High-throughput sequencing analysis of *Euphorbia fischeriana* Steud provides insights into the molecular mechanism of pharmaceutical ingredient biosynthesis

Ming Jiang¹ · Hui Li¹

Received: 12 June 2018 / Accepted: 9 October 2018 / Published online: 12 October 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

High-throughput sequencing is an effective approach to analyse the bioinformation on the molecular biological and whole genome levels, especially in non-model plants for which reference genome sequences are unavailable. In this study, high-throughput sequencing analysis of *Euphorbia fischeriana* Steud was conducted on the Illumina HiSeq 2000 platform. A total of 9,6481,893 raw reads were generated and assembled into 304,217 transcripts and 186,384 unigenes. Of the 186,384 unigenes, 77.45% were annotated in at least one database, and some pathways involved in the biosynthesis of the terpenoid backbone were closely linked to the main anticancer components. In addition, 7452 transcription factors and 76,193 SSRs were detected. This study may provide a candidate pathway for terpenoid backbone biosynthesis in this medicinal plant.

Keywords *Euphorbia fischeriana* Steud · High-throughput sequencing analysis · Diterpenoid biosynthesis

Introduction

Euphorbia fischeriana Steud is a traditional perennial herbaceous plant with milky juice, distributed mainly on hilly slopes, sandy grasslands, sparse pine forests in northeastern China and belongs to the large family of Euphorbiaceae. The first book to describe *Euphorbia fischeriana* Steud was the Chinese *Shennong Bencao Jing* written about 1000 years ago. The dried root of the plant is commonly known as “Lang-Du-Da-Ji”, which has been used for treatment of edema, ascites, tuberculosis and cancer for a long time (Jian et al. 2018). The root contains various but overlapping pharmaceutical ingredients with medicinal effects. Many secondary metabolites with therapeutic efficacy, such as diterpenoids, triterpenoids, steroids, phenolic acids, and anthraquinones; volatile oils; and other key enzymes in metabolic processes have been identified in *Euphorbia fischeriana* Steud (Sun and Liu 2011). For example, jolkinolides A and B showed cytotoxic activities towards HeLa cells and Ehrlich ascites, and 17-hydroxyjolkinolide

B is a potent STAT3 signaling inhibitor and was suggested to be a promising anticancer drug candidate (Uemura et al. 1997; Wang et al. 2009). Of these, diterpenoids are the major active constituents of this medicinal herb which have notable biological effects against tumor cells, and their anticancer activity has attracted more attention in recent years (Shen et al. 2017).

Diterpenoids are a class of terpenes whose structure is based on the diterpene (C₂₀) backbone (Kirby et al. 2010). There are two biosynthetic pathways, the MVA (mevalonate) pathway and the DOXP/MEP (1-deoxy-D-xylulose 5-phosphate/2-C-methyl-D-erythritol 4-phosphate) pathway, for isopentenyl-PP in terpenoid backbone biosynthesis (Ha et al. 2017). A number of metabolite profiling studies have been conducted to identify diverse diterpenoid lactone compounds in the roots of *Euphorbia fischeriana* Steud and other *Euphorbia* species (Wang et al. 2013, 2017; Kuang et al. 2016). However, very little is known about the biosynthetic pathway of diterpenoids, although much is known about its biochemical structure. To date, transcriptome analysis has only been conducted on prostratin from *Euphorbia fischeriana* Steud, a protein kinase C activator and effective in the treatment of HIV-infected patients, providing insights into the diterpenoid pathway (Barrero et al. 2011).

In recent years, several other *Euphorbia* species have been used as medicinal plants to treat cancer and tuberculosis.

✉ Hui Li
lihui314220@126.com

¹ Department of Science and Technology, Qiqihar Medical University, Qiqihar 161006, Heilongjiang, China

However, the literature shows that the pharmacodynamic effect and chemical components differ on the basis of the cultivars and environmental conditions. Thus, this is an opportunity to research the genetic background of *Euphorbia* species using molecular markers. The genetic diversity of germplasm resources of *Euphorbia fischeriana* Steud has typically been analysed with ISSR markers (Li et al. 2014). However, ISSR markers suffer from poor codominance. Compared to ISSR markers, simple sequence repeat (SSR) markers show several advantages, i.e., reproducibility, multi-allelic loci, codominance and analytically simple (He et al. 2013; Liu et al. 2015).

