929 transcripts (Transcripts) were obtained. Subsequently, the acquired high-quality sequencing data was analyzed by using Corset software (Davidson and Oshlack, 2014) to compare all reads with transcripts and perform hierarchical clustering, and 102,014 non-redundant genes (Unigenes) were obtained. These transcripts and genes had similar distribution patterns in terms of sequence length (Figure 1A; Figure 1B), and the quantitative difference between transcripts and non-redundant genes was mainly reflected in shorter sequences (<500 bp) (Figure 1C). This result showed that the quality of this sequencing result was high, mainly because there was no significant difference between the number of transcripts and non-redundant genes in the longer sequence (>1000 bp).

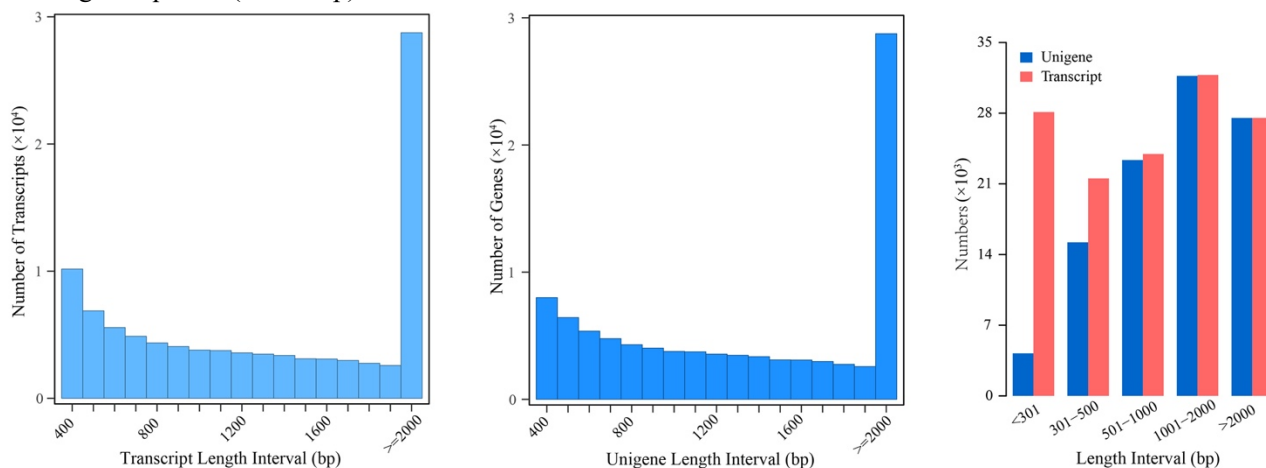


Figure 1 Lengths distribution from transcriptomic data

Note: A, Lengths distribution of transcripts; B, Lengths distribution of unigenes; C, Statistical results of transcripts and unigenes

1.2 Gene annotation and functional classification

In order to analyze the gene function information involved in the sequence obtained by transcriptome sequencing more scientifically and systematically, the 102,014 spliced genes in this study were annotated separately using different public databases. The statistical information of genes annotated in different databases was shown in Table 1, among them, 79.94% of genes had annotated results in at least one database. The number of genes annotated from the Nr database was the largest, accounting for 76.52% of all genes. The gene annotation results from five databases including Nr, Nt, pfam, GO and KOG were selected and analyzed, the results showed that 13,035 genes were specifically annotated in the Nr database, far more than the 741 genes specifically annotated in the Nt database and the 8 genes specifically annotated in the KOG database, and there were no genes specifically annotated in the pfam database or GO database (Figure 2A). Therefore, the annotation information of transcriptome data in the Nr database was more comprehensive. The annotation results in the Nr database were further analyzed, nearly half (47.3%) of the sequences had a high degree of similarity (over 80%, Figure 2B) with the target sequence, and there were a large number of sequences (accounting for 60.3%) with the comparison results e value less than 1e-60 (Figure 2C). Among the species annotation results of these sequences, Zhongguo coffee (*Coffea canephora*) was the most compared species (accounting for 39.7%), and there was a larger proportion of sequences (44.1%) aligned to other (Other) species (Figure 2D). In the KOG database classification, there were 25 KOG categories annotated with different numbers of genes, containing 28,696 genes in total, among them, the two most annotated to genes categories were O: post-translational modification: protein switches and molecular chaperones (3519 genes) and R: overall function prediction (3421 genes, Figure 3), this result was similar to the transcriptome result of jute drought stress (Jin et al., 2019). At the same time, from a metabolic perspective, these genes were more enriched in three metabolic items such as the "translation" of the "genetic information processing" category, "carbohydrate metabolism" of the "metabolism" category, and "folding, arrangement and degradation" of the "genetic information processing" category (Figure 4). Finally, since the transcription factors were often located in the upstream of gene expression pathways, they can regulate the

