Genomics 111 (2019) 1360–1372

Contents lists available at ScienceDirect

Genomics

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

Orginal Article

Identification of genes involved in steroid alkaloid biosynthesis in *Fritillaria imperialis* via de novo transcriptomics

Mahsa Eshaghi[a], Behrouz Shiran[a,b,*], Hossein Fallahi[c], Rudabeh Ravash[a], Bojana Banović Đeri[d]

[a] Department of Plant Breeding and Biotechnology, Faculty of Agriculture, Shahrekord University, P.O. Box 115, Shahrekord, Iran
[b] Institute of Biotechnology, Shahrekord University, P.O. Box 115, Shahrekord, Iran
[c] Department of Biology, School of Sciences, Razi University, Bagh-e-Abrisham Kermanshah, Iran
[d] Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Vojvode Stepe 444a, P.O. Box 23, Belgrade 11010, Serbia

ARTICLE INFO

Keywords:
Fritillaria imperialis
Steroidal alkaloids
de novo transcriptomics
Expression profiles

ABSTRACT

Crown imperial (CI) has been used in traditional medicine. Today it is known that such beneficial effects are due to its richness in steroidal alkaloids (SA). Using de novo transcriptomics, orthologues/paralogues finder, phylogenetic analysis and tissue- and developmental stage-specific expression analysis, we identified ten genes and several TFs involved in the biosynthesis of SA in CI. The comparative analysis of ten genes expression profiles revealed the possibility of their co-regulation, which may imply the possibility of their organization in metabolic gene clusters. Having in mind convergent evolution of steroidal biosynthetic pathways in flowering plants and records of convergent evolution of specific proteins, observed expression patterns open a reasonable interest to investigate the possibility of the existence of genes cluster organization in SA pathway in the family Liliaceae or at least in some species of genus *Fritillaria*. Obtained results support transcriptomics as useful approach in elucidating genes underlying complex biochemical pathways.

1. Introduction

Plant steroidal alkaloids are unique natural products of great chemical diversity, interesting to study because of their wide pharmaceutical range of activities and vast bioengineering potential [1]. Many studies devoted to identification of genes and enzymes involved in biosynthesis of steroidal alkaloids in various plants led to finding that their production includes the main route of triterpene or steroid biosynthesis, starting with acetyl coenzyme A via mevalonic acid, isopentenyl pyrophosphate, farnesyl pyrophosphate, squalene, cycloartenol, and cholesterol as intermediates [1,2]. Further conversion in the steroidal alkaloids is carried out by two large multigene families, cytochromes P450 (CYP450), responsible for oxidation steps (like hydroxylations, dealkylation, epoxidation, dehydration, carbon-carbon bond cleavage) and glycosyltransferases (GTFs) responsible for intermolecular transfer of sugar groups [3–5]. Multiple steroidal biosynthetic pathways involving similar genes are present in monocots (i.e. Liliaceae, Melanthiaceae, Poaceae) and dicots (i.e. Apocynaceae, Buxaceae, Solanaceae), but they seem to be of independent origin, at least in the case of solanidine- and verazine-like alkaloids, suggesting their convergent evolution [6,7,11]. Moreover, steroidal alkaloid genes are found to be organized in clusters in the Solanaceae family [8,10,56], but so far there are no records on the presence of similar clusters in other families. Biochemical analysis accompanied by gene expression studies supported these genes to be involved in the steroidal alkaloid biosynthesis pathway [9–11,57–59]. Interestingly, such biosynthesis is shown to be restricted to specific cells/tissues and plant developmental stages in order to enable careful orchestration of their production with the main homeostasis processes in plant [3]. Therefore the content of different steroidal alkaloids varies in different plant tissues, some of which have been found to be the highest in root, bulb and rhizome in *Veratrum californicum*, tubers in potato ad fruit in tomato,

Abbreviations: GO, Gene ontology; CYP450, Cytochrome P450; RNA-Seq, RNA sequencing; NGS, Next generation sequencing; RIN, RNA Integrity Number; TPM, Transcript per million; RSEM, RNA-Seq by Expectation Maximization; FPKM, Fragments per kilo base per million mapped reads; KEGG, Kyoto encyclopedia of genes and genomes; EC, Enzyme commission; GAPDH, Glyceraldehyde-3-phosphate dehydrogenase; MYB, myeloblastosis protein; MADS, MCM1, AG, DEFA and SRF; bHLH, basic Helix-Loop-Helix; bZIP, basicleucine Zipper; PSPG, Plant secondary product glycosyltransferase; UGTs, UDP-glycosyltransferases; HMG, Hydroxymethylglutaryl-CoA reductase; MK, Mevalonate kinase; MVD, Diphosphomevalonate decarboxylase; IPP, Isopentenyl diphosphate isomerase; FPS, Farnesyl diphosphate synthase; SQS, Squalene synthase; CAS, Cycloartenol Synthase; SMT, 24-methylenesterol *C*-methyltransferase; CYP51, Sterol 14-demethylase; nt, Nucleotide; bp, Base pair; TFs, Transcription factors; TEs, Transposable Elements; OMA, Orthologous Matrix; CI, Crown Imperial; SA, Steroidal Alkaloids