Genomic approaches have been used to discover the important genes and key enzymes involved in medicinal plant secondary metabolism pathways. However, the genome of *Euphorbia fischeriana* Steud is still unavailable. Transcriptome analysis is an efficient way to discover and characterize novel enzymes and transcription factors from *Euphorbia fischeriana* Steud and can be used to identify SSR loci from ESTs rapidly with high throughput. Here, we performed de novo transcriptome sequencing of *Euphorbia fischeriana* Steud root using Illumina paired-end RNA sequencing technology. Based on the sequencing results, we identified candidate genes involved in the terpenoid backbone pathway and SSR loci information, providing insights into this important medicinal plant.

Materials and methods

Plant materials

A *Euphorbia fischeriana* Steud cultivar with high levels of pharmaceutical ingredients was used in this study. The pharmaceutical ingredients consist of diterpenoids, triterpenoids, tannins, and anthraquinone steroids, among other compounds. Live plants of *Euphorbia fischeriana* Steud were collected in July 2015 from Jiagedaqi of Hei longjiang province of China. The plants were then grown under natural light at 25 °C in a greenhouse of Qiqihar Medical University, Qiqihar. After 150 days of growth after planting, three fresh roots of *Euphorbia fischeriana* Steud were harvested, immediately frozen in liquid nitrogen and stored at −80 °C for subsequent study.

RNA isolation and quality assessment

For transcriptome analysis, total RNA was extracted from the fresh roots of *Euphorbia fischeriana* Steud using Trizol reagent (TransGen Biotech, Beijing, China) according to the manufacturer's protocol. RNA samples were then treated with RNase-free DNase I (Takara, Dalian, China) for 4 h at 37 °C to prevent DNA contamination. The purified RNA

concentrations were quantified using a Nanodrop spectrophotometer (Thermo Nanodrop Technologies, Wilmington, DE, USA), and the quality was initially checked on a 1% denaturing gel with Tris-(hydroxymethyl)-aminomethane boric acid (TBE) buffer before proceeding. Then, the quality of the total RNA was further examined by an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) with a lab chip kit (Zhang et al. 2016). Finally, 10 µg of the RNA that passed the quality control checks for each sample was prepared for construction of a cDNA library and quantitative real-time PCR (qRT-PCR) validation.

cDNA library construction, quality control, and Illumina sequencing

The transcriptome library for sequencing was prepared using a NEXTflex™ Rapid Directional RNA-Seq library prep kit according to the manufacturer's protocol. From the total RNA, poly(A) mRNA was isolated and enriched using oligo(dT)-attached magnetic beads and then chemically broken into short fragments in fragmentation buffer by an RNA fragmentation kit (Ambion, Austin, TX, USA). Next, the cDNA was synthesized using the mRNA fragments as templates. Short fragments were purified and dissolved EB buffer for end reparation and single nucleotide A (adenine) addition. The short fragments were connected with adapters, and first-strand cDNA was synthesized with reverse transcriptase and random hexamer primers. Then, second-strand cDNA was synthesized in a buffer containing DNA polymerase I and RNase H (Invitrogen, Carlsbad, CA, USA). dNTPs were added at this step to maintain directionality. After end repair and the addition of a poly(A) tail, the fragments were purified by agarose gel electrophoresis, and suitable fragments (approximately 180 bp) were isolated as a template for PCR amplification and connected to the sequencing adaptors. Then, the sequences were sequenced on the Illumina HiSeq™ 2000 platform (Novogene Bioinformatics Technology Company, Beijing, China). After sequencing, raw image data were transformed by base calling into sequence data, which were called raw data or raw reads and were stored in the fastq format.

