Transcriptome analysis provides insights into key gene(s) involved in steroidal alkaloid biosynthesis in the medicinally important herb Fritillaria taipaiensis

Xiaoyue Wang
Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Medicinal Plant Development https://orcid.org/0000-0003-3928-0997

Hong-mei Luo
Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Medicinal Plant Development

Xue-min Wei
Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Medicinal Plant Development

Pei Cao
Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Medicinal Plant Development

Zi-tong Gao
Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Medicinal Plant Development

Jian-ping Han (happymyra2007@163.com)

Research article

Keywords: Fritillariae Cirrhosae Bulbus, Fritillaria taipaiensis, steroidal alkaloids, biosynthesis pathway, CYP450s

Posted Date: June 26th, 2019

DOI: https://doi.org/10.21203/rs.2.10690/v1

Abstract

Background Fritillariae Cirrhosae Bulbus, which is a traditional antitussive and expectorant medicine in China, is derived from six original plants. Among these six species, all except for Fritillaria taipaiensis and Fritillaria unibracteata var. wabuensis have been recorded in the National Protected Plants III. Their major bioactive compounds, including imperialine and peimisine, are steroidal alkaloids, derived from the steroid synthesis pathway. However, research on the F. taipaiensis genome, transcriptomes and steroid-synthesis-related genes has been lacking for a long time. In this study, the first transcriptome sequencing of F. taipaiensis was performed. Results A total of 94,396,694 (~11.4 Gb) high-quality reads obtained using paired-end Illumina sequencing were de novo assembled into 190,350 transcripts. Functional annotation with multiple public databases identified an array of genes involved in steroidal alkaloid biosynthesis and other secondary metabolite pathways, including steroid and terpenoid backbone biosynthesis, and important oxidoreductases. The large number of differentially expressed transcripts together with CYPs suggests the involvement of these candidates in tissue-specific expression. The expression analysis revealed that bulbs may be the main site of the upstream steroidal alkaloid biosynthesis pathway, and we speculated that the downstream reactions, including the oxidation-reduction reaction catalyzed by CYPs, might also occur primarily in bulbs. Conclusion In conclusion, the comprehensive transcriptome dataset created in this study will serve as a resource for the identification of potential candidates for the genetic manipulation of targeted bioactive metabolites and will also contribute to the development of functionally relevant molecular marker resources to expedite molecular breeding and conservation efforts for F. taipaiensis.

Background

Fritillariae Cirrhosae Bulbus is a kind of traditional Chinese medicine and a famous food that has been used in China for thousands of years due to its antitussive, expectorant and anti-asthma activities [1]. Fritillariae Cirrhosae Bulbus is derived from the dried bulbs of six species from the *Fritillaria* genus, including *Fritillaria cirrhosa* D. Don, *F. unibracteata* Hsiao et K. C. Hsia, *F. przewalskii* Maxim., *F. delavayi* Franch., *F. taipaiensis* P. Y. Li and *F. unibracteata* Hsiao et K. C. Hsia var. *wabuensis* (S. Y. Tang et S. C. Yue) Z. D. Liu, S. Wang et S. C. Chen. As haze pollution has worsened in recent years in China, the demand for Fritillariae Cirrhosae Bulbus has increased dramatically [2]. The various plants of this medicine are mainly distributed in the western region of China, including Tibet, Yunnan, Sichuan, Qinghai and Gansu provinces, at an altitudinal range of 1800 to 4700 m [3]. Due to the increasing demand for Fritillariae Cirrhosae Bulbus and the long growing period of its plant constituents, many populations remain at risk of being depleted [4]. Among the six species mentioned above, all except for *F. taipaiensis* have been recorded in the National Protected Plants . Artificial cultivation may be an important way to resolve the current imbalance between supply and demand. According to a previous study, *F. taipaiensis* is the most suitable species for cultivation at low altitudes [5].

To date, a variety of isosteroidal alkaloids, mainly including imperialine, peimisine, peimine, and peiminine, have been isolated and described from *Fritillaria* species [6-9]. Previous investigations on the

pharmacology and phytochemistry suggested that the steroidal alkaloids, especially isosteroidal alkaloids, were responsible for the antitussive, expectorant, anti-inflammatory and antinociceptive effects of these herbs [10-14]. *F. taipaiensis* is no different from *F. cirrhosa* as a commodity due to their similar chemical constituents [15, 16]. Steroidal alkaloids are mainly synthesized from squalene through isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) via cytosolic mevalonate (MVA) and plastidial 2-C-methyl-D-erythritol 4-phosphate (MEP/DOXP) pathways, respectively. Furthermore, several hydroxylation steps catalyzed by CYP450s contribute to the structural and functional diversification of various steroidal alkaloids [17, 18]. However, due to the nonavailability of genomic resources, this medicine,*F. taipaiensis*, is less studied at the molecular level. Until now, the steroidal alkaloid biosynthesis pathway in this medicinal plant has been unclear.

RNA sequencing (RNA-Seq) is a common approach to transcriptome profiling that uses next-generation sequencing (NGS) technologies. Compared with other methods, this approach has several significant advantages, such as being able to reveal the precise location of transcription boundaries at a single-base resolution and avoiding background signals with the detection of digital signals, while requiring less sample RNA and not requiring the design of specific probes with existing genomic sequences. Thus, RNA-Seq has been shown to be an efficient tool for genome-wide transcriptome profiling and elucidation of important candidates involved in complex biosynthetic pathways, irrespective of genome complexity, even in the case of nonmodel plant species [19]. RNA-Seq has been used to discover key enzyme genes in various species. Meanwhile, many tools have been developed to increase the data credibility, such as trinity and cd-hit.

In this study, a comprehensive transcriptome of *F. taipaiensis* was sequenced using the RNA-Seq approach to identify key genes involved in the complex steroidal alkaloid pathway, and qRT-PCR was used to verify gene expression levels. Current data provide the first genome-wide transcriptional insights into the functions of steroidal alkaloid genes and their differential expression in bulbs or leaves of *F. taipaiensis* of different ages. In the future, the outcome of available data will serve as a resource to expedite cutting edge research for the upscaling of targeted secondary metabolite production through genetic engineering of *F. taipaiensis*. Efforts were also made to identify potential transcription factors and CYPs that play key roles in the regulation and diversification of secondary metabolites.

Methods & Materials

Sample collection for RNA-Seq and HPLC-ELSD

The cultivated plant materials were collected from Hongchiba Natural Forest Park, Wuxi County, Chongqing municipality, China (31°31'49"N, 109°6'29"E) at an altitude of 1816 m. Three genotypes for different year-old tissues located at a distance of 10 m from each other were randomly considered. Bulb or leaf tissues were collected from each genotype, and the details are presented in Table 1. The collected samples of *F. taipaiensis* were first washed under running tap water. Then, the surface of samples was sterilized with 75% ethanol for RNA-Seq. All samples were stored in RNA Later liquid (Beijing HT-biotech

Co., Ltd., China) and frozen immediately in dry ice and stored at -80°C until RNA isolation. Samples for HPLC-ELSD were dried at 45°C in a drying oven (101FXB-2, Shanghai, China) until a constant weight was achieved. Subsequently, the dried samples were powdered to a homogeneous size, sieved through a No. 4 mesh and then stored at room temperature.

RNA isolation, sequencing and *de novo* assembly

Total RNA was extracted from individual samples using a Plant Universal Total RNA Kit (BioTeke Biotech Beijing Co., China) according to the manufacturer's instructions. The concentration of RNA was determined using an Agilent 2100 Bioanalyzer system (Agilent Technologies Inc., USA), and integrity was checked on a denaturing agarose gel. Equimolar concentrations of RNA from three genotypes for each tissue were pooled together for RNA-Seq library preparation to remove biological bias.

RNA-Seq libraries were prepared using the NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs, Inc., UK) according to the manufacturer's instructions. The libraries were purified using Beckman AMPure XP beads (Beckman Coulter, Inc., USA), and an Agilent 2100 Bioanalyzer (Agilent Technologies, USA) was used for library size estimation. An ABI 7500 real-time PCR system (Applied Biosystems Co., USA) with KAPA SYBR green fast qPCR Kits (Kapa Biosystems, USA) was used to quantify the libraries. Furthermore, for cluster generation, 10 pM of these libraries were loaded onto the flow cell of a cluster station using a TruSeq PE Cluster Kit (Illumina Inc., USA). Clonally amplified clusters were used for paired-end (PE) (2 × 125) sequencing using HiSeq 2500 with a TruSeq SBS Kit v3-HS (Illumina). Trimmomatic (v 0.30) was used to filter raw reads, and reads with 99% probability of no error (minimum phred score 20 for each base) were utilized for assembly [20]. . Trinity (v r20140717) was used for *de novo* transcriptome assembly with default parameters and a minimum transcript length of 200 base pairs [21]. Cd-hit was used to cluster biological sequences to reduce sequence redundancy and obtain unigenes [22]. TransDecoder (v r20130225), which comes included in the Trinity software distribution, was used to perform open reading frame annotation [21].

Functional annotation and classification

To find the putative functions of assembled transcripts of *F. taipaiensis*, a similarity search using BLASTx [23] was performed against publicly available protein databases, including NCBI nonredundant protein database (Nr) and Swiss-Prot with an e-value cut-off of $\leq 1e^{-5}$. *F. taipaiensis* transcripts were classified into three major categories: biological process, cellular component and molecular function according to Gene Ontology (GO) terms using WEGO software (http://wego.genomics.org.cn/). Transcripts were further functionally categorized into different classes using the COG database (ftp://ftp.ncbi.nih.gov/pub/COG/COG). The transcription factor (TF)-encoding transcripts were identified based on a similarity search against the Plant Transcription Factor Database (http://planttfdb.cbi.pku.edu.cn). Biochemical pathways were assigned to the transcripts by the

bidirectional best hit method on the KAAS (KEGG Automatic Annotation Server) (http://www.genome.jp/tools/kaas/). CYPs present in *F. taipaiensis* were identified from Swiss-Prot.

Identification of differentially expressed genes

To measure the expression pattern of transcripts in each tissue, high-quality reads were mapped onto the final *de novo* assembled transcripts of *F. taipaiensis* using RSEM v 1.2.4 [24]. The expression level for each transcript was measured in terms of fragments per kilobase per million reads (FPKM) by normalizing read counts. Furthermore, the edgeR package was used to evaluate the differential gene expression using read counts in pairwise comparisons of different year-old bulbs [25]. Based on statistical analysis, genes having a p-value < 0.05 and log2-fold change ≥ 2 were considered differentially expressed genes. A heatmap representing the tissue-specific gene expression pattern (log2-fold change) for different pathways was generated using Multiple Experiment Viewer (MEV v4.9.0).