* Corresponding author at: Department of Plant Breeding and Biotechnology, Faculty of Agriculture, Shahrekord University, P.O. Box 115, Shahrekord, Iran.
E-mail address: beshiran45@gmail.com (B. Shiran).

https://doi.org/10.1016/j.ygeno.2018.09.008
Received 11 May 2018; Received in revised form 15 August 2018; Accepted 14 September 2018
Available online 21 September 2018

while others have been found to be the highest in green plant tissues [9,11,12].

Fritillaria is one of the largest genera of the Liliaceae family that includes ca. 140 species recognized for their horticultural and medical importance [13]. One of its species, *Fritillaria imperialis* (crown imperial), has been used for treating many illnesses such as a cough, sore throat, asthma, bronchitis, scrofula, gland tumor, dysuria and hemoptysis in traditional medicine [14]. Although this species is known for numerous secondary metabolites like persicanidine and imperialine with anti-acetylcholinesterase and anti-butyrylcholinesterase inhibitory effects [7], the genes involved in its steroidal alkaloid biosynthesis pathway have not yet been identified. On the other hand, studies in two other species, one representing a close relative of crowns imperial *Fritillaria cirrhosa* D. Don [15] of the Liliaceae family, and, the other representing more distant relative of crown imperial *Veratrum californicum* [11] of the Melanthiaceae family, resulted in identifying genes encoding for 3-hydroxy-3-methylglutaryl-CoA reductase (HMG), farnesyl pyrophosphate synthetases (FPSs), aminotransferases, γ-aminobutyrate transaminase and numerous CYP450 to be involved in the production of steroidal alkaloids. However, all processes in plant alkaloid biosynthesis pathway would not be possible without corresponding transcription factors, which are identified to belong mostly to WRKY, ERF, AP2 and zinc finger families [16–18].

Since transcriptome approach was verified to be effective in discovering genes involved in various metabolic pathways in different plant species like carnation - *Dianthus caryophyllus L.* [19], *Siraitia grosvenorii* [20], *Bupleurum Chinese* [21], *Chlorophytum borivilianum* [22], *Senna* and *Cassia angustifolia* [23], we decided to use a de novo transcriptome approach to assess the roles of candidate genes of the steroidal alkaloid biosynthetic pathway in crown imperial. The workflow of the conducted study is summarized in Fig. 1. Knowing that *F. imperialis* is native to southern Iran, Turkey and Cyprus [24], but that its populations number has been continuously reducing due to improper grazing, pasture degradation and pest attacks [25,26], the biotechnological and molecular studies will become important tool for improving sustainable production and utilization of crown imperial, with its rich bioactive content.

The enzymes/transcripts identified in our research could be targeted to be applied in bioengineering, biocatalysis and in developing novel biocatalysts. Furthermore, having in mind the current trends of transferring the deduced biosynthetic pathways to microorganisms and/or easily transformable plants as well as chemical synthesis of valuable natural products [27,28], the potential of applications of this kind of findings is vastly increasing.

2. Material and methods

2.1. Plant material collection and preparation

Fritillaria imperialis plants were growing at the Koohrang, Central region Zagros (50° 13′ 49/4″(E), 32°34′ 06/8″(N)), Chaharmahalva Bakhtiari province, Iran, at the altitude 2563 m above sea level, with an average annual precipitation of 1414 mm. The sampling distance between plants was 100–500 m and two biological replicas per sample were used. Each sample contained three to four plants pooled together for each tissue.

Six tissues (leaf, bulb, anther, ovary, sepal and petal) were harvested at two developmental stages (stem elongation, flower development) and two tissues (bulb and fruit) were collected as seed head stage (Fig. 2) in April 2015 and in March–May 2016. The collected samples were immediately frozen in liquid nitrogen and stored at −80 °C until processing.

The stem elongation stage was defined as the phase in which leaves were in a rosette and the maximum length of the plant was up to 60 mm. The flower development phase was characterized by bright red flower, umbel inflorescence and maximum length of 32.5 mm for pedicle. Finally, the seed head stage was distinguished by oblong-cylindrical capsule fruit with the maximum width of 28.5 mm.