Raw sequencing processing and de novo assembly

To acquire high-quality reads, the raw sequences were generated by removing the adaptor sequences, empty reads and low-quality reads containing N. The clean reads were de novo assembled using the Trinity method (Grabherr et al. 2011). The clean reads produced from a library were assembled into contigs with the Inchworm program. The contigs were linked to transcripts based on the paired-end sequence information. The transcripts were subsequently clustered according to their nucleotide identity. We took the longest transcript in a cluster

unit as the unigene of each cluster. The unigenes were input into the non-redundant UniGene database for annotation.

Functional annotation and classification of non-redundant unigenes

To further annotate the unigenes, all assembled unigenes were first searched against the NCBI non-redundant protein database (Nr) and Swiss-Prot protein database using BLASTX alignment (E value $< 1e^{-5}$) and the NCBI non-redundant nucleotide sequence (Nt) database using BLASTn (E value $< 1e^{-5}$). Gene names were assigned to each unigene based on the BLAST hit (highest score), and the best alignment results were used to determine the sequence direction. The Blast2GO program was used to obtain the GO annotation of unigenes based on the Nr annotation, and GO functional classification was performed using the WEGO software (Conesa et al. 2005; Ye et al. 2006). GO terms were grouped into three components: biological processes (BPs), cellular components (CCs), and molecular functions (MFs). Further functional analysis of the transcriptome was performed with the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (Kanehisa et al. 2012), Pfam database, and Cluster of Orthologous Groups (COG) database. These enhance our understanding of gene classification in a species-independent manner and provides an overview of gene functions of species.

Identification of transcription factors

The protein sequences of all plant transcription factors and the assignment of these sequences to different transcription factor families were downloaded from the Plant Transcription Factor Database (PlnTFDB) (<http://plntfdb.bio.uni-potsdam.de/v3.0/downloads.php>). Red transcripts were subjected to BLASTx analysis against the PlnTFDB peptide sequences with an E value cutoff of $< 10^{-5}$.

SSR detection

Microsatellite Identification Tool (MISA) software (<http://pgrc.ipk-gatersleben.de/misa/>) was used to identify the SSRs in the transcriptome. Six types of SSRs, including mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide repeats, were detected among the unigenes with lengths of > 1000 bp.

Results

Transcriptome sequencing and de novo assembly

In this study, the transcriptome profiles of *Euphorbia fischeriana* Steud were determined by Illumina HiSeq 2000

sequencing, and 9,6481,893 raw reads with an average GC content of 44.32% were obtained from fresh roots. After removing adaptor sequences containing N reads or low-quality reads, 91,413,326 (94.75%) high-quality reads were obtained. Assembly of the high-quality reads yielded 304,217 transcripts with a mean length of 1007 bp and an N50 length of 1849 bp, and these transcripts were further clustered into 186,384 unigenes with an average length of 1452 bp and an N50 length of 2048 bp (Table 1). The length distributions of the transcripts and unigenes are shown in Fig. 1.

Functional annotation and classification of non-redundant unigenes

For functional annotation, the sequences of 186,384 assembled unigenes were searched against the Nr, Nt, Swiss-Prot, Pfam, KOG, GO, and KEGG databases. According to the BLASTX results, a total of 144,356 (77.45%) unigenes were annotated with putative functions based on hits to at least one database, and 25,473 (13.66%) unigenes were annotated in all databases. Among them, 128,280 (68.82%) of the putative proteins showed similarity to sequences in the Nr database. A total of 106,939 (57.37%), 54,459 (29.21%), 101,182 (54.28%), 97,722 (52.43%), 101,997 (54.72%), and 43,602 (23.29%) of the unigenes had functional annotations in the KO, Swiss-Prot, Pfam, GO, and KOG databases (Table 2). Particularly, the predicted results from the Nr database revealed that 73,435 (39.4%) and 49,578 (26.6%) unigenes showed significant homology with sequences from *Jatropha curcas* and *Populus euphratica*, respectively (Fig. 2).

