Gene 660 (2018) 18–27

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

The transcriptome analyses of *Tagetes erecta* provides novel insights into secondary metabolite biosynthesis during flower development

Guodong Feng[a,1], Shengxiong Huang[a,*,1], Ying Liu[a,1], Fangming Xiao[b], Jian Liu[a], Zheng Zhang[a], Qiang Chen[a], Yun Mao[a], Xulv Cao[a], Yang Wang[a], Danyang Chen[a], Yu Zhou[a], Fengyuan Yu[a], Guoqing Liu[a], Yongsheng Liu[a], Xiangli Niu[a,*]

[a] School of Food Science and Engineering, Hefei University of Technology, Hefei 230009, China
[b] Department of Plant Sciences, University of Idaho, Moscow, ID 83844-2339, USA

ARTICLE INFO

Keywords:
Tagetes erecta
RNA-seq
Carotenoids
Reference gene
Expression profile

ABSTRACT

Genomic and transcriptomic sequences of *Tagetes erecta* are very limited in public databases, despite its nutritional and economical value. In this study, nine cDNA libraries were constructed from leaves, immature and mature flowers and sequenced using Illumina Hiseq 2000. Over 45.66 Gb nucleotides were generated and a total of 72,499 unigenes were assembled, 34,892 (48.13%) of which were annotated in NR, Swiss-prot, COG, GO, KOG, KEGG and Pfam. 11,721 differentially expressed genes were identified in leaves, immature flowers and mature flowers. The differentially expressed genes between immature flowers and mature flowers were mainly involved in photosynthesis and secondary metabolite biosynthesis. Additionally, the catalysis-related unigenes and their expression involved in carotenoids biosynthesis were determined. Using our newly identified reference genes as internal control, the expression profiles of carotenoids biosynthetic genes were verified by real-time qRT-PCR, and four of the unigenes were cloned for full length sequences confirmation. Furthermore, gene expansions occurred among the catalytic gene families in carotenoids biosynthesis pathway, which might explain the high pigment content in *T. erecta*.

1. Introduction

Tagetes erecta, commonly called marigold, is an economically important plant because of its remarkably high pigment content and other agriculture/medicine-beneficial secondary metabolites. Due to its versatile roles, a wide range of researches have been conducted in *T. erecta*, such as secondary metabolites extraction and characterizations, evaluation of its phytoremediation and stress tolerance (Hadden et al., 1999; Wei et al., 2015; Goswami and Das, 2017).

The flowers of *T. erecta* are rich in carotenoids, flavonoids and other phenolic compounds, and thus be considered as major natural sources of xanthophylls, including zeaxanthin, lutein and lutein esters (Hadden et al., 1999; Liu et al., 2011). *T. erecta* plants have been cultivated in many countries for xanthophylls extraction and application in drugs for eye-related diseases and edible alternative in food. Moreover, the methanol extracts of *T. erecta* flowers exhibit high antibacterial activity against both Gram-negative (*Escherichia coli*) and Gram-positive (*Bacillus subtilis*) bacterial strains, and fungi (*Alternaria alternate*) (Ayub et al., 2017; Saani et al., 2017). Its alcoholic extracts also show antioxidant capacity, with quercetagetin identified as the strongest antioxidant compound (Gong et al., 2012; Ayub et al., 2017).

Additionally, *T. erecta* and other *Tagetes* species can produce thiophenes, which exhibit strong biocidal activity and are natural pesticides (Jacobs et al., 1995; Marotti et al., 2010; Gupta et al., 2012). Evidence demonstrated that their thienyl compounds indicated profound antioxidant activity and cytotoxicity at relatively lower concentrations (Gupta et al., 2012). Furthermore, *T. erecta* plants can accumulate metals in above-ground tissues, suggesting their potential for cadmium, copper, plumbum and zinc phytoremediation (Sinhal et al., 2010; Goswami and Das, 2017).

Several studies on functional characterization of *T. erecta* genes have been performed. Five MADS-Box genes, *TePI*, *TeAP3-1*, *TeAP3-2*, *TeTM6-1* and *TeTM6-2* have been cloned and their putative functions were investigated. Yeast hybrid showed TePI can form homodimer and

Abbreviations: Cq, Quantification Cycle; DEG, Differentially Expressed Gene; FDR, False Discovery Rate; FPKM, Fragments Per Kilobase of exon model per Million mapped reads; GO, Gene Ontology; NCBI, National Center for Biotechnology Information; qRT-PCR, Quantitative Reverse Transcription Polymerase Chain Reaction

* Corresponding authors.
E-mail addresses: huangshengxiong@163.com (S. Huang), niu_xiangli@163.com (X. Niu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.gene.2018.03.051
Received 24 August 2017; Received in revised form 13 March 2018; Accepted 16 March 2018
Available online 21 March 2018
0378-1119/

heterodimers with *TeAP3-1*, *TeAP3-2*, *TeTM6-1* and *TeTM6-2*, respectively. Over-expression of *TePI* in tobacco showed altered floral morphology (Ai et al., 2017). However, little is known for the genomic and transcriptomic sequences of *T. erecta*. With the advancement of sequencing technology and decrease of sequencing cost, there is a good chance to reveal its transcriptomic sequences and gene profiling during different development stages. Recently, a transcriptomic study was reported to determine the differential gene expression profiling in male sterile and male fertile flower buds, with interesting findings that twelve MADS-box genes are involved in male sterility (Ai et al., 2016).