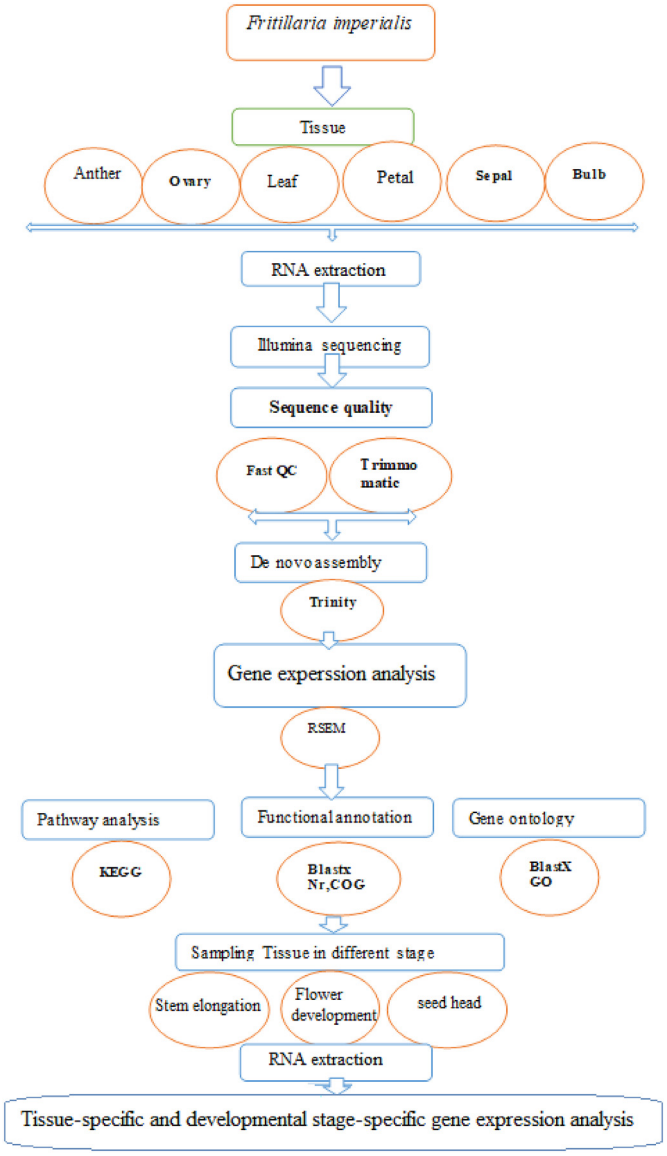

Fig. 1. Schematic overview of the RNA-seq analysis workflow in *Fritillaria imperialis*. Pooled RNA from different tissues in plant flowering stage was used for cDNA library construction and sequenced by Illumina's HiSeq platform. Upon de novo assembly, annotation and other analysis of transcriptome data, the expression of ten genes identified as possibly associated with steroidal alkaloid biosynthesis were assessed with qPCR in seven tissues (leaf, ovary, sepal, petal, anthers, bulb and fruit) and three developmental stages (stem elongation, flower development and seed head stage).

2.2. RNA isolation

Total RNA was isolated from 50 to 100 mg of plant tissue using TRIzol reagent [29] and dissolved in RNase-free water. The purity and quality of isolated RNA were assessed using a spectrophotometer and agarose gel electrophoresis. Equal amounts of RNA samples isolated from different tissues during flower development stage in the first year (2015) were pooled together and the integrity of pooled RNA sample was tested prior to sequencing for RIN (RNA Integrity Number) value to be equal or > 8 (Agilent 2100 Bioanalyzer) [30].

Fig. 2. Side view of three developmental stages of *Fritillaria imperialis* used in this study: (A) Stem elongation stage, (B) Flower development stage and (C) Seed head stage.

Table 1
The list of gene-specific primers used in quantitative Real Time PCR expression analysis. Following the primer-gene name list contains specified primer's sequence with corresponding annealing temperature and the length of expected PCR product.