Analysis and functional classification of unigenes

Enrichment analysis was performed to determine the biological functions of potential unigenes using the WEGO software. According to the gene ontology (GO) function analysis (Fig. 3), a total of 101,997 unigenes generated biological

Table 1 Length distribution of assembled contigs and unigenes from *Euphorbia fischeriana* Steud

Nucleotide length (bp)	Transcripts	Unigenes
0–500	146,379 (48.12%)	34,926 (18.74%)
500–1000	57,779 (18.99%)	52,210 (28.01%)
1000–2000	55,189 (18.14%)	54,429 (29.20%)
> 2000	44,870 (14.75%)	44,819 (24.05%)
Total number	304,217	186,384
Total length (bp)	306,406,082	270,594,099
N50 length (bp)	1849	2048
N90 length (bp)	372	691
Mean length (bp)	1007	1452

Fig. 1 Overview of the *Euphorbia fischeriana* Steud transcriptome assembly. **a** Length distribution of transcripts. **b** Length distribution of unigenes

process terms (47.67%), cellular component terms (29.55%), and molecular function terms (22.78%), and these unigenes were classified into three categories, including 56 functional groups. In-depth analysis of the cellular component terms revealed that most genes belonged to binding (10.57%) (GO:0005488) and catalytic activity (8.87%) (GO:0003824). Cellular processes (10.63%) (GO:0009987), metabolic processes (10.06%) (GO:0008152), single-organism processes (8.05%) (GO:044699) and biological regulation (3.57%) (GO:0065007) dominated the molecular function terms. Cell

(5.77%) (GO:0005623), cell part (5.77%) (GO:0044464) and organelle (3.94%) (GO:0043226) terms were the most represented categories under the biological process terms. The GO analysis revealed that most of unigenes were involved in catalytic activity and metabolic processes, which is closed linked to metabolism of terpenoid substances. These annotations provide a valuable resource for understanding *Euphorbia fischeriana* Steud root. A total of 54,459 unigenes were annotated with the KOG database and subdivided into 26 KOG classification groups (Fig. 4). The largest group was

Table 2 Summary of the annotation of the unigenes in the transcriptome of *Euphorbia fischeriana* Steud

Public protein database	Number of annotated unigenes	Percentage of annotated unigenes (%)
NR annotation	128,280	68.82
NT annotation	106,939	57.37
KO annotation	54,459	29.21
Swiss-prot annotation	101,182	54.28
PFAM annotation	97,722	52.43
GO Annotation	101,997	54.72
KOG annotation	43,602	23.39
Annotation in all databases	25,473	13.66
Annotated in at least one database	144,356	77.45
Total	186,384	100

NR non-redundant protein database, NT non-redundant nucleotide sequence database, KO KEGG (Kyoto Encyclopedia of Genes and Genomes pathway) Orthology, Swiss-Prot Swiss-Prot protein database, PFAM The Pfam database is a large collection of protein families, GO gene ontology, KOG Eukaryotic orthologous groups

Fig. 2 Annotation and classification of the assembled unigenes in the non-redundant databases, and the species distribution of BLAST hits for each unigene in the Nr database

the cluster for posttranslational modification (5894), followed by the clusters for general function prediction only (5706), ribosomal structure and biogenesis (3947) and RNA processing and modification (3635).

To further understand the biological functions of unigenes, the KEGG pathway database was used, and according to the enrichment results, a total of 54,459 matched sequences were assigned to 133 relevant metabolic pathways, the majority of which were related to carbon metabolism (ko01200) and amino acid biosynthesis (ko01230) (Fig. 5). Among these two groups, diterpenoid biosynthesis (ko00904), sesquiterpenoid and triterpenoid biosynthesis (ko00909) and terpenoid backbone biosynthesis (ko00900)

were the closest connections to the biosynthesis of *Euphorbia fischeriana* Steud pharmaceutical ingredients.

Terpenoid biosynthesis which belongs to the secondary metabolisms is a dynamic and complex process catalysed by a series of enzymes, the GO/KOG/KEGG analysis revealed that these unigenes were involved in a number of important processes and pathways, which may be contributed to the pharmaceutical ingredient of *Euphorbia fischeriana* Steud.