expression of a series of downstream genes to influence the degree of *Lisianthus* response to drought stress in a greater extent. Correspondingly, a total of 2,929 genes in the transcriptome sequencing results might encode transcription factors which belong to 79 different transcription factor families (Figure 5). Among these transcription factor families, the top three families with the largest number of predicted genes were bHLH family (256 genes), MYB_related family (216 genes) and bZIP family (191 genes).

Table 1 Statistical numbers on successfully annotated genes against different databases

Databases	NR	NT	KO	SwissProt	PFAM	GO	KOG
Number of Genes	78070	50844	34057	60288	55628	55628	25389
Percentage (%)	76.52	49.84	33.38	59.09	54.52	54.52	24.88

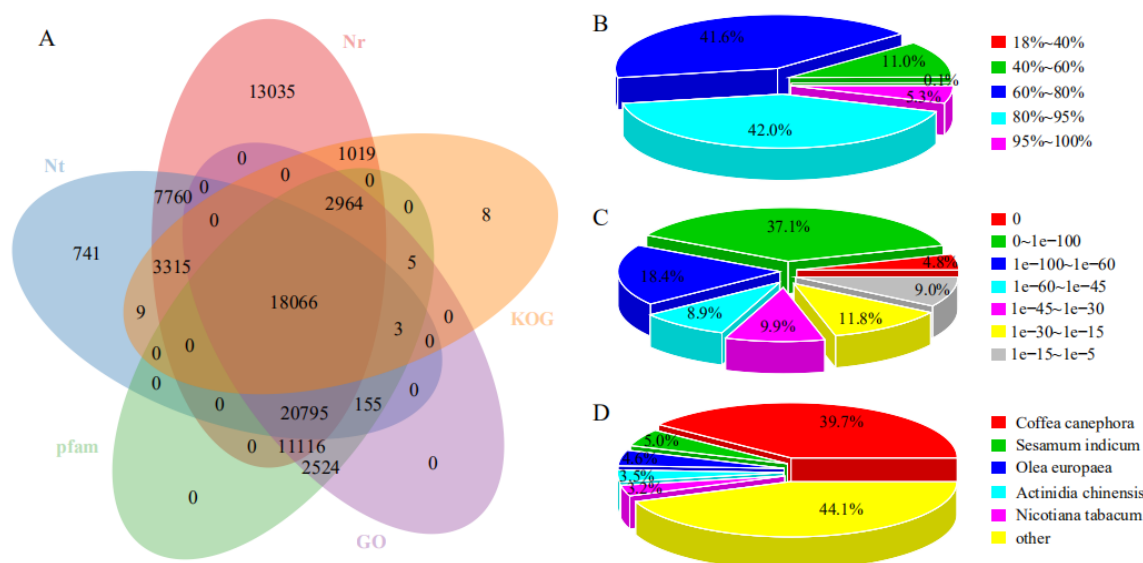


Figure 2 Annotation of transcriptome

Note: A, Numbers of unigenes annotated by different databases; B, Distribution of sequence similarities against the Nr database; C, Distribution of e-values against the Nr database; D, The most annotated species from our transcriptome data