Quantitative real-time PCR (qRT-PCR) analysis

gDNA Eraser (Takara Bio Co., USA) treatment was performed on total RNA to remove DNA contamination. First-strand cDNA was synthesized from 1 μg of the total RNA using the PrimeScript RT reagent kit (Takara, Japan) according to the manufacturer's instructions and was diluted by 10X. Gene-specific primers (Table 2) were designed using Primer3 Input (http://primer3.ut.ee) and synthesized by SinoGenoMax Co., Ltd, China. The relative expression of steroidal alkaloid pathway genes in seven samples was analyzed on a CFX96 Real-time System (Bio-Rad Lab., USA) using SYBR® *Premix Ex Taq*™ (Tli RNaseH Plus) (Takara, Japan). qRT-PCR was performed with three technical replicates based on which standard error was calculated. *β*-actin was used as a reference gene for establishing equal amounts of cDNA in each reaction. The relative gene expression and fold change was calculated using the $2^{-\Delta\Delta CT}$ method with rhizome as a control tissue [26].

Chemicals and reagents

Alkaloid standards, including peimisine (CAS-19773-24-1), sipeimine (CAS-61825-98-7), peimine (CAS-23496-41-5) and peiminine (CAS-18059-10-4), were purchased from Chengmust Biotech Co., Ltd. (Chengdu, Sichuan Province, China). The purities of these compounds were determined to be more than 98.0% by HPLC. Methanol for liquid chromatography was purchased from Fisher Scientific (Thermo Fisher Scientific, USA). Other reagents were of HPLC grade or the highest grade commercially available.

Solution preparation

Stock solutions (500 μg/ml) of the alkaloid standards peimisine, sipeimine, peimine and peiminine were prepared. Total alkaloids of samples were prepared as follows: the dried bulb powders were soaked in 4

mL ammonium hydroxide for 2 h and then reflux extracted with 60 mL of a mixture of chloroform:methanol (80:20, v/v) at 80℃ for 3 h. After extraction, the total extracted solution was filtered through qualitative filter paper. Subsequently, the filtrate was concentrated to dryness in an evaporating dish under a fume hood. The residue was dissolved to exactly 2 mL with methanol using a volumetric flask. The solution above was filtered through a 0.22 μm membrane filter prior to injection into the HPLC system.

Instrumentation and HPLC conditions for determining alkaloids in *F. taipaiensis*

Chromatographic analysis was performed on a Waters 2695 Series high-performance liquid chromatography system (Milford，MA，USA) equipped with a binary pump, an autosampler, an online degasser and a column temperature controller. An Alltech ELSD 2000 evaporative light scattering detector (Alltech，Deerfield，IL，USA) connected to an Allchrom Plus Client/server software and an Agilent Extend-C18 column (250 mm × 4.0 mm, 5 μm) were used for chromatographic separation at 30℃. The mobile phases were composed of methanol (A) and water containing 0.02% triethylamine (B). The gradient elution process used was 0–15 min, 35–55% A; 15–30 min, 55–72% A; 30–37 min, 72–75% A; 37–40 min, 75–80% A; 40–50 min; 80–85% A; 50–55 min, 85–90% A; and 55–60 min, 90–35% A, followed by a 10 min post run for re-equilibration of the column. The flow rate was kept constant at 1.0 mL/min, and ten microliters of each sample solution and standard solution were injected for analysis. With these conditions, representative HPLC chromatograms of the standards and the alkaloids were obtained.

Results

RNA sequencing and *de novo* assembly

In this study, Illumina HiSeq 2500 was used to sequence the cDNA libraries of bulb tissues for elucidating secondary metabolite biosynthesis in *F. taipaiensis*. The PE sequencing of three different libraries resulted in 108,731,004 raw reads, ranging from 33 to 37 million reads for each library. After quality filtering removal of adaptor sequences, ambiguous and low-quality reads, 94,396,694 (~11.4 Gb) high-quality reads, 150 bp in length with quality scores of >Q20, were obtained. *De novo* assembly of clean reads resulted in 190,350 nonredundant (NR) assembled transcripts (~101.20 Mb). The length of the transcripts varied from 201 to 15,727 bp, with an average of 532 bp, N50 (698 bp) and GC content of 46.15%.

Functional annotation and classification

Functional annotation of the assembled transcripts allows for insight into the particular molecular functions, cellular components, and biological processes in which the putative proteins are involved. As a

nonmodel plant, to obtain the optimum annotations, assembled transcripts of *F. taipaiensis* were aligned with five public protein databases. Annotations of 190,350 transcripts identified 94,172 (49.47%; NCBI's nr), 51,212 (26.90%; GO), 66,901 (35.15%; Swiss-Prot), 48,205 (25.32%; COG) and 18,493 (9.72%; KEGG) transcripts showing putative functions. Interestingly, 15,833, 735, 681 and 3 transcripts were uniquely annotated to Nr, Swiss-Prot, COG and KEGG, respectively (Figure 1). Due to the nonavailability of genomic and transcriptomic resources in targeted species, 93,694 (49.22%) transcripts could not be annotated to any of the searched databases.

GO has been widely used for functional analysis and inferring the biological significance of genomic and transcriptomic datasets, in which there are three ontologies: molecular functions, biological processes and cellular components. A total of 51,212 transcripts were assigned 4954 GO terms, of which 1929 terms were categorized into molecular functions, 2529 were categorized into biological processes and 496 were categorized into cellular components. Among the molecular functions, the most abundant of the second-level GO terms was related to catalytic activity (69.35%), followed by binding (12.80%) and transporter activity (11.05%). Among the cellular components, organelle parts (26.80%) and cell parts (26.30%) were the most represented in the second level, followed by protein-containing complexes (23.64%) and membrane parts (11.00%). In biological processes, cellular processes (22.37%) and metabolic processes (20.66%) were abundant, followed by single-organism processes (19.28%) and biological regulation (16.04%) (Figure 2).

Further, to assess the competence of *de novo* assembly and effectiveness of the annotation process, alignment of the transcripts with the COG database resulted in 48,205 annotated transcripts. Of these, 43,095 transcripts were uniquely classified into 25 COG categories, while the remaining 5,110 were annotated with multiple COG functions and hence cannot be classified into any category. The general function prediction with 5,152 (11.95%) transcripts was clearly the major COG category, followed by posttranslational modification, protein turnover, chaperones (5,122 transcripts, 11.89%) and signal transduction mechanisms (4,520 transcripts, 10.49%). Interestingly, a total of 1,439 (3.34%) transcripts were assigned to the secondary metabolite biosynthesis, transport and catabolism categories, which may influence the secondary metabolite biosynthesis of *F. taipaiensis* (Figure 3).

KEGG annotations provide background information on active metabolic processes within an organism, hence enabling further understanding of the biological function of the transcripts [27]. To elucidate active biosynthesis pathways in *F. taipaiensis*, annotation of NR data with the KEGG database showed 18,493 transcripts comprising 4,272 unique KO identifiers. Of these, 12,494 transcripts with 2,724 unique KO identifiers were assigned to six main categories representing 359 biological pathways. The highest number of KO identifiers was involved in metabolism (12,503), followed by human diseases (6,228), genetic information processing (5,260), organismal systems (3,931), environmental information processing (2,795) and cellular processes (2,598). Pathways with the largest number of KO identifiers were signal transduction (2,698), carbohydrate metabolism (2,681), translation (2,245), global and overview maps (1,857), and amino acid metabolism (1,613). Interestingly, a significant number of genes

involved in the biosynthesis of other secondary metabolites (455) and the metabolism of terpenoids and polyketides (366) were also identified (Figure 4).

Age-related bulb-specific differential gene expression

To understand the key putative regulators involved in steroidal alkaloid biosynthesis, specific gene expression was measured using DESeq in the Bioconductor program. Transcripts with log2-fold change (FC) > 2 and a p-value ≤ 0.05 were considered differentially expressed genes (DEGs). Pairwise comparison of transcripts in bulbs of different ages resulted in 4607 DEGs, including 1883 in one-year-old bulbs vs two-year-old bulbs (709 upregulated and 1174 downregulated), 2555 in one-year-old bulbs vs four-year-old bulbs (668 upregulated and 1887 downregulated), and 2412 in two-year-old bulbs vs four-year-old bulbs (1381 upregulated and 1031 downregulated). According to the annotated COG functions, 1813 of 4607 DEGs were uniquely classified into 22 COG categories (without cell motility, extracellular structures, and nuclear structure categories), and 220 DEGs were not classified into any category due to having multiple functions. Of these, the general function prediction with 225 DEGs (12.41%) and posttranslational modification, protein turnover, and chaperones with 224 DEGs (12.36%) were clearly the main categories, followed by signal transduction mechanisms (133 DEGs, 7.33%) (Figure 5). A total of 1,453 DEGs were assigned 1438 GO terms, of which 546 terms were categorized into molecular functions, 172 were categorized into biological processes and 720 were categorized into cellular components. Interestingly, some CYP 450 genes were found to be differentially expressed, including CYP71 and CYP734 families. Furthermore, a total of 520, 891 and 985 unique DEGs were obtained in one-year-old bulbs vs two-year-old bulbs, one-year-old bulbs vs four-year-old bulbs and two-year-old bulbs vs four-year-old bulbs, respectively.

Secondary metabolic pathway analysis

Metabolic pathway analysis enables us to understand the interactions of genes in particular pathways and their related biological functions. A total of 19 pathways (174 entries) related to secondary metabolite biosynthesis were identified from the KEGG database. The identification of these pathways helped us to analyze secondary metabolite biosynthesis in *F. taipaiensis*. Of these, five major pathways, carotenoid, steroid, ubiquinone and other terpenoid-quinone, terpenoid backbone, and phenylpropanoid biosynthesis pathways, were well represented in our data. Out of 515 transcripts involved in these pathways, 33 recorded specific differential expression in bulb tissues from different growth years. Furthermore, most of the genes involved in these pathways were highly expressed in one-year-old bulbs.