Detection of genes related to terpenoid backbone biosynthesis

Unigenes assigned to KEGG IDs were mapped to KEGG pathways. Several key enzymes in the mevalonate and non-mevalonate pathways of terpenoid backbone biosynthesis were identified in the *Euphorbia fischeriana* Steud unigene sequences. We found 33 transcripts encoding candidate genes in the two pathways and related them to 19 *HMGR* (3-hydroxy-3-methylglutaryl-CoA reductase) genes, 6 *CMK* (4-diphosphocytidyl-2-C-methyl-D-erythritol kinase) genes, 1 *PMK* (phosphomevalonate kinase) gene, 1 *DXR* (1-deoxy-D-xylulose 5-phosphate reductoisomerase) gene and 1 *DXS* (1-deoxy-D-xylulose 5-phosphate synthase) gene.

Identification of transcription factors

Transcription factors (TFs) play important roles in plant development and response to the environment (Zhang 2003). Some important TFs were identified in *Euphorbia fischeriana* Steud. A total of 7452 TFs were found in the roots of *Euphorbia fischeriana* Steud (Table 3): the most notable groups were MYB, bZIP, bHLH, NAC, WRKY and AP2, and other plant common TFs, such as ARF, MADS, Trihelix, HSF, TCP and PLATZ, were also identified in *Euphorbia fischeriana* Steud (Fig. 6).

SSR markers

Microsatellites are short tandem repeats used as markers, because they are highly polymorphic and present in diverse groups of lower eukaryotes (Bretagne et al. 1997; Field and Wills 1996). Using the MISA software to search for potential SSRs from 186,384 unigenes, 76,193 SSRs were identified in *Euphorbia fischeriana* Steud, 53,146 of which contained SSR markers and 16,101 of which contained more than one SSR. The total number of SSRs in *Euphorbia fischeriana* Steud was 76,193. Monomer SSRs were dominant at 58.81%. Dimer and trimer SSRs were found to make up 18.70% and 18.67%, respectively. In addition, tetramer, pentamer and hexamer SSRs composed 2.17%, 0.91% and 1.74% of the total abundance, respectively. The SSR characteristics are presented in Fig. 7.

Fig. 3 Gene ontology categories of the assembled unigenes from the root of *Euphorbia fischeriana* Steud. Unigenes were assigned to three categories: biological processes, cellular components, or molecular functions

Fig. 4 Distribution of Kyoto Encyclopedia of Gene and Genomes (KEGG) pathways in *Euphorbia fischeriana* Steud

Discussion

Euphorbia fischeriana Steud is a perennial herbaceous and very important in our life. The roots of *Euphorbia fischeriana* Steud had long been not only used as a traditional

Chinese medicine, but also as a usual pesticide. Many chemical investigations of this plant suggested that it contains diterpenoids, triterpenoids, steroids, phenolic acids, anthraquinones, volatile oils and so on. Among these pharmaceutical ingredient the diterpenoids are the most

Fig. 5 KEGG-based functional classification of *Euphorbia fischeriana* Steud sequences. The numbers beside each bar indicate the actual number of unigenes classified in that descriptive term

important, which showed significant antitumor activities against several tumor lines.

To date, the genome of *Euphorbia fischeriana* Steud is still unavailable, and only a few reports on the identification of *Euphorbia fischeriana* Steud sequences; however, high-throughput sequencing is an effective strategy for the exploration of genome information using amounts of genome-wide transcription data, especially in non-model plants for which reference genome sequences are unavailable (Schmittgen and Livak 2008). These analyses could identify the candidate genes involved in complex biosynthetic pathways, thus, this work provides genomic resources for discovering candidate genes and for additional research on the molecular mechanism of pharmaceutical ingredient biosynthesis in *Euphorbia fischeriana* Steud.

In this study, the roots sequences of a *Euphorbia fischeriana* Steud cultivar with high levels of pharmaceutical ingredients were sequenced on the Illumina HiSeq™ 2000 sequencing platform. A total of 91,413,326 high-quality reads were produced, and these reads were assembled into 186,384 unigenes. The mean length and N50 showed that these reads are high-quality and up to standard, and would be used for further analysis.