In this study, we carried out the RNA-seq-based comprehensive transcriptional analyses on leaves, immature flowers and mature flowers of *T. erecta*, seeking to reveal the molecular mechanisms underlying the secondary metabolite biosynthesis and flower development. We performed profound transcriptome investigation, including transcriptome sequencing, sequence assembly and annotation, analyses of differentially expressed genes, validation of new reference genes, and expression profiling of genes involved in carotenoids biosynthesis pathway. Our findings of expanded genes and related expression profiles shed light on the molecular basis of pigment biosynthesis in *T. erecta* flowers, potentially providing gene candidates for further genetic manipulation for improvement of agronomic traits in *T. erecta* and related *Tagetes* species.

2. Plant materials and methods

2.1. Plant materials

Leaves, immature and mature flowers were collected from 70-day-old plants of *T. erecta* 'Juwang', growing in the farm of Hefei University of Technology, Anhui Province, China. To reflect the developmental stage, the initial flowers were defined as DP0 (Days Post 0.5 cm) representing the size of 0.5 cm in diameter. And flowers of DP6 and DP21 were defined as immature and mature flowers for RNA-seq, respectively. All the tissue samples were immediately frozen in liquid nitrogen and stored at −80 °C for further use.

2.2. RNA extraction, library construction and transcriptome sequencing

Leaf or flower samples, collected as previously described, were grounded to fine powder with liquid nitrogen. According to the manufacturers' protocols, total RNAs were isolated using the Trizol reagent (Invitrogen, USA), followed by treatment with RNase-free DNase I (Promega, USA). The quality of RNAs was checked using Agilent 2100 Bioanalyzer.

Nine Illumina RNA-Seq libraries were generated and sequenced by the HiSeq 2000 system, according to the manufacturer's instructions (Illumina, USA). Three biological replicates were included for each sequencing tissue. The raw sequence data were deposited in BioProject database of NCBI (http://www.ncbi.nlm.nih.gov) with accession PRJNA431782.

2.3. De novo sequence assembly and annotation

The raw reads were cleaned by removing adapter sequences, low quality sequences (reads containing > 50% bases with Q value ≤10), and reads with unknown nucleotides 'N' larger than 5%. The cleaned reads were assembled to unigenes using Trinity. The unigenes sequences were respectively annotated in public databases, NR, Swiss-Prot, COG, GO and KEGG. The default cutoff e-value was < 1–5E. Gene Ontology (GO) annotations of each gene were performed by Blast2GO software (version 4.1.5), using the default parameters (Conesa et al., 2005).

2.4. Quantification of gene expression profiles

The clean paired-end RNA-seq reads were first aligned to ribosomal RNA and tRNA sequences to remove possible contaminations of these sequences. For each sample, the reads were subsequently aligned to the assembled unigene sequences using Bowtie (Langmead et al., 2009). Following the alignment, raw counts for each unigene were normalized to the FPKM (fragments per kilobase of exon model per million mapped reads) value (Trapnell et al., 2010).

2.5. Analyses of differentially expressed gene

Differential expression analyses between two different types of tissues were performed, using the R package DEseq (Anders and Huber, 2010). The *P* value was adjusted using the FDR (False Discovery Rate) value. The absolute value of log2 (fold change) ≥1 and FDR value < 0.01 were set as the threshold to evaluate the significant differential expression.

2.6. Identification of stably expressed genes from RNA-seq data

Based on the transcriptome expression profiles, the value of mean, standard deviation and coefficient of variation of each gene across three different types of samples were determined by using Excel 2010 (Microsoft, USA). Stably expressed genes were identified according to the optimized parameters as follows: mean value, between 15 and 110; coefficient of variation < 0.1.

2.7. Validation of new reference genes

Seven genes, stably expressed across leave and flower samples in RNA-seq data, were selected as candidate reference genes for the real-time qRT-PCR. Gene-specific primers were designed using software Primer Premier 5.0 and the quantification cycle (Cq) values of each gene were exported from StepOne Plus software V2.3 (Applied Biosystems, USA). Using Cq data as input, the expression stability was analyzed and ranked by two independent algorithms, geNorm (Vandesompele et al., 2002) and NormFinder (Andersen et al., 2004), respectively.

2.8. cDNA synthesis and real-time qRT-PCR

cDNA synthesis was carried out in a total volume of 20 μl containing 2 μg RNA by using the TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix (Transgen Biotech, China). The real time qRT-PCR was performed in StepOne Plus Real-Time PCR System (Applied Biosystems, USA) using the TransStart Top Green qPCR SuperMix (Transgen Biotech, China). The 20 μl reaction mixture contained 8.6 μl of diluted template (including 0.5 μl of the generated first-strand cDNA), 10 μl of Supermix, 0.4 μl of passive reference dye, and 0.5 μl of 10 μM gene-specific primers. The reactions were performed with the following cycling program: 95 °C for 1 min, 40 cycles at 95 °C for 10 s, and 65 °C for 15 s. Melting curve analysis was performed ranging from 60 to 95 °C, and PCR products were separated on agarose gel and sequenced to confirm the specificity of the PCR reaction and verify the accuracy of the amplification. Each sample was quantified in triplicates.

2.9. Gene cloning and sequencing

The full length of cDNA of four carotenoids biosynthesis genes, *TePDS1* (*T. erecta phytoene desaturase 1*), *TeLYCB* (*T. erecta 1ycopene β-cyclase*), *TeVDE1* (*T. erecta violaxanthin deepoxidase 1*) and *TeVDE2* (*T. erecta violaxanthin deepoxidase 2*), were cloned by nested PCR using primers designed from the assembled unigenes. The PCR products were ligated to pBlunt vector (Transgen Biotech, China) and verified by DNA sequencing.