Steroidal alkaloid pathway gene expression and content analysis

In plants, steroidal alkaloids are mainly synthesized from lanosterol and cycloartenol via cholesterol and sitosterol, respectively [28, 29]. . The genes related to steroidal alkaloids via cholesterol have not been characterized so far; therefore, genes related to steroidal alkaloid biosynthesis via sitosterol were considered in this study, which comprises two parts: terpenoid backbone and steroid biosynthesis, according to KEGG classification. Terpenoid backbone biosynthesis primarily involves the synthesis of dimethylallyl pyrophosphate (DMAPP) and isopentenyl diphosphate (IPP) from the mevalonate pathway (MVA) and 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate (MEP/DOXP) pathway. A total of 15 genes (27 transcripts) involved in the MVA and MEP/DOXP pathways were identified, including acetyl-CoA C-acetyltransferase (ACAT), hydroxymethylglutaryl-CoA synthase (HMGS), hydroxymethylglutaryl-CoA reductase (NADPH) (HMGR), mevalonate kinase (MVK), phosphomevalonate kinase (PMVK), diphosphomevalonate decarboxylase (MVD), 1-deoxy-D-xylulose-5-phosphate synthase (DXS), 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR), 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (ispD), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (ispE), 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (ispF), (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ispG) and 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase (ispH). Additionally, DMAPP and IPP can interconvert under the catalysis of isopentenyl diphosphate delta-isomerase (IDI). Subsequently, farnesyl diphosphate (FPP) is produced through sequential head-to-tail condensation of IPP and DMAPP catalyzed by geranylgeranyl diphosphate synthase (GGPS) and farnesyl diphosphate synthase (FDPS) [30] Interestingly, not all the genes involved in the current known pathway were identified in this study. For instance, there are two types of HMGR involved in the MVA pathway, which can be differentiated according to the consumption of NADPH, and only one type was identified.

In total, 14 genes corresponding to 22 transcripts involved in the steroid biosynthesis pathway were identified in our study. In this biosynthesis pathway, farnesyl diphosphate farnesyltransferase (FDFT1), which is also called squalene synthase (SQS), catalyzes the first two-step reaction in which two identical molecules of farnesyl pyrophosphate (FPP) are converted into squalene. Then, squalene can be converted into 2,3-oxidosqualene based on oxidation by squalene epoxidase (SQLE), which is the branching point between triterpenoid and steroid [31] . During steroid biosynthesis, the chair-boat-chair-boat conformational change of 2,3-oxidosqualene results in the formation of cycloartenol, which ultimately leads to the synthesis of steroidal alkaloids through various modifications catalyzed by CYPs, isomerase, methyltransferases, reductase, etc.

To validate the RNA-Seq data and study the specific expression among different tissues and growth years, 19 genes involved in the steroidal alkaloid biosynthesis pathway were selected for quantitative real-time PCR (qRT-PCR) analysis, and the results are shown in Figure 6. In terpenoid backbone biosynthesis, the expression levels of all MEP/DOXP pathway genes, including *DXS*, *DXR*, *ispD*, *ispE*, *ispF*, *ispG* and *ispH,* were found to be clearly higher in leaves than in bulbs, whereas six genes involved in the MVA pathway, including *ACAT*, *HMGR*, *HMGS*, *MVK*, *PMVK* and *MVD,* showed diametrically opposite trends. However, the *IDI* gene in the next step, which can convert DMAPP and IPP to each other, showed no clear trend in bulbs or leaves. Interestingly, GGPS and FDPS were described as having the same reaction to FPP, but they showed completely different trends: GGPS was highly expressed in leaves,

whereas FDPS was clearly more highly expressed in bulbs than in leaves. In the steroid biosynthesis pathway, FDFT1 and SQLE were both highly expressed in leaves, while the expression level of CAS was similar in leaves or bulbs, and it was relatively high in the three-year-old bulb.

The HPLC-ELSD method was used in this study to quantify 4 alkaloids from 11 samples of *F. taipaiensis* in 5 growth stages; the analyte contents are shown in Table 3. Among the four alkaloid standards, only peimisine was detected in bulbs of *F. taipaiensis,* and the content ranged from 0.0230% to 0.0660%. The quantitative results showed that the peimisine content in *F. taipaiensis* increased with age from one to three years. However, when the plants were cultivated for over 4 years, the content decreased with age. Therefore, the peimisine content of the three-year-old bulb was the highest in either the 2016 or 2017 samples. Meanwhile, some samples could be regarded as the same sample taken in different years due to the continuous sampling within two years, such as three-year-old bulbs in 2016 and four-year-old bulb in 2017s. A longitudinal comparison was carried out within these samples (Figure 7). The results showed that the peimisine content in the three-year-old bulb (2016) was significantly higher than that of the four-year-old bulb (2017). Meanwhile, alkaloids in four-year-old leaves were extracted and detected, but no peak was obtained, which means that the steroidal alkaloid content in leaves was below the detection limit.

Cytochrome P450s and some other oxidoreductase genes

Plant cytochrome P450 (CYP450s) is a B-type cytochrome protease gene superfamily with heme as a prosthetic group. CYP450s catalyze a wide variety of monooxygenation/hydroxylation reactions in metabolic reaction pathways, including those for cinnamates, fatty acids, hormone homeostasis, lignin, terpenoids, sterols, alkaloids, and saponins [32] . A total of 205 CYP genes corresponding to 620 transcripts classified under 58 families were identified in *F. taipaiensis*. Among the various CYPs, the CYP71 family was the most abundant. In total, 42 CYP genes related to 17 families were found to be differentially expressed in at least one pairwise comparison. The tissue-specific expression of these CYP genes revealed that 21, 13, and 8 genes were highly expressed in one-year-old, two-year-old, and four-year-old bulbs, respectively.

Furthermore, some modification genes that may be involved in the steroidal alkaloid biosynthesis pathway were analyzed by qRT-PCR (Figure 8). Sterol 14-demethylase (CYP51G1), steroid 22-alpha-hydroxylase (CYP90B1, DWF4), sterol delta-14 reductase (FK), delta (24)-sterol reductase (DIM, DWF1) and 7-dehydrocholesterol reductase (DWF5) were highly expressed in bulbs. Especially for CYP90B1 and DWF1, both genes were barely expressed in leaves but had high expression levels in the bulbs, and the expression levels were fifty to one thousand times that of the leaves. Interestingly, similar to the CAS gene, CYP90B1 showed the highest expression in three-year-old bulbs in both 2016 and 2017 samples. Moreover, the five genes above shared one common trait: the expression level was lowest in the perennial bulbs. In addition, acyl-CoA synthetase (ACS12) and CYP734A6 were more highly expressed in leaves than bulbs, but 3-epi-6-deoxocathasterone 23-monooxygenase (CYP90D1) had no clear trends of

expression levels in bulbs or leaves. The qRT-PCR result matches the RNA-Seq sequencing result by 55.56%.

Transcription factors

TFs are major regulatory elements that play significant roles in gene expression, plant secondary metabolism and response to environmental stress by binding to specific cis-regulatory elements of the promoter regions. TF families, including bHLH, bZIP, MYB, NAC, and WRKY, were reported to be involved in the regulation of secondary metabolites and abiotic and biotic stress responses in many plant species [33, 34] . In our study, a total of 1,362 transcripts were assigned to 62 TF families. Among these, WRKY (165), AP2/ERF (161), bHLH (158), MYB-related (84), GATA (51), NAC (20), and bZIP (18) were found to be the most abundant and were possibly involved in the regulation of various physiological processes and biochemical pathways in *F. taipaiensis*.

Discussion

Fritillariae Cirrhosae Bulbus is a well-known traditional Chinese medicine derived from six *Fritillaria* species. Among these, steroidal alkaloid biosynthesis genes of only *F. cirrhosa* have been described because it is the primary plant source for Fritillariae Cirrhosae Bulbus. Sun et al. identified five genes, including HMGR and FPPS, that are involved in steroid biosynthesis from the expressed sequence tags (ESTs) database of *F. cirrhosa* bulbs [35] . However, due to the low throughput of EST, genes of *Fritillaria* species involved in the alkaloid biosynthesis pathway presented a research gap. Until now, however, no study has carried out transcriptome sequencing of the other five species involved in Fritillariae Cirrhosae Bulbus.

The RNA-Seq approach based on NGS technology has a wide variety of applications, including in plant, animal and human cells [36-40] . Moreover, data analysis software has been developed for *de novo* assembly, which overcame the shortcomings of short reads, making assembly of transcriptome results from nonreference genomic species reliable [21] . In this case, the time was ripe for transcriptome sequencing of *F. taipaiensis*, and we present the first Illumina-based RNA-seq derived transcriptome for this species. The pairwise reads (33 million to 37 million) obtained from this study are sufficient for reliable *de novo* transcriptome characterization and accurate quantification of gene expression patterns.

In plants, a variety of metabolites are derived formally from isoprenoid units, including terpenoids, steroidal alkaloids, carotenes, phytol and gibberellin. Isosteroidal alkaloids, belonging to steroidal alkaloids, are the main pharmacodynamic components of *F. taipaiensis*. In higher plants, IPP, as the common precursor of all isoprenoid compounds, is synthesized via two different pathways: MVA in the cytoplasm and MEP/DOXP in plastids [41]. From the leaf and bulb samples, it was evident that chloroplasts are common to leaves, but few are found in bulbs; this difference may explain why the genes in the MEP pathway were upregulated in leaves, whereas genes in the MVA pathway showed higher expression levels in bulbs.

FDPS can catalyze the synthesis of geranyl diphosphate (GPP), FPP and geranylgeranyl pyrophosphate (GGPP) from IPP and DMAPP, which can be further synthesized into various metabolites. It is generally accepted that the MVA pathway mainly synthesizes sesquiterpenoids, triterpenoids and steroids [42]), while the MEP pathway is mainly related to the synthesis of monoterpenes, diterpenes, phytol, gibberellin, and carotenoid. In this study, the results of the HPLC-ELSD experiment demonstrated that isosteroidal alkaloids mainly accumulate in bulbs of *F. taipaiensis*. Furthermore, previous studies have found that genes in the MEP pathway play a major role in coordinating the synthesis of chloroplast-localized isoprenoid-derived compounds, such as carotenoid and chlorophyll [43] . Thus, it is speculated that the IPP required by the isosteroidal alkaloids mainly comes from the MVA pathway in bulbs of *F. taipaiensis*.