As we all know, there are two biosynthesis pathway of terpenoid in high plants: mevalonic acid (MVA) pathway which is also called cytosolic pathway, and methylerythritol phosphate (MEP) pathway. These pathway synthesis two

most important precursor of terpenoid, isopentyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). In addition, *HMGR* genes, *CMK*, *PMK*, *DXR*, *DXS*, *GGPS* (geranyl pyrophosphate synthase), *FPPS* (farnesyl pyrophosphate synthase), *AATC* (acetoacetyl-CoA) and other genes are the major genes which were vital for the accumulation of terpenoid. Previous research has shown that 17-Hydroxyjolinolide A, 17-hydroxyjolinolide B, jolinolide B, jolinolide A, and fischeriana are the main components isolated from the root of *Euphorbia fischeriana* Steud and are important for anticancer effects. These components are all diterpenoids; that is to say, they may have common synthetic pathways. From the GO, KOG and KEGG analysis, these unigenes were abundant in catalytic activity and metabolic processes, and the biosynthesis of terpenoid need a number of secondary metabolism enzymes, and the further results showed that ko00904, ko00909 and ko00900 were closely linked with the biosynthesis of terpenoids (Figs. 3, 4, 5). In addition, some *CMK*, *PMK*, *DXR*, *DXS* were found in unigenes, these results provide candidate pathways and related genes for investigation of the molecular mechanisms of anti-cancer activity, and suggest that these genes of terpenoid were activated and promoted in root of *Euphorbia fischeriana* Steud.

The families of the plant-specific TFs are defined by their characteristic DNA-binding domains (DBDs), such as AP2/ERF, WRKY, NAC, MYB and so on. In addition, many

Table 3 Analysis of the transcription factor families in *Euphorbia fischeriana* Steud

Transcription factor family	Number of annotated uni-genes
MYB	454
bZIP	333
bHLH	265
NAC	262
WRKY	225
AP2	205
ARF	131
MADS	130
Trihelix	99
HSF	74
TCP	30
PLATZ	23
Others	5221
Total	7452

MYB transcription factors characterized by a highly conserved DNA-binding domain-MYB motif, *bZIP* basic region-leucine zipper, *bHLH* basic helix-loop-helix, *NAC* NAM/ATAF/CUC, *WRKY* tW-box or W-box like, *AP2* transcription factors with one or two AP2/EREBP domain, *ARF* auxin response factor, *MADS* MADS-box family, *Trihelix* helix-loop-helix-loop-helix, *HSF* heat shock factors, *TCP* Teosinte branched1/Cycloidea/Proliferating cell factor, *PLATZ* factor family with plant AT-rich sequence- and zinc-binding domain

transcription factors are involved in the biosynthesis of terpenoid and regulated related genes expression. Interestingly, in this study, several important transcription factors, including MYB, NAC, bHLH, HSF and so on, were significantly

found in root *Euphorbia fischeriana* Steud (Fig. 6), and these transcription factors may regulate terpenoid biosynthesis and anticancer effects. Boer et al. (2011) found that *ORCA3* is a member of AP2, which regulated the contents of alkaloids of terpenoid indole (TIA). *CAD1* was a key gene for artemisinin biosynthesis, and target gene of *GaWRKY1* which regulated the process of sesquiterpene in cotton (Ma et al. 2009).

At genome stage, molecular genetic markers are important for *Euphorbia fischeriana* Steud breeding and genetic linkage maps. Due to their functionality, abundance, high polymorphism and excellent reproducibility, SSRs are useful resources for genome analysis (Wayne et al. 1996). In this study, a total of 53,146 potential SSRs were identified from 186,384 unigenes and will play an important role in *Euphorbia fischeriana* Steud molecular breeding and field cultivation.

Conclusion

The results showed that a total of 186,384 unigenes were identified from the roots of *Euphorbia fischeriana* Steud using high-throughput sequencing technology. We also found some important biosynthesis pathways and a number of related enzymes genes and transcription factor, which were related to the anticancer effects of this plant. In addition, 53,146 SSRs were identified. This study not only provides candidate genes but also reveals insights into the molecular mechanism of terpenoid backbone ingredient biosynthesis in *Euphorbia fischeriana* Steud. These results provide comprehensive transcript information on the effective components in *Euphorbia fischeriana* Steud.