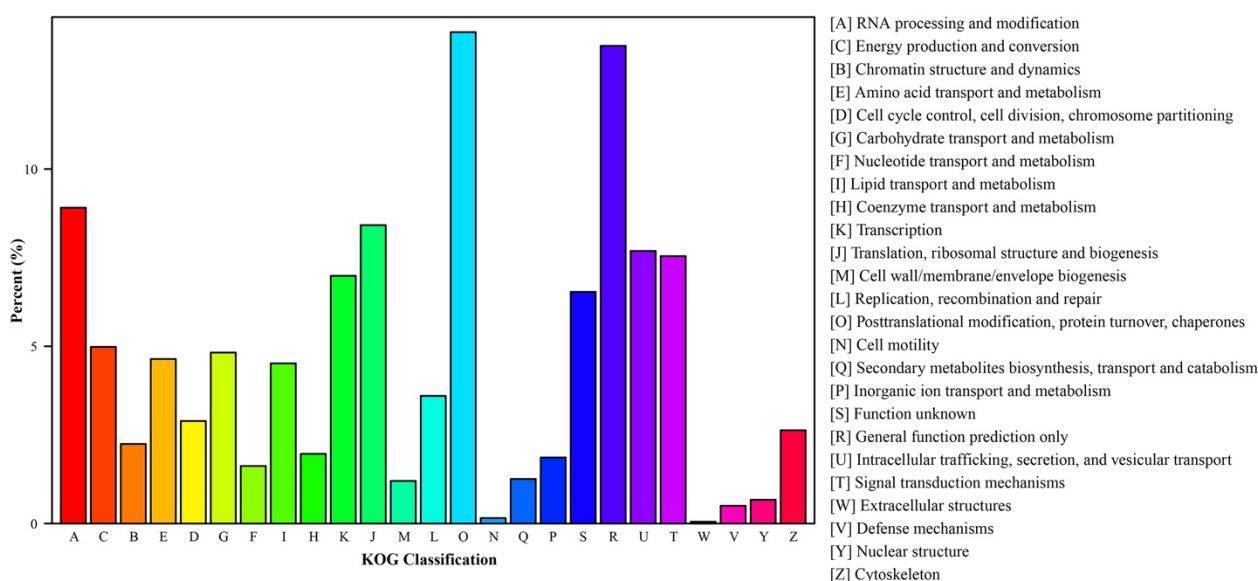


Figure 3 The KOG classification of our transcriptome data

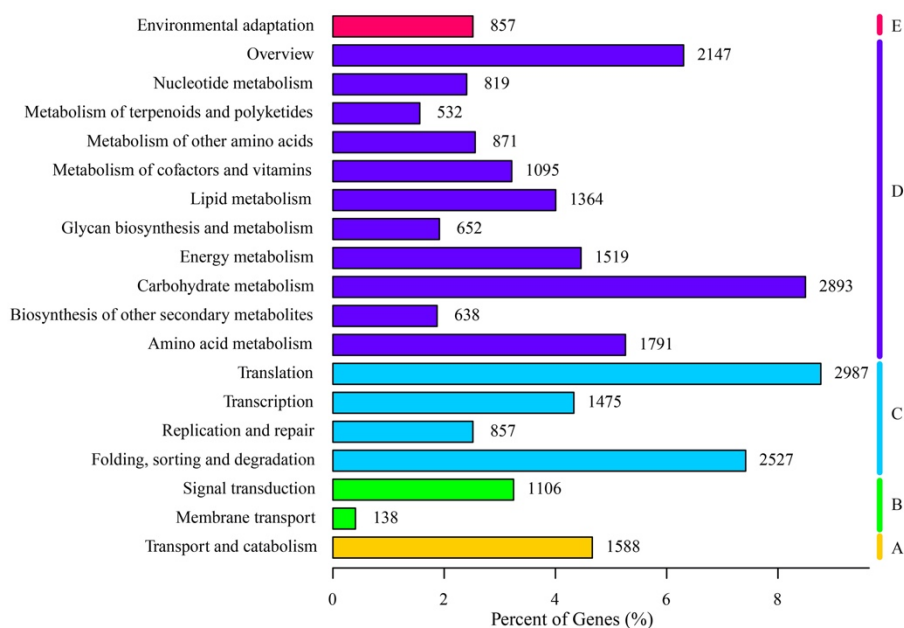


Figure 4 The KEGG classification of our transcriptome data

Note: A total of five categories were included: A, Cellular processes; B, Environmental information processing; C, Genetic information processing; D, Metabolism; E, Organismal systems