The *CAS* gene, which catalyzes the formation of the sterol cycloartenol by cyclization of 2,3-epoxysqualene, is a key enzyme gene in the steroid biosynthesis pathway [44, 45] . This catalytic reaction could provide precursors for the biosynthesis of sterols and triterpenoids and is considered to be a key site for the biosynthesis branch of sterols and triterpenoids. More interestingly, the relative peimisine content was highest in the three-year-old bulbs, which matched the expression trend of the *CAS* gene in this study. Thus, it could be speculated that the high expression of enzymes related to the steroid biosynthesis pathway leads to a high accumulation of steroidal alkaloids in the third year in bulbs.

The precursors, such as cycloartenol, need to undergo a series of redox reactions to form the secondary metabolites. Thus, some oxidoreductases, such as the cytochrome P450 enzymes (CYP450), play an essential catalytic role in redox reactions. The CYP450 superfamily is a large and diverse group of membrane-bound hemoproteins belonging to monooxygenase, which is the first enzyme class to be classified as a superfamily [46, 47]; it is also an ancient superfamily, including more than 1,000 families and 2,500 subfamilies [48] . In plants, CYP450s have multiple subfamily members, including CYP51, CYP71, CYP99, CYP701, CYP736, etc. CYP90B1/DWF4, which belongs to family 90, subfamily B, has been confirmed to catalyze the C22-alpha-hydroxylation step in brassinosteroid biosynthesis by converting campestanol to 6-deoxocathasterone and 6-oxocampestanol to cathasterone. Fujita et al. confirmed that CYP90B1 could catalyze the C-22 hydroxylation of various C_{27}, C_{28} and C_{29} sterols [49] . Interestingly, this enzyme gene was discovered in this study, and qRT-PCR results demonstrated that it was only highly expressed in bulbs, and the expression level corresponded with the trend of peimisine accumulation. Thus, we hypothesize that this gene may catalyze C-22 hydroxylation in the biosynthesis of *F. taipaiensis* alkaloids. CYP51G1, a type of multifunctional oxidase known to be involved in sterol and steroid biosynthesis, was also identified in this study [50, 51] . CYP51G1/CYP51A2, known as sterol 14-demethylase, is involved in sterol biosynthesis, belongs to family 51, subfamily A and catalyzes the reaction removing methyl from the C-14 of certain phytosterols. Kim et al. noted that this enzyme plays a crucial role in controlling plant growth and development through a sterol-specific pathway [52].

In addition to the CYP450 enzyme genes, some other oxidoreductase genes have also been verified by qRT-PCR, and *DWF1/DIM* has been the most prominent among all genes. DWF1/DIM is a bifunctional protein that is necessary for both the isomerization and reduction in 24-methylenecholesterol in the brassinosteroid biosynthesis pathway [6] . Sumeer et al. noted that increasing the transcript levels of

DWF1 could enhance the accumulation of withanolides—a group of steroidal lactones, possibly by affecting downstream phytosterol biosynthesis [53] . According to the significant difference in the expression levels of this gene between different tissues, we suspect that DWF1 also has a certain function in the steroidal alkaloid synthesis pathway of *F. taipaiensis*. However, no studies have shown whether these genes will play a role in the biosynthesis of peimisine and other alkaloids. Thus, we will continue to focus on whether this occurs and what roles the above enzyme genes may play in the synthesis of sterol alkaloids in subsequent studies of *F. taipaiensis*.

As shown before, relative steroid accumulation is highest in the third year and decreases in the fourth year. However, in actual production of medicinal materials, we must consider not only the content of active ingredients per gram but also the final yield of the medicine. Thus, biomass must be an important indicator. As shown in Figure 9, the four-year-old bulbs of *F. taipaiensis* are at least twice the size of three-year-old bulbs. Thus, the reason why the alkaloid content in the four-year-old bulbs was relatively low may be that the accumulation of secondary metabolites cannot keep up with the increasing bulb mass. For production, the larger mass and relatively high secondary metabolite accumulation are more suitable for harvesting. Based on the above results and discussion, we recommend harvesting *F. taipaiensis* in the fourth year.

Conclusions

Overall, the lack of a reference genome for *F. taipaiensis* has made it difficult to determine the exact number of genes involved in the metabolic pathways of this plant species. Here, we conducted Illumina sequencing to profile the transcriptome of *F. taipaiensis*. To the best of our knowledge, this is the first attempt at using Illumina PE sequencing technology to produce bulb and leaf transcriptomes of Fritillariae Cirrhosae Bulbus through *de novo* sequencing and assembly without a reference genome. A large number of candidate transcripts were matched with known enzymes in public databases, providing a substantial gene resource for further research in *F. taipaiensis*. Our study also provides reference sequences for evolutionary analyses of metabolomes among the members of the Fritillaria family, which is rich in ethnomedicinal plants.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

The datasets generated during the current study are available in the NCBI Sequence Read Archive database (The data is being uploaded).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China [grant number 81473303] and [grant number 81673552], and the CAMS Innovation Fund for Medical Sciences [grant number 2016-I2 M-3-016]

Authors' contributions

JH conceived the study and participated in its design. XYW, XMW and PC contributed samples and performed the experiments. XYW and ZG analyzed the data. XYW, JH and HL drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

References

1. Commission CP: Pharmacopoeia of the People's Republic of China, vol. Part I. Beijing: China Medical Science Press; 2015.

2. Guo S, Hu M, Zamora ML, Peng J, Shang D, Zheng J, Du Z, Wu Z, Shao M, Zeng L *et al*: Elucidating severe urban haze formation in China. *Proceedings of the National Academy of Sciences of the United States of America* 2014, 111(49):17373-17378.

3. Edita DFRPSAAS: Flora Reipublicae Popularis Sinicae, vol. Tomus 14. Beijing: Science Press; 1980.

4. Wang D, Chen X, Atanasov AG, Yi X, Wang S: Plant Resource Availability of Medicinal Fritillaria Species in Traditional Producing Regions in Qinghai-Tibet Plateau. *Frontiers in pharmacology* 2017, 8:502.

5. Duan B, Chen X, Huang L, Lu Q, Li X, Chen S: Resources study of Fritillaria taipaiensis. *Modern Chinese Medicine* 2010, 12(4):3.

6. Chan S-W, Li S-L, Lin G, Li P: Pharmacokinetic Study and Determination of Imperialine, the Major Bioactive Component in Antitussive Fritillaria cirrhosa, in Rat by High-Performance Liquid Chromatography Coupled with Evaporative Light-Scattering Detector. *Analytical Biochemistry* 2000, 285(1):172-175.

7. Li S-L, Li P, Lin G, Chan S-W, Ho Y-P: Simultaneous determination of seven major isosteroidal alkaloids in bulbs of Fritillaria by gas chromatography. *Journal of Chromatography A* 2000, 873(2):221-228.

8. Ruan X, Yang L, Cui WX, Zhang MX, Li ZH, Liu B, Wang Q: Optimization of Supercritical Fluid Extraction of Total Alkaloids, Peimisine, Peimine and Peiminine from the Bulb of Fritillaria thunbergii Miq, and Evaluation of Antioxidant Activities of the Extracts. *Materials* 2016, 9(7).

9. Liu J, Wang S, Xin G-z, Li H-j, Li P: Determination of Peimisine in Fritillaria taipaiensis and Fritillaria unibracteata by HPLC-ELSD. *Chinese Pharmaceutical Journal* 2010, 45(13):3.

10. Wang D, Wang S, Chen X, Xu X, Zhu J, Nie L, Long X: Antitussive, expectorant and anti-inflammatory activities of four alkaloids isolated from Bulbus of Fritillaria wabuensis. *Journal of ethnopharmacology* 2012, 139(1):189-193.

11. Wang D, Zhu J, Wang S, Wang X, Ou Y, Wei D, Li X: Antitussive, expectorant and anti-inflammatory alkaloids from Bulbus Fritillariae Cirrhosae. *Fitoterapia* 2011, 82(8):1290-1294.

12. Xu F, Xu S, Wang L, Chen C, Zhou X, Lu Y, Zhang H: Antinociceptive Efficacy of Verticinone in Murine Models of Inflammatory Pain and Paclitaxel Induced Neuropathic Pain. *Biological and Pharmaceutical Bulletin* 2011, 34(9):1377-1382.

13. Wang D, Yang J, Du Q, Li H, Wang S: The total alkaloid fraction of bulbs of Fritillaria cirrhosa displays anti-inflammatory activity and attenuates acute lung injury. *Journal of ethnopharmacology* 2016, 193:150-158.

14. Wu X, Chan SW, Ma J, Li P, Shaw PC, Lin G: Investigation of association of chemical profiles with the tracheobronchial relaxant activity of Chinese medicinal herb Beimu derived from various Fritillaria species. *Journal of ethnopharmacology* 2018, 210:39-46.

15. Xiao P-G, JIANG Y, LI P, LUO Y-B, LIU Y: The botanical origin and pharmacophylogenetic treatment of Chinese materia medica Beimu *Acta Phytotaxonomica Sinica* 2007, 45(4):473-487.

16. Wang L, Peng R, Li L: Research Advances in Fritillaria taipaiensis P. Y. Li. *Journal of Anhui Agri⍰ Sci⍰* 2011, 39(36):22309-22310, 22327.

17. Nomura T, Bishop GJ: Cytochrome P450s in plant steroid hormone synthesis and metabolism. *Phytochemistry Reviews* 2006, 5(2-3):421-432.

18. Ohnishi T, Yokota T, Mizutani M: Insights into the function and evolution of P450s in plant steroid metabolism. *Phytochemistry* 2009, 70(17-18):1918.

19. Unamba CI, Nag A, Sharma RK: Next Generation Sequencing Technologies: The Doorway to the Unexplored Genomics of Non-Model Plants. *Frontiers in plant science* 2015, 6:1074.

20. Bolger AM, Lohse M, Usadel B: Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014, 30(15):2114-2120.

21. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al*: Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011, 29(7):644-652.

22. Fu L, Niu B, Zhu Z, Wu S, Li W: CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 2012, 28(23):3150-3152.

23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. *Journal of Molecular Biology* 1990, 215(3):403-410.