Fig. 6 Genome-wide distribution of different transcription factor families in *Euphorbia fischeriana* Steud. A bar graph represents the transcription factors belonging to various families. Transcripts were subjected to a BLASTx search against all the transcription factors in the PInTFDB database with an *E* value cut-off of 1×10^{-5} . If the number of transcripts encoding a particular transcription factor family was < 20 , then those families were included in the 'Others' category

Fig. 7 Types and number of SSRs distributed in *Euphorbia fischeriana* Steud

Funding This work was supported by the Fundamental Research Funds for Education Department of Heilongjiang Province (no. 2016-KYYWF-0869).

Compliance with ethical standards

Ethical approval This article does not include any studies with human participants or animals performed by any of the authors.

Informed consent This article does not involve any informed consent.

Conflict of interest The authors declare that they have no conflict of interest.

References

- Barrero RA, Chapman B, Yang Y, Moolhuijzen P, Keeble-Gagnère G, Zhang N, Tang Q, Bellaqrd MI, Qiu D (2011) De novo assembly of *Euphorbia fischeriana* root transcriptome identifies prostratin pathway related genes. *BMC Genom* 12:600–613. <https://doi.org/10.1186/1471-2164-12-600>
- Bretagne S, Costa JM, Besmond C, Carsique R, Calderone R (1997) Microsatellite polymorphism in the promoter sequence of the elongation factor 3 gene of *Candida albicans* as basis for typing system. *J Clin Microbiol* 35:1777–1780. <http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC229840&blobtype=pdf>
- Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomicsresearch. *Bioinformatics* 21:3674–3676. <https://doi.org/10.1093/bioinformatics/bti610>
- Boer KD, Tilleman S, Pauwels L, Vanden RB, De VS, Vanderhaeghen R, Hilson P, Hanmill JD, Goossens A (2011) Apetala2/ethylene response factor and basic helix-loop-helix tobacco transcription factors cooperatively mediate jasmonate-elicited nicotine biosynthesis. *Plant J* 66:1053–1065. <https://doi.org/10.1111/j.1365-3113X.2011.04566.x>
- Field D, Wills C (1996) Long, polymorphic microsatellites in simple organisms. *Proc Biol Sci* 263:209–215. <https://doi.org/10.1098/rspb.1996.0033>
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, Chen ZH, Mauceli E, Hacohen N, Gnirke A, Rhind N, Palma FD, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652. <https://www.nature.com/articles/nbt.1883.ris>
- Ha JM, Lee TY, Kim MY, Oliya BK, Gwag JG, Lee YH, Lee SH (2017) Comprehensive transcriptome analysis of *Lactuca indica*, a traditional medicinal wild plant. *Mol Breed* 37:112–123. <https://doi.org/10.1007/s11032-017-0711-z>
- He GH, Meng RH, Newman M, Gao GQ, Pittman RN, Prakash CS (2013) Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). *BMC Plant Biol* 3:3–8. <https://doi.org/10.1186/1471-2229-3-3>
- Jian BY, Zhang H, Han CC, Liu JC (2018) Anti-cancer activities of diterpenoids derived from *Euphorbia fischeriana* Steud. *Molecules* 23:387–397. <https://doi.org/10.3390/molecules23020387>
- Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40:D109–D114. <https://doi.org/10.1093/nar/gkr988>
- Kirby J, Nishimoto M, Park JG, Withers ST, Nowroozi F, Behrendt D, Rutledge EJ, Fortman JL, Johnson HE, Anderson JV, Keasling JD (2010) Cloning of casbene and neocembrene synthases from Euphorbiaceae plants and expression in *Saccharomyces cerevisiae*. *Phytochemistry* 71:1466–1473. <https://doi.org/10.1016/j.phytochem.2010.06.001>
- Kuang XZ, Li W, Kanno Y, Yamashita N, Nemoto K, Asada Y, Koike K (2016) ent-Atisane diterpenoids from *Euphorbia fischeriana* inhibit mammosphere formation in MCF-7 cells. *J Nat Med* 70:1–7. <https://doi.org/10.1007/s11418-015-0940-6>
- Li W, Li L, Chen ZJ (2014) Genetic diversity of germplasm resources of *Euphorbia Fischeriana* Steud based on ISSR makers. *J Liaoning Univ Tradit Chin Med*. 16:50–52. <https://doi.org/10.13194/j.issn.1673-842x.2014.03.018> (in Chinese)
- Liu TM, Zeng LB, Zhu SY, Chen XJ, Tang QM, Mei SY, Tang SW (2015) Large-scale development of expressed sequence tag-derived simple sequence repeat markers by deep transcriptome sequencing in garlic (*Allium sativum* L.). *Mol Breed* 35:204–212. <https://doi.org/10.1007/s11032-015-0399-x>
- Ma DM, Pu GB, Lei CY, Ma LQ, Wang HH, Guo YW, Chen JL, Du ZG, Wang H, Li GF, Ye HC, Liu BY (2009) Isolation and characterization of AaWRKY1, an *Artemisia annua* transcription factor that regulates the amorpho-4, 11-diene synthase gene, a key gene of artemisinin biosynthesis. *Plant Cell Physiol* 50:2146–2161. <https://doi.org/10.1093/pcp/pcp149>
- Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* 3:1101–1108. <https://doi.org/10.1038/nprot.2008.73>
- Shen L, Zhang SQ, Liu L, Sun Y, Wu YX, Xie LP, Liu JC (2017) Jolkinolide A and jolkinolide B inhibit proliferation of A549 cells and activity of human umbilical vein endothelial cells. *Med Sci Monit* 23:223–237. <https://www.medscimonit.com/abstract/index/idArt/902704>
- Sun YX, Liu JC (2011) Chemical constituents and biological activities of *Euphorbia fischeriana* Steud. *Chem Biodivers* 8:1205–1214. <https://doi.org/10.1002/cbdv.201000115>
- Uemura D, Katayama CJ, Hirata Y (1997) Crystal and molecular structure of jolkinolide B, a novel oxidolactone diterpene. *Tetrahedron Lett* 3:283–284. [https://doi.org/10.1016/S0040-4039\(01\)92614-0](https://doi.org/10.1016/S0040-4039(01)92614-0)
- Wang Y, Ma XQ, Yan SS, Shen SS, Zhu HL, Gu Y, Wang HB, Qin GW, Yu Q (2009) 17-Hydroxy-jolkinolide B inhibits signal transducers