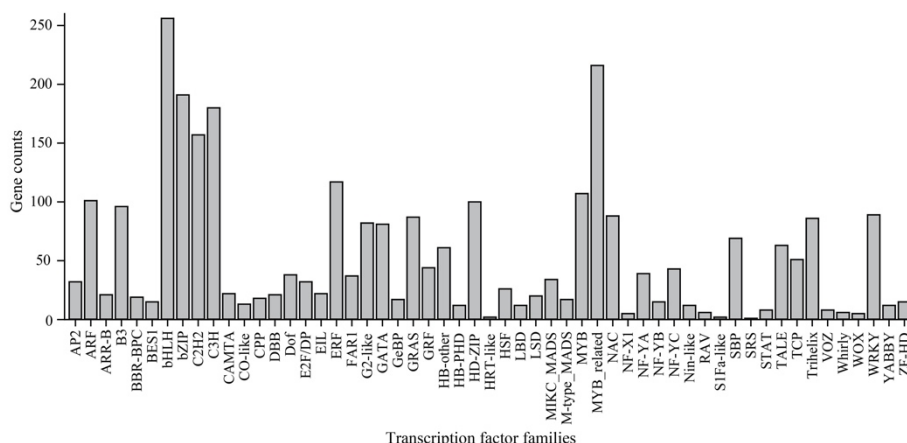


Figure 5 Numbers of transcription factors from different families

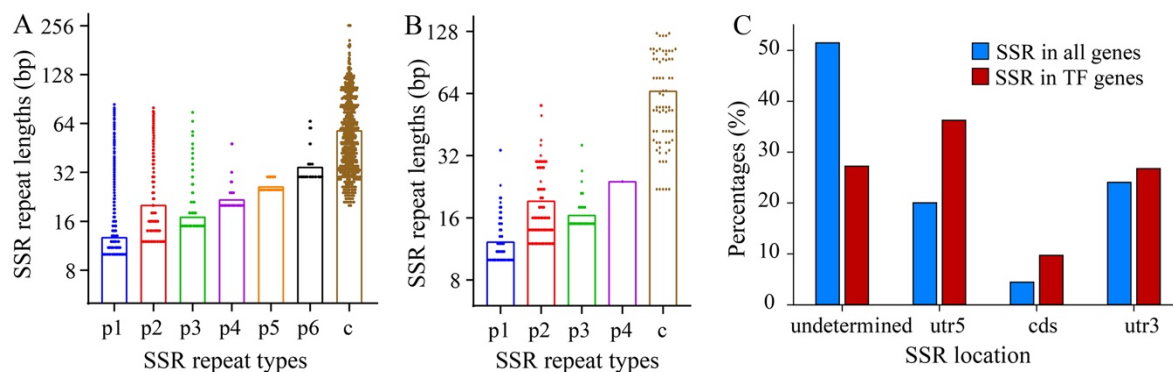


Figure 6 Statistical on simple sequence repeats (SSR)

Note: A, Sequence lengths distribution of SSRs amongst the whole transcriptome data; B, Sequence lengths distribution of SSRs from transcription factor coding genes; C, Location of SSRs on varied districts of unigenes. The SSR units included mononucleotides to hexanucleotides (p1 to p6), and complex units (c). these SSRs may locate on the 5' untranslated regions (utr5), 3' untranslated regions (utr3), coding sequences (cds), or currently unknown positions (undetermined)

1.3 Development of molecular markers

The sequence obtained from transcriptome sequencing was analyzed by simple repeat sequence (SSR), a total of 25,468 SSRs were found in 21,329 gene sequences (20.91% of all gene sequences). These SSR sequences mainly included repetitions of varying degrees from single base to six bases, and complex repetitive sequences. As shown in Figure 6A, except for two-base repeats, the average total length of SSR repeats increased as the complexity of the repeat unit increased, and the length of the repeats of the complex repeat unit was the longest. Among the 2,929 genes encoding transcription factors, 833 sequences (28.44%) owned different SSRs, the total length of these SSR sequences also had the aforementioned similar rule (Figure 6B). Finally, the analysis showed that SSR might be located in different positions of the gene, indicating that for all genes, more than half (51.46%) of the repetitive sequences might span two adjacent gene structures (5' untranslated region: utr5; coding region: cds; 3'untranslated region: utr3), and the proportion of SSR located in the coding region was the least (Figure 6C). For genes encoding transcription factors, the SSR span the two gene structures was significantly reduced (27.25%), and the proportion of repetitive sequences located in the coding region was still the least (Figure 6C). Subsequent detection and research on specific genes or transcription factors could be conducted more specifically through developing specific primers for these SSRs.

2 Discussion

Lisianthus is an important ornamental plant, however, the molecular biology research foundation for this species is relatively lacking. So far, only one subtractive library for salt stress was constructed in the early stage (Wang et al., 2008), genes possibly differentially expressed were identified. However, the throughput of subtracted libraries is generally low, and compared with current high-throughput sequencing, it is far from meeting the needs of research. In terms of molecular resource development, there was only transcriptome sequencing for flowering (Kawabata et al., 2012) in the early stage, in this transcriptome sequencing, only 65% of the 63,401 obtained-contigs were compared in the NCBI database, which was less than the 76.52% in this study (Table 1), it shows that with the development of sequencing technology and the maturity of splicing methods, the sequence annotation situation of this sequencing result has been improved. However, the species with the highest proportion of the transcriptome annotation results comparison (Figure 2D) was *Coffea canephora* (Rubiaceae), which may be due to the weak molecular research foundation of close-source species with *Lisianthus*. Rubiaceae and Gentianaceae belong to the order of Gentianaceae, so this *Coffea canephora* is probably the species closest to *Lisianthus* and has certain molecular research resources. In addition, the alignment results of the *Lisianthus* transcriptome sequence were scattered in other species (Figure 2D).