24. Li B, Dewey CN: RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011.

25. Robinson MD, McCarthy DJ, Smyth GK: edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010, 26(1):139-140.

26. Livak KJ, Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001, 25(4):402-408.

27. Kanehisa M, Goto S: KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 2000, 28(1):27-30.

28. Mehrafarin A, Qaderi A, Rezazadeh S, Naghdi Badi H, Noormohammadi G, Zand E: Bioengineering of important secondary metabolites and metabolic pathways in fenugreek (Trigonella foenum-graecum L.). *Journal of Medicinal Plants* 2010, 9:1-18.

29. Vaidya K, Ghosh A, Kumar V, Chaudhary S, Srivastava N, Katudia K, Tiwari T, Chikara SK: De Novo Transcriptome Sequencing in L. to Identify Genes Involved in the Biosynthesis of Diosgenin. *Plant Genome* 2013, 6(2):1-11.

30. Naoumkina MA, Modolo LV, Huhman DV, Urbanczykwochniak E, Tang Y, Sumner LW, Dixon RA: Genomic and Coexpression Analyses Predict Multiple Genes Involved in Triterpene Saponin Biosynthesis in Medicago truncatula. *Plant Cell* 2010, 22(3):850-866.

31. Abe I: Enzymatic synthesis of cyclic triterpenes. *Natural Product Reports* 2007, 24(6):1311.

32. Mizutani M, Ohta D: Diversification of P450 Genes During Land Plant Evolution. *Annual Review of Plant Biology* 2010, 61(1):291-315.

33. Fujita M, Fujita Y, Noutoshi Y, Takahashi F, Narusaka Y, Yamaguchi-Shinozaki K, Shinozaki K: Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in

the stress signaling networks. *Current Opinion in Plant Biology* 2006, 9(4):436.

34. Patra B, Schluttenhofer C, Wu Y, Pattanaik S, Yuan L: Transcriptional regulation of secondary metabolite biosynthesis in plants. *Biochimica Et Biophysica Acta* 2013, 1829(11):1236-1247.

35. Sun C, Sun Y, Song J, Li C, Li X, Zhang X, Li Y, Hu S, Luo H, Zhu Y: Discovery of genes related to steroidal alkaloid biosynthesis in Fritillaria cirrhosa by generating and mining a dataset of expressed sequence tags (ESTs). *Journal of Medicinal Plants Research* 2011, 5(21):5307-5314.

36. Zeisel A, Muñozmanchado AB, Codeluppi S, Lönnerberg P, La MG, Juréus A, Marques S, Munguba H, He L, Betsholtz C: Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 2015, 347(6226):1138.

37. Pang T, Ye CY, Xia X, Yin W: De novosequencing and transcriptome analysis of the desert shrub,Ammopiptanthus mongolicus, during cold acclimation using Illumina/Solexa. *Bmc Genomics* 2013, 14(1):488.

38. Alves-Carvalho S, Aubert G, Carrère S, Cruaud C, Brochot AL, Jacquin F, Klein A, Martin C, Boucherot K, Kreplak J: Full-length de novo assembly of RNA-seq data in pea (Pisum sativum L.) provides a gene expression atlas and gives insights into root nodulation in this species. *Plant Journal for Cell & Molecular Biology* 2015, 84(1):1.

39. Björklund ÅK, Forkel M, Picelli S, Konya V, Theorell J, Friberg D, Sandberg R, Mjösberg J: The heterogeneity of human CD127+ innate lymphoid cells revealed by single-cell RNA sequencing. *Nature Immunology* 2016, 17(4):451.

40. Spyropoulou EA, Haring MA, Schuurink RC: RNA sequencing on Solanum lycopersicum trichomes identifies transcription factors that activate terpene synthase promoters. *Bmc Genomics* 2014, 15(1):402.

41. Wang LJ, Fang X, Yang CQ, Jianxu LI, Chen XY: Biosynthesis and Regulation of Secondary Terpenoid Metabolism in Plants. *Scientia Sinica* 2013, 43(12):1030-1046.

42. Newman JD, Chappell J: Isoprenoid Biosynthesis in Plants: Carbon Partitioning Within the Cytoplasmic Pathway. *Crc Critical Reviews in Biochemistry* 1999, 34(2):95.

43. Meier S, Tzfadia O, Vallabhaneni R, Gehring C, Wurtzel ET: A transcriptional analysis of carotenoid, chlorophyll and plastidial isoprenoid biosynthesis genes during development and osmotic stress responses in Arabidopsis thaliana. *BMC Syst Biol* 2011, 5:77.

44. Corey EJ, Matsuda SP, Bartel B: Isolation of an Arabidopsis thaliana gene encoding cycloartenol synthase by functional expression in a yeast mutant lacking lanosterol synthase by the use of a chromatographic screen. *Proceedings of the National Academy of Sciences of the United States of America* 1993, 90(24):11628-11632.

45. Zhao H, Tang Q, Mo C, Bai L, Tu D, Ma X: Cloning and characterization of squalene synthase and cycloartenol synthase from Siraitia grosvenorii. *Acta Pharmaceutica Sinica B* 2017, 7(2):215-222.

46. Palrasu M, Nagini S: Cytochrome P450 structure, function and clinical significance: A review. *Current Drug Targets* 1969, 18(999).

47. Nebert DW, Nelson DR, Adesnik M, Coon MJ, Estabrook RW, Gonzalez FJ, Guengerich FP, Gunsalus IC, Johnson EF, Kemper B: The P450 superfamily: updated listing of all genes and recommended nomenclature for the chromosomal loci. *Dna* 1989, 8(1):1-13.

48. Gotoh O: Evolution of cytochrome p450 genes from the viewpoint of genome informatics. *Biological & Pharmaceutical Bulletin* 2012, 35(6):812.

49. Fujita S, Ohnishi T, Watanabe B, Yokota T, Takatsuto S, Fujioka S, Yoshida S, Sakata K, Mizutani M: Arabidopsis CYP90B1 catalyses the early C-22 hydroxylation of C27, C28 and C29 sterols. *Plant J* 2006, 45(5):765-774.

50. O'Brien M, Chantha SC, Rahier A, Matton DP: Lipid signaling in plants. Cloning and expression analysis of the obtusifoliol 14alpha-demethylase from Solanum chacoense Bitt., a pollination- and fertilization-induced gene with both obtusifoliol and lanosterol demethylase activity. *Plant Physiology* 2005, 139(2):734-749.

51. Lepesheva GI, Waterman MR: Sterol 14alpha-demethylase cytochrome P450 (CYP51), a P450 in all biological kingdoms. *Biochim Biophys Acta* 2007, 1770(3):467-477.

52. Kim HB, Schaller H, Goh C-H, Kwon M, Choe S, An CS, Durst F, Feldmann KA, Feyereisen R: Arabidopsis cyp51 Mutant Shows Postembryonic Seedling Lethality Associated with Lack of Membrane Integrity. *Plant Physiology* 2005, 138:2033-2047.

53. Razdan S, Bhat WW, Dhar N, Rana S, Pandith SA, Wani TA, Vishwakarma R, Lattoo SK: Molecular characterization of DWF1 from Withania somnifera (L.) Dunal: its implications in withanolide biosynthesis. *Journal of Plant Biochemistry and Biotechnology* 2016, 26(1):52-63.

Tables

Table 1 Information for *Fritillaria taipaiensis* bulb and leaf samples

Sample name	Collected year	Growth age	Organ
2016-B1	2016	one year	bulbs
2016-B2	2016	two years	bulbs
2016-B3	2016	three years	bulbs
2016-B4	2016	four years	bulbs
2016-L1	2016	one year	leaves
2016-BN	2016	perennial	bulbs
2016-LN	2016	perennial	leaves
2017-B1	2017	one year	bulbs
2017-L1	2017	one year	leaves
2017-B2	2017	two years	bulbs
2017-L2	2017	two years	leaves
2017-B3	2017	three years	bulbs
2017-L3	2017	three years	leaves
2017-B4	2017	four years	bulbs
2017-L4	2017	four years	leaves
2017-BN	2017	perennial	bulbs
2017-LN	2017	perennial	leaves

Table 2 Gene-specific primers for qRT-PCR

No.	Gene name	Primer pairs (5'-3')	
		Forward primer	Reverse primer
1	FtActin	GGAGAAGCTTGCGTATGTGG	TAGGTGGTCTCATGGATGCC
2	ACAT	TGATGTTGTTGTGGCTGGTG	CATAACTCGGCACACACTCC
3	HMGS	GATTCGTGCTACAACCGCTT	TCCTCAATCGAACTGGCCTT
4	HMGR	CTGGATTGAAGGGAGAGGCA	CATGGTGATGCAGTGAGAGC
5	MVK	TTTTGCTGCTGTTGACGTCA	AACACATCCTCCTCCACCAG
6	PMVK	CCGTAGTTGCAGCTCTCCTA	GACAGCAGCACACACATCAA
7	MVD	TCATTTCCCGACCGCCTATT	GGATTAGGGTTTCGTCACGC
8	DXS	TCAACTGGCCCCGTTCTTAT	ACCTGTACCACCTCCCATTG
9	DXR	TCGTTACGGGGATAGTTGGG	ATGTTCTGAATCGGCTGGGA
10	ispD	ACAACCCCTAATCTCTCGCC	GTGCGATTCTGATTCACCCC
11	ispE	GCTTTGTGGGCTGCTAATCA	TTGGACATGCTTCTTGTGGC
12	ispF	TCGGTTTTCATGAAGGAGGC	CTCTCCGAGACTGTCAACCT
13	ispG	AGTATGTCTCGTGCCCATCC	CCACCGACATACCCAAAATC
14	ispH	AGTATTCCGCCACAACCTCA	AACTGCTTCCTCGCCTCATA
15	IDI	TGATGTACCGGTTGCTGAGT	CAGCCTTCCTCACCAGTTCT
16	GGPS	ACGGCCGTAATCTGGGTTTA	AAACCACGCTCGACAACTTC
17	SQLE	AAAACAAAGCAGGCGAGGAG	GGCATGGAAGTCGACAGTTC
18	FDPS	CCGCGGATTCTTAAACGACA	GGCCAAGTTTTCACCAGACA
19	FD	CGACGTGGAGTTGTTGCTAG	GACTGCTGCGATACATGTGG
20	CAS	ATCTTCTGGTGGTTGGGGAG	CCCGCTATTTCTTGTTGGGG
21	CYP90B1	TTCAACACATGGGAGGACGA	GAGGTCGATAGGGATGGCAA
22	CYP51G1	TCACCCTCCGCTAATAATGC	GATCGTATGTGTCCGGGTCT
23	DWF1	CGAGCAACATAACCGACCAG	ATCGAACATGCGCCAAAAGT
24	FK	TCGTTCAGTTTGCCTTGTGG	CACGGAACAAGCTTGCAGTA
25	DWF5	CACTTTCCATACTCGCAGCC	TTGTTTCGCCTTTCTCGGTC
26	CYP90D2	TCGTTGTCATTCACCCAAGA	GAAGTCCCCGTCTCCTTACC
27	ACS12	TCCCGGATTCAGAGTTGGAG	CCGAAGCCTCTCCCTGTTTA
28	CYP734A6	ATGACTATCGTACAGGGCCG	AAGCAAATGGGAGGTGGTCT