- and activators of transcription 3 signaling by covalently cross-linking Janus Kinases and induces apoptosis of human cancer cells. *Cancer Res* 69:7302–7310. <https://doi.org/10.1158/0008-5472.CAN-09-0462>
- Wang HB, Chen W, Zhang YY, Wang XY, Liu LP, Chen Y (2013) Four new diterpenoids from the roots of *Euphorbia fischeriana*. *Fito-terapia* 91:211–216. <https://doi.org/10.1016/j.fitote.2013.09.003>
- Wang C, Zhang X, Yan X, Ye W, Ma S, Jiang Y (2017) Chemical profiling of *Euphorbia fischeriana* Steud by UHPLC-Q/TOF-MS. *J Pharm Biomed Anal* 151:126–132. <https://doi.org/10.1016/j.jpba.2017.12.051>
- Wayne P, Michele M, Chaz A, Michael H, Julie V, Scott T (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol Breed* 2:225–238. <https://doi.org/10.1007/BF00564200>
- Ye J, Fang L, Zheng HK, Zhang Y, Chen J, Zhang ZJ, Wang J, Li ST, Li RQ, Bolund L, Wang J (2006) WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res* 34:W293–W297. <https://doi.org/10.1093/nar/gkl031>
- Zhang JZ (2003) Overexpression analysis of plant transcription factors. *Curr Opin Plant Biol* 6:430–440. [https://doi.org/10.1016/S1369-5266\(03\)00081-5](https://doi.org/10.1016/S1369-5266(03)00081-5)
- Zhang B, Zhang W, Nie RE, Li WZ, Segraves KA, Yang XK, Xue HJ (2016) Comparative transcriptome analysis of chemosensory genes in two sister leaf beetles provides insights into chemosensory speciation. *Insect Biochem Mol Biol* 79:108–118. <https://doi.org/10.1016/j.ibmb.2016.11.001>