Putative gene name	Primer sequences 5′- > 3′	Annealing (°C)	Amplicon (bp)
Fim-HMG	F:CACCAACAAGCTCTTCCTCG	58.85	198
	R:GAGGAGACGAAGGACTGGAC	58.85	
Fim-MK	F:AACACATCCTCCTCCACCAG	59.01	191
	R:TCGAGTCACCTGCTTTGGAT	59.02	
Fim-MVD	F:TCCTCCCTATCAACGACAGC	58.89	171
	R:GGTTCCGGATCTCCCTCAAA	59.09	
Fim-IPP	F:TGCTCCATAGGGCTTTCAGT	58.71	191
	R:CAACTTCCTTTGGGCAGCAT	59.03	
Fim-FPS	F:CTGGAACGTGCAAATGAGGG	59.48	196
	R:CTTCCTGGACTGCTTTGCTC	58.84	
Fim-SQS	F:AGGGTCATGGACCAAACAAA	60.2	177
	R:GCATCCCCTTTTATTCAGCA	60.0	
Fim-CAS	F:TCCTGGTATGGATCTTGGGC	58.86	160
	R:CTCCCCAACCACCAGAAGAT	59.00	
Fim-CYP51-G1	F:CGCATGCCGTTGTCTAGATC	59.14	173
	R:GAGGCGGAGGACTACTTCTC	58.97	
Fim-SMT	F:CCTGAATCACCTCCACATGC	58.61	171
	R:GCTTCGACGGTGTTTACTCC	58.93	
Fim-UGT	F:GATAGGAGGCTTGGGGAAGG	59.23	161
	R:AAGACCACCATCCTCACCAC	59.31	
Fim-GAPDH	F:GTGGTCCAGGGGTCTTACTC	59.1	172
	R:CTGACTTCAACAGCGACACC	59.14	

2.3. RNA seq

The cDNA library was created by using ® TruSeq™ RNA Sample Preparation Kit according to manufacturers' instructions. Constructed paired-end reads, with an average size of 101 bp, were sequenced by Illumina HiSeq 2000 platform (Macrogen Company, Korea). The raw transcriptome data were submitted to NCBI sequence reads archive (SRA) under the accession number SRR6242026 (SRA website).

2.4. Bioinformatics analysis of transcriptome sequencing results and de novo assembly

Sequence quality was checked by using FastQC and Trimmomatic tool (Version 0.32) [31]. Low-quality reads, primer and adapter sequences were filtered out and only the reads with the quality score higher than 20 (Q > 20) (last two base pairs from each read were trimmed so as to diminish the sequencing error, which is usually greater in the 3′ end of reads) were used for de novo assembly in Trinity v1.3 (k-mer size was set to 25using a PC running Linux OS (Ubuntu 16.04.4 LTS) [32]. To form longer contiguous sequences (contigs), the cleaned pair-end reads with mutual overlapping were combined together into contigs. Finally, contigs were additional assembled and the ones that could not be extended on either end were determined as unique transcripts. Furthermore, the unique assembled transcripts were further subjected to the process of sequence splicing redundancy removal using sequence clustering software Trinity v1.3 to obtain non-redundant transcripts named unigenes.

The downstream data analysis, which included the quantification of transcript abundances in RNA-Seq data (estimating transcripts abundance, fraction of transcripts in transcript per million (TPM) and fragments per kilo base per million mapped reads (FPKM)) was conducted by using RSEM software tool (Version1.2.31). RSEM calculate the fraction of transcripts as the value between 0 and 1 that can be multiplied by million to acquire a measure TPM [33], which is more comparable across samples and species than FPKM. STAR2.6 aligner was used within RSEM as an expression calculator [34]. Transcripts' GC content was calculated by using Perl script available in EMBOSS program infoseq version 6.6.0.0.

2.5. Assignment of Gene Ontology (GO)

Unique sequences were assessed for gene ontology (GO) terms at the protein level by using basic local alignment search tool X algorithm (BLASTX version20140820) against the public protein databases (non-redundant, UniProt). De novo transcripts were annotated using BLASTX with a cut-off *E*-value of 10^{-5} [35]. Functional categorization, including biological processes, molecular functions and cellular components, was determined by GO terms available via http://www.geneontology.org.

2.6. Transcription factor families investigation

For the identification of transcription factor families represented in *Fritillaria imperialis* transcriptome, BLASTX with cut-off E-value of 10^{-5} was used to search against the known plant transcription factor database (http://plntfdb.bio.unipotsdam.de/v3.0/blastform.php).

2.7. Identifications of metabolic pathway enzymes

Identification of enzymes involved in metabolic pathways was carried out by exploring public pathway-related database, including Kyoto Encyclopedia of Genes and Genomes (KEGG), using BLASTX [36]. KAAS (KEGG automatic annotation server Ver. 2.1) was used to automatically determine functional properties and biological roles for sequences with BLASTX scores above the threshold value of 60 (Default) (http://www.genome.jp/tools/kaas/) [37]. For identifying cytochrome P450 (CYP450) in the annotated data we used intramural data mining tool and search in P450 database (http://drnelson.uthsc.edu/cytochromeP450.html) to identify the family number [38]. Also, to obtain the gathering of glycosyltransferases in