Table 3 Peimisine content in bulbs of different ages

Sample name	parallel sample 1	parallel sample 2	parallel sample 3	Average	standard deviation
2016-B1	0.0456%	0.0445%	0.0443%	0.0448%	6.6704E-06
2016-B2	0.0632%	0.0633%	0.0615%	0.0626%	1.0131E-05
2016-B3	0.0651%	0.0657%	0.0673%	0.0660%	1.1572E-05
2016-B4	0.0339%	0.0334%	0.0330%	0.0334%	4.4820E-06
2016-BN	0.0228%	0.0229%	0.0232%	0.0230%	1.9286E-06
2017-B1	0.0521%	0.0522%	0.0509%	0.0517%	7.4971E-06
2017-B2	0.0563%	0.0559%	0.0558%	0.0560%	2.9405E-06
2017-B3	0.0630%	0.0642%	0.0634%	0.0635%	6.0700E-06
2017-B4	0.0341%	0.0346%	0.0347%	0.0345%	3.6540E-06
2017-BN	0.0304%	0.0308%	0.0305%	0.0306%	1.9027E-06

Figures

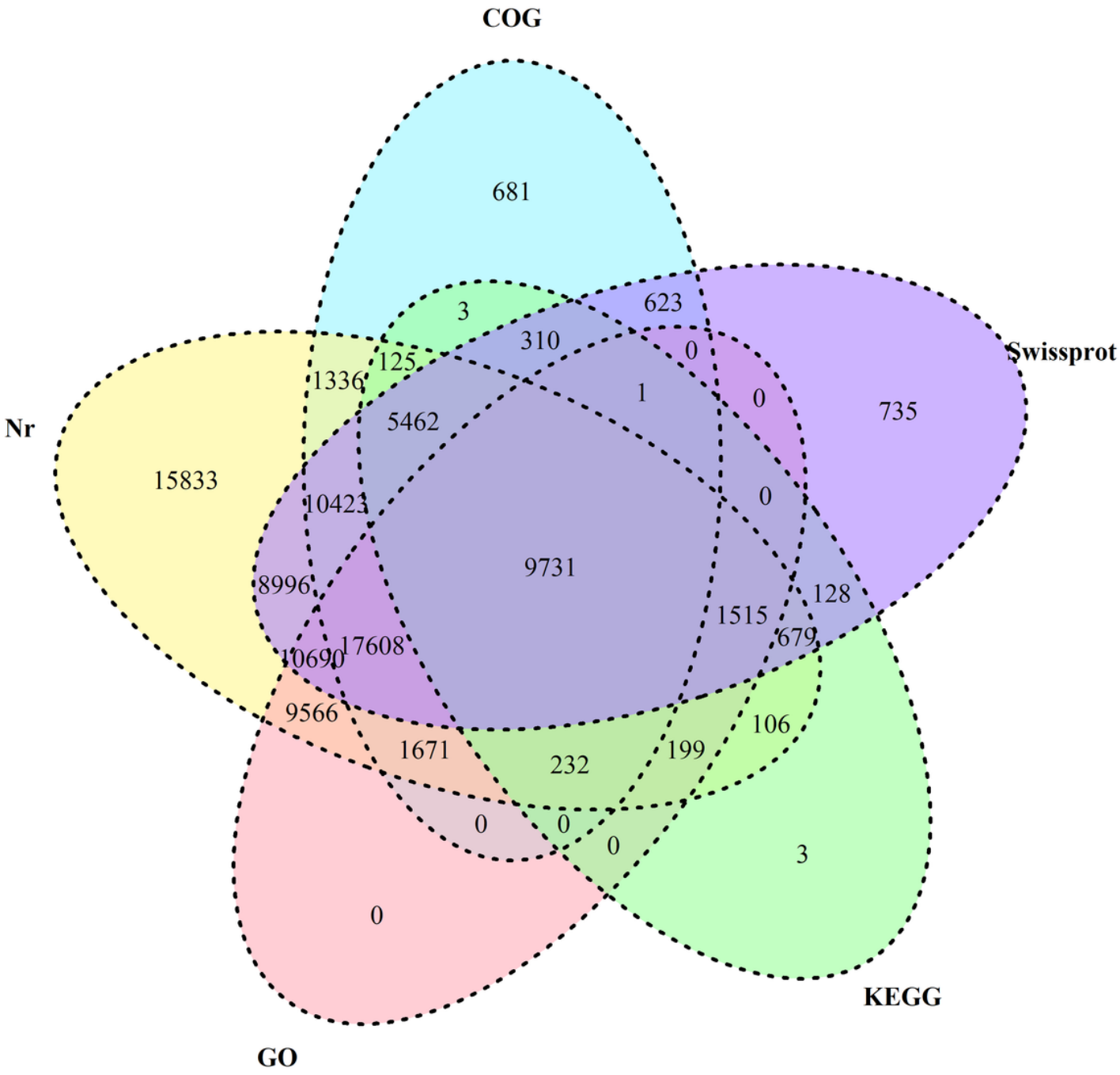


Figure 1

Unigene annotation results from five different databases

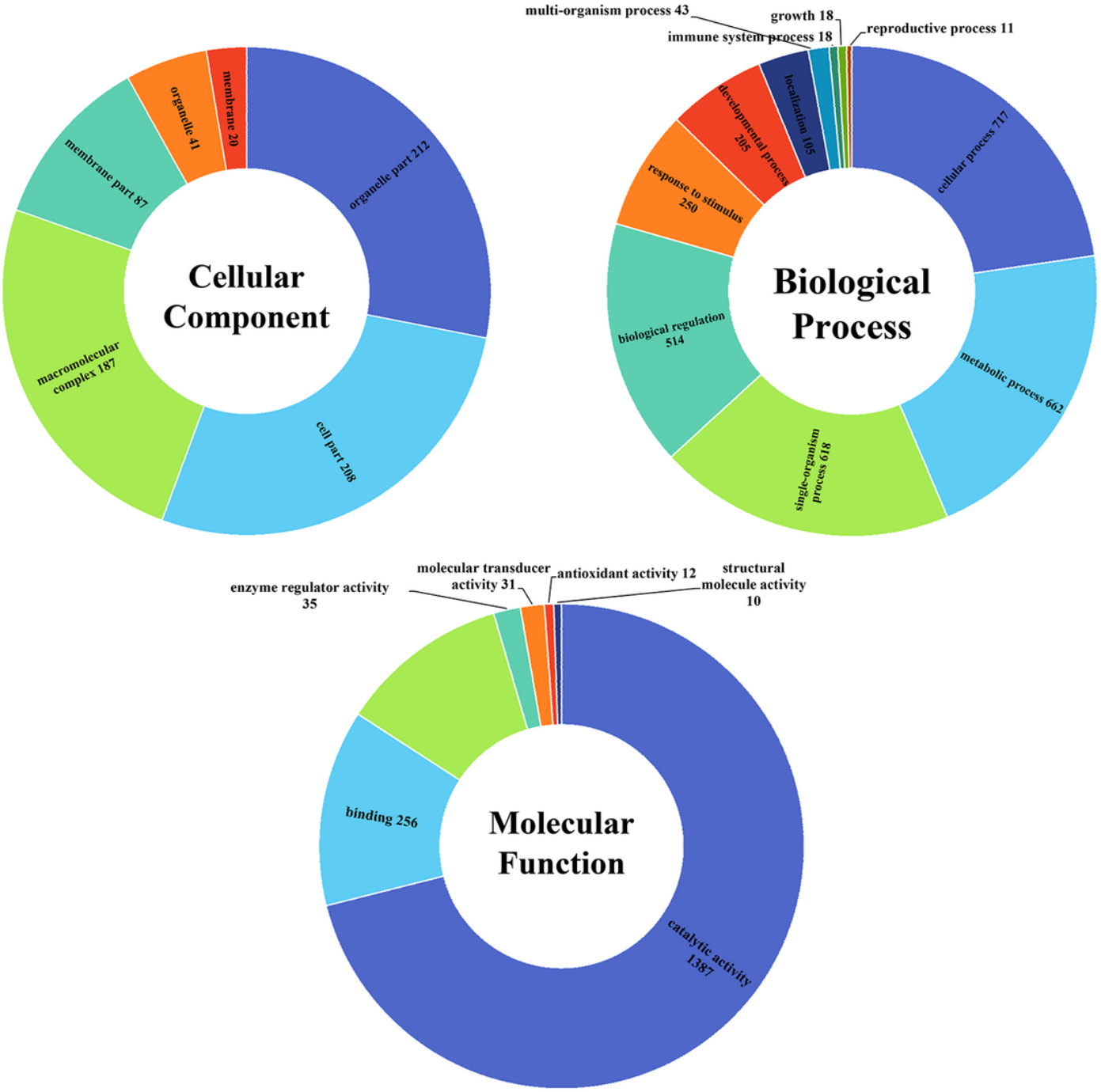


Figure 2

GO term enrichment results within three ontologies (Level 2, Num≥10)