Studies had been conducted to identify *Lisianthus* MADS family genes (Ishimori and Kawabata, 2014) and study their functions (Li et al., 2015). However, without the support of high-throughput sequencing results such as transcriptomes or genomes, it is often more difficult to conduct gene function studies or gene family identification (Nakano et al., 2011). In this study, transcriptome sequencing was performed on *Lisianthus* under drought stress, and sequencing data was initially analyzed, the relevant research results can provide corresponding data support for the later study of the molecular mechanism of *Lisianthus* in response to drought stress. At the same time, some researchers had completed the sequencing of *Lisianthus* plastids (Yan et al., 2019), which relevant molecular resources provided the basis for the regulation of agronomic traits such as flower morphology through plastid genetic information (Jin and Daniell, 2015). The sequence comparison of this plastids sequencing results (Yan et al., 2019) with the corresponding sequencing results of other species in the Gentianaceae showed that *Lisianthus* was more distantly related to these species. This result confirmed from the side that in this transcriptome sequencing annotation result, the most compared species was *Coffea canephora* (39.7%) in the Rubiaceae of the order Gentiana, while most of the remaining sequence information was scattered to other species (Figure 2D). Finally, the results of this transcriptome sequencing will provide molecular data for subsequent studies such as identifying and mining drought-response-related genes (Wang et al., 2018) or drought stress-related transcription factors (Xu et al., 2015).

3 Materials and Methods

3.1 Plant material

Lisianthus (*Eustoma grandiflorum*) variety "Xuelai" was purchased in the market. The plant was subjected to drought stress when grows to about 8cm. After 36 hours of treatment, liquid nitrogen was used to take whole plant samples for total RNA extraction.

3.2 Total RNA extraction and library construction

The samples had been stored in an ultra-low temperature refrigerator at -80°C before being used for RNA extraction. The sample was taken out and fully ground into powder in a liquid nitrogen environment. Tiangen's RNA extraction kit was used to complete total RNA extraction and construct a transcriptome library.

3.3 Sequencing data processing and transcript assembly

The direct data obtained by sequencing samples on the computer were raw reads that need to be quality controlled, removed uncertain base sequencing results and poor quality sequencing information due to the inclusion of more than 10% adapters, the remaining readings were high-quality sequencing information. The processing methods for high-quality sequencing information were mainly sequence splicing (Grabherr et al., 2011) and hierarchical clustering (Davidson and Oshlack, 2014), and the resulting gene information was non-redundant genes.

3.4 Gene function annotation and classification

The non-redundant genes obtained through the above steps were compared with four sequence databases for annotation.

3.5 Transcription factor prediction

While annotating non-redundant gene sequences, online tools(<http://plantfdb.gao-lab.org/prediction.php>) were used to predict the possible encoding products of these genetic information, and if they encode transcription factors, they can be classified, and it would be classified if it encodes a transcription factor.

3.6 SSR analysis

There were some sequence features with clear rules at the base sequence level of genes, these sequences were often repeated multiple times based on simple sequence units (single to multiple bases, or even complex base units), which were called simple sequence repetitions (SSR). Online tools (<http://pgrc.ipk-gatersleben.de/misa/misa.html>) were generally used to predict these SSRs. SSRs that were repeated ten times or more with a single base as a unit, six or more times with a two-base as a unit, and five or more times with a three-base to six-base as the repeat unit were all included in the statistical scope, if the complex repeating sequence contained different repeating units listed above, each repeating unit met the above requirements.

Authors' contributions

An Xia is the executor of the experimental design and experimental research of this study, completing data analysis and writing the first draft of the paper; Zhu Qiang, Lou Xuping, Li Lufeng, Chen Jie, Liu Tingting, Li Wenlue, Luo Xiahong, Zhu Guanlin, Yu Lijun are the experimental design participants; An Xia is the creator and person in charge of the project, directing experimental design, data analysis, essay writing and revision. All authors read and approved the final manuscript.

Acknowledgement

This research was funded by the Provincial Science and Technology Commissioner Project (Terrace Scenic Area Farmhouse Landscape Improvement Demonstration and Leisure Product Innovation).

References

Davies K.M., Marie Bradley J., Schwinn K.E., Markham K.R., and Podivinsky E., 1993, Flavonoid biosynthesis in flower petals of five lines of lisianthus (*Eustoma grandiflorum* Grise.), *Plant Science*, 95(1): 67-77
[https://doi.org/10.1016/0168-9452\(93\)90080-J](https://doi.org/10.1016/0168-9452(93)90080-J)

- Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen Nir., Gnirke A., Rhind N., Palma F.D., Birren B.W., Chad Nusbaum., Lindblad-Toh K., Friedman N., and Regev A., 2013, Trinity: reconstructing a full-length transcriptome without a genome from rna-seq data, *Nature Biotechnology*, 29, 644
<https://doi.org/10.1038/nbt.1883>
PMid:21572440 PMCID:PMC3571712
- Götz S., García-Gómez J.M., Terol J., Williams T.D., Nagaraj S.H., Nueda M.J., Robles M., Talo' N.M., Dopazo J., and Conesa A., 2008, High-throughput functional annotation and data mining with the blast2go suite, *Nucl Acids Research*, 36(10): 3420-3435
<https://doi.org/10.1093/nar/gkn176>
PMid:18445632 PMCID:PMC2425479
- Ishimori M., Kawabata S., 2014, Conservation and Diversification of Floral Homeotic MADS-box Genes in *Eustoma grandiflorum*, *J. Japan, Soc. Hort. Sci.* 83(2): 172-180
<https://doi.org/10.2503/jjshs1.CH-098>
- Jin G.R., Chen J., and An X., 2019, Transcriptome profiling of jute under drought stress, *Fenzi Zhiwu Yuzhong (Molecular Plant Breeding)*
- Jin S., and Daniell H., 2015, The Engineered Chloroplast Genome Just Got Smarter, *Trends in plant science*, 20(10): 622-640
<https://doi.org/10.1016/j.tplants.2015.07.004>
PMid:26440432 PMCID:PMC4606472
- Kawabata S., Li Y., and Miyamoto K., 2012, EST sequencing and microarray analysis of the floral transcriptome of *Eustoma grandiflorum*, *Scientia Horticulturae*, 144: 230-235
<https://doi.org/10.1016/j.scienta.2011.12.024>
- Li K.H., Chuang T.H., Hou C.J., and Yang C.H., 2015, Functional Analysis of the FT Homolog from *Eustoma grandiflorum* Reveals Its Role in Regulating A and C Functional MADS Box Genes to Control Floral Transition and Flower Formation, *Plant Mol Biol Rep*, 33(4): 770-782
<https://doi.org/10.1007/s11105-014-0789-y>
- Nakano Y., Kawashima H., Kinoshit T., Yoshikawa H., and Hisamatsu T., 2011, Characterization of FLC, SOC1 and FT homologs in *Eustoma grandiflorum*: effects of vernalization and post-vernalization conditions on flowering and gene expression, *Physiologia plantarum*, 141(4): 383-393
<https://doi.org/10.1111/j.1399-3054.2011.01447.x>
PMid:21241311
- Wang B., Wang Q.F., Tang T.X., Tang W.J., Li L.P., Sui L.B., Sun C., Zhang H., Xia Z.T., and Lin L.B., 2018, Discovery of drought-tolerant genes in *Brassica napus* by transcriptome analysis, *Jiyinzuxue Yu Yingyong Shengwuxue (Genomics and Applied Biology)*, 37(11): 4775-4786
- Wang J.G., Zhang K., Xu Q.J., and Li Y.H., 2008, Construction and Analysis of *Eustoma grandiflorum* Subtracted cDNA Library, *Acta Horticulturae Sinica*, 35(7): 1075-1080
- Xu J., Tang Q., Zhu H., Yan X., Tang L., and Meng W., 2015, Research Advance of the Transcription Factors Related to Rice (*Oryza sativa* L.) Drought Stress Responses, *Genomics and Applied Biology*, 34(11): 2525-2531
- Yan J., Cao Q., Wu Z., Chen S., Wang J., Zhou D., and Xie J., 2019, Complete plastome sequence of *Eustoma grandiflorum* (Gentianaceae), a popular cut flower, *Mitochondrial DNA Part B*, 4(2): 3163-3164
<https://doi.org/10.1080/23802359.2019.1667893>
PMid:33365900 PMCID:PMC7706805