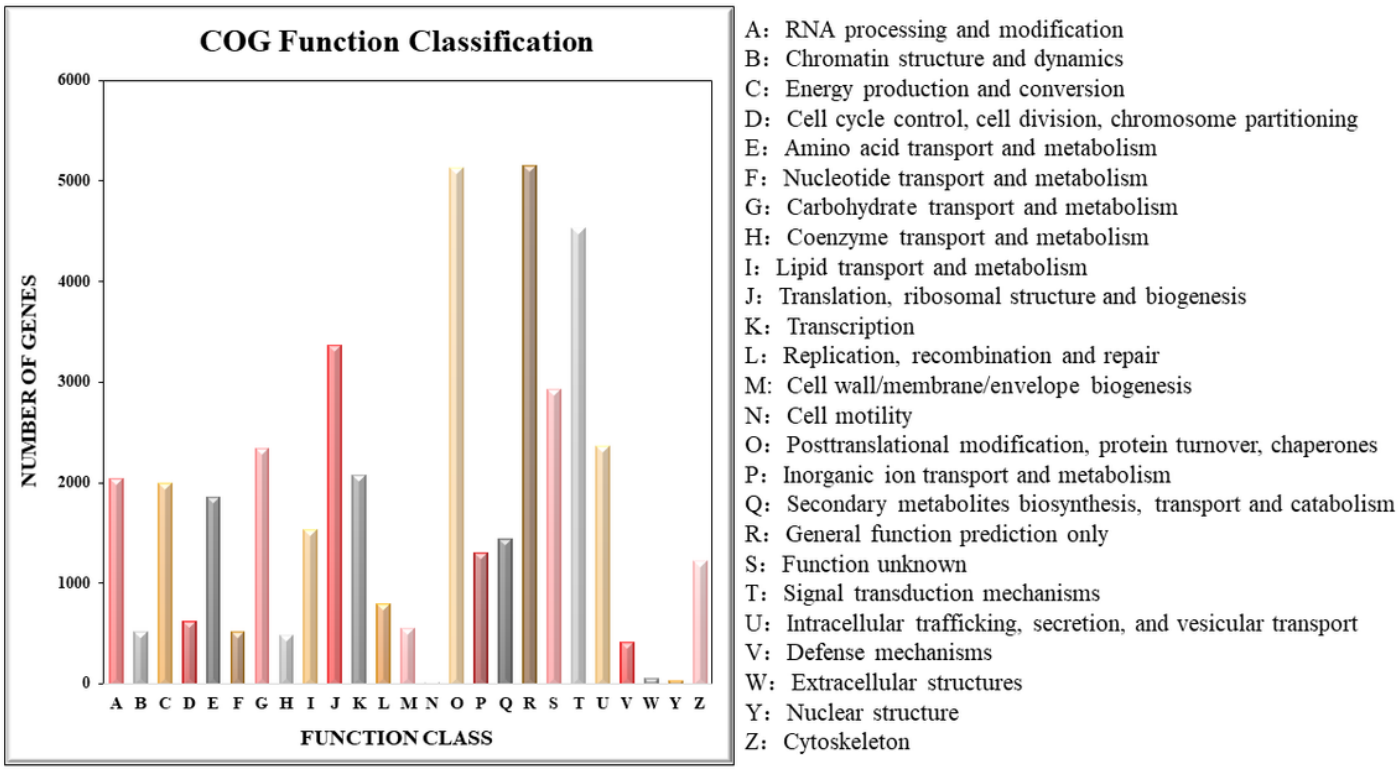


Figure 3

COG functional classification results of transcriptome data

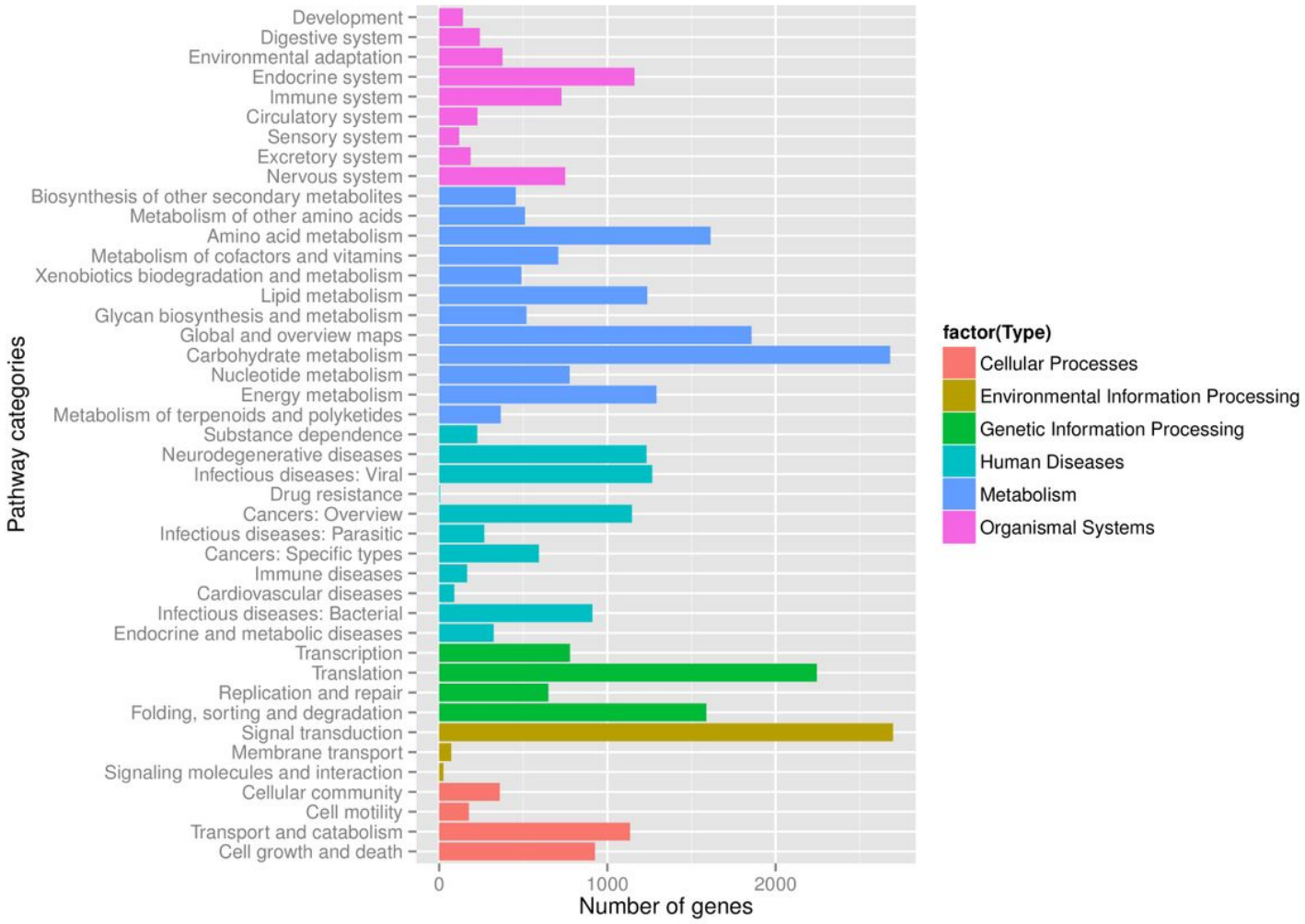
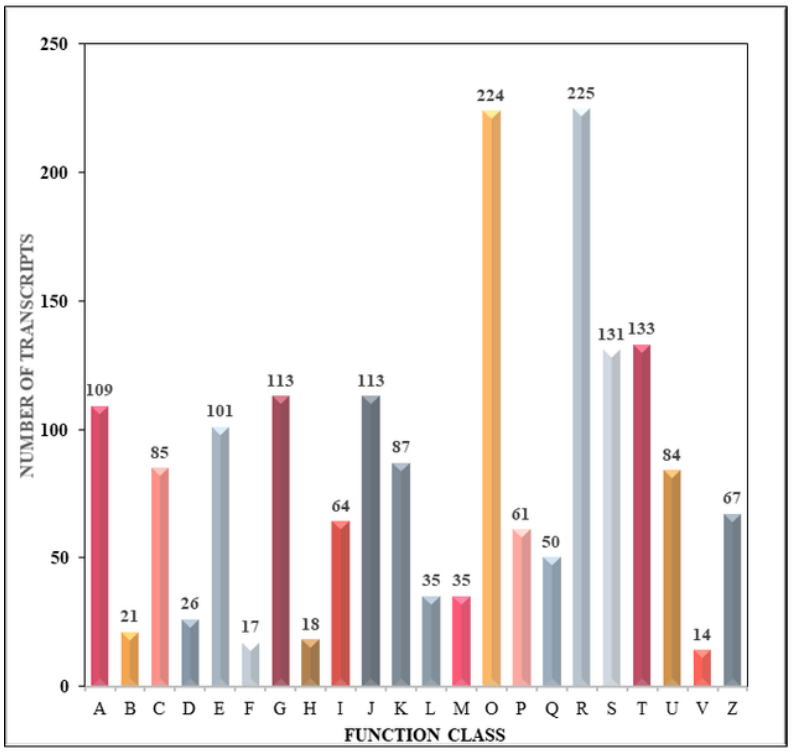


Figure 4

KEGG pathway gene categories



A: RNA processing and modification
B: Chromatin structure and dynamics
C: Energy production and conversion
D: Cell cycle control, cell division, 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
O: Posttranslational modification, protein turnover, 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
Z: Cytoskeleton

Figure 5

COG functional classification results for differentially expressed genes

Figure 6

Relative quantitative results from qRT-PCR of 19 genes related to the steroidal alkaloid biosynthesis pathway

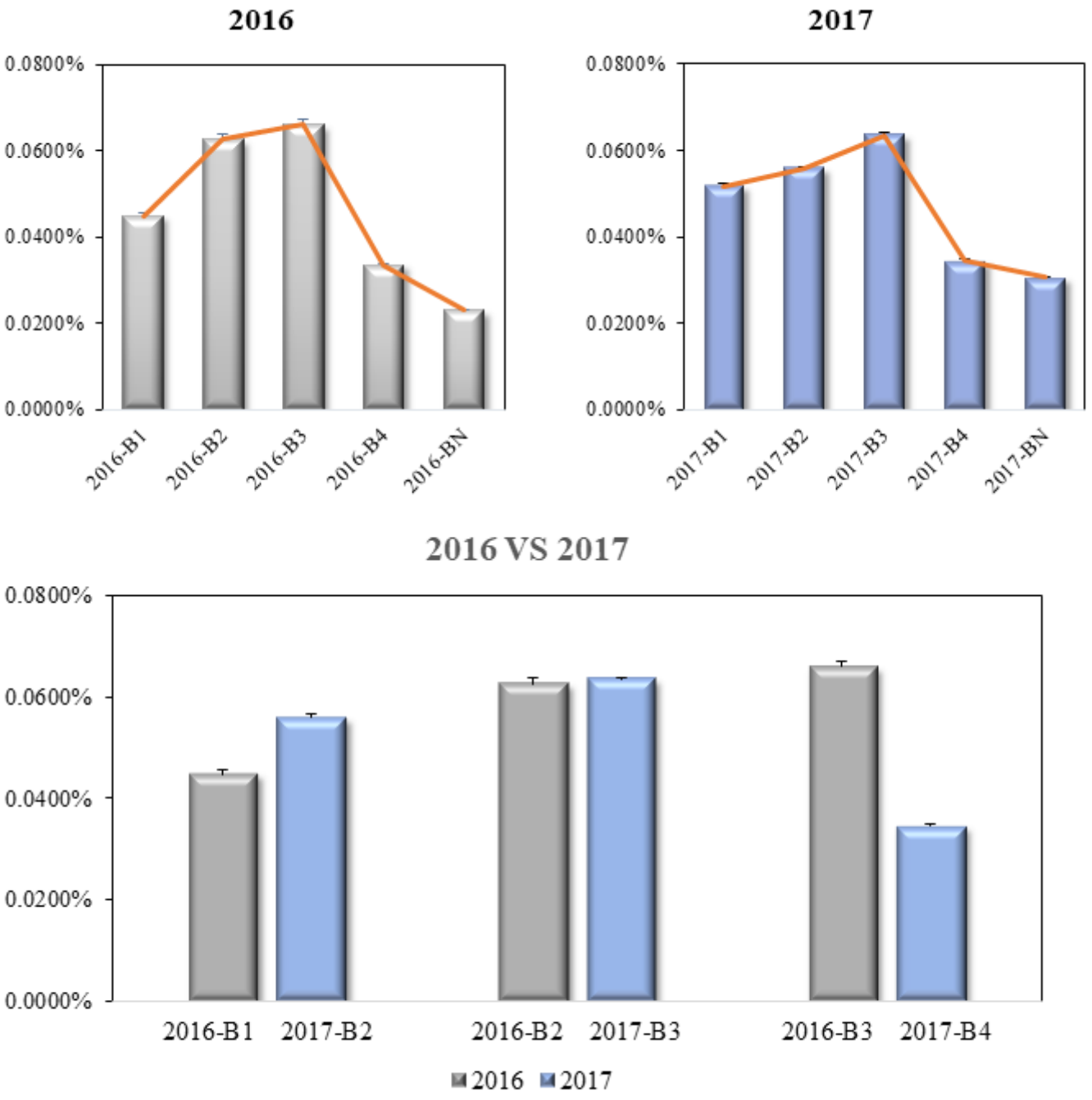


Figure 7

Comparison of content determination among bulbs of different ages

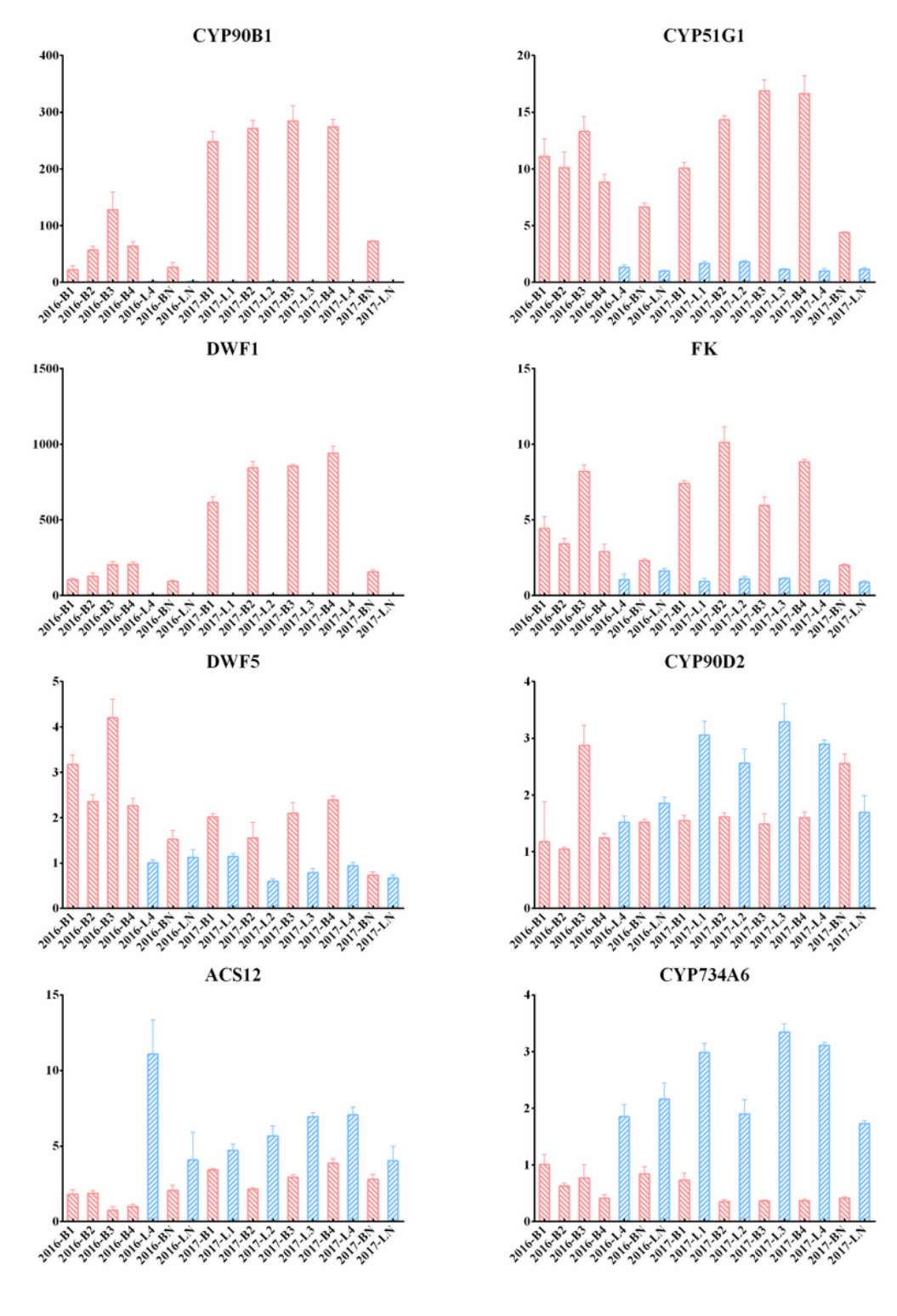


Figure 8

Relative quantitative results from qRT-PCR of 8 oxidoreductase genes

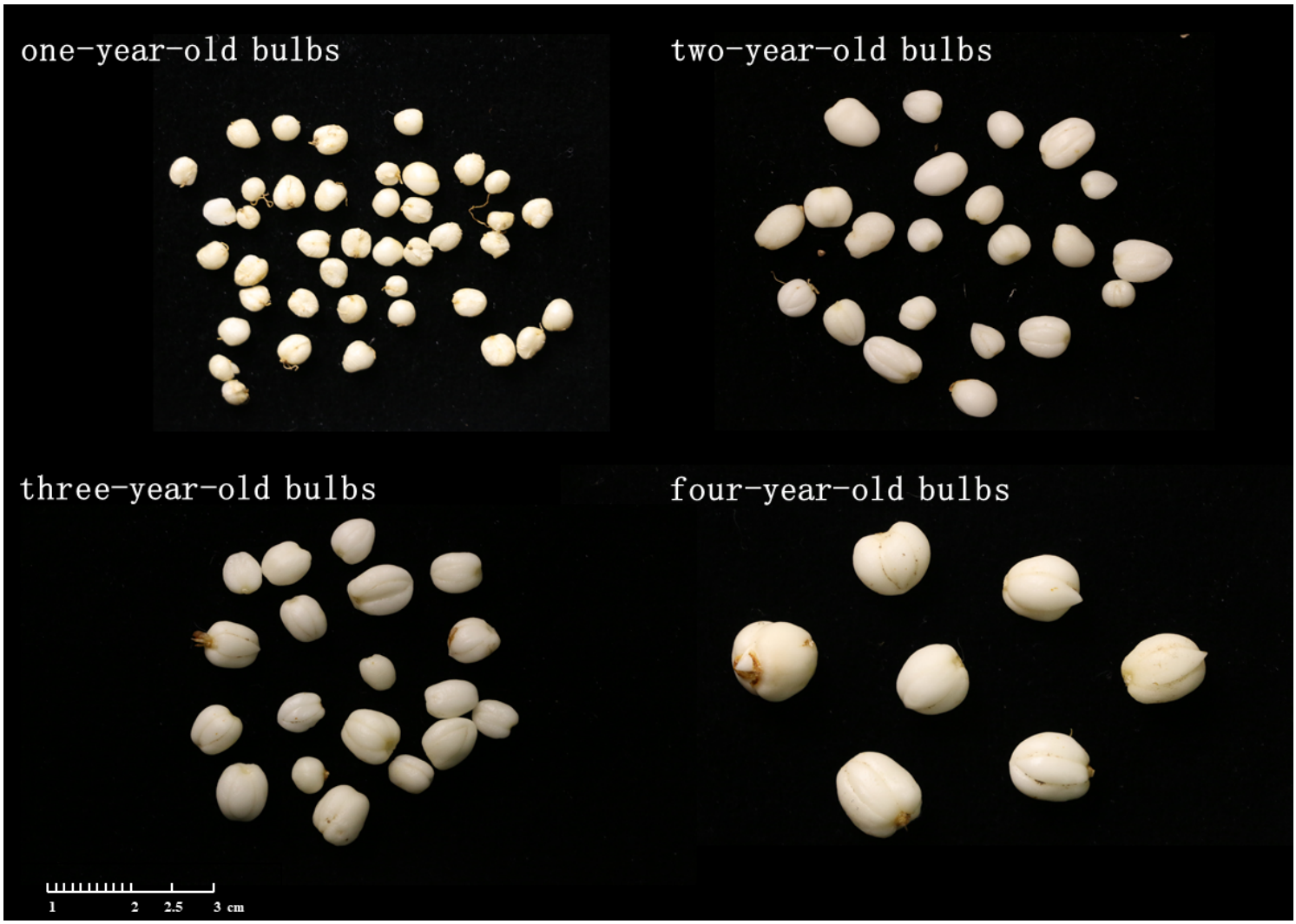


Figure 9

Size comparison of bulbs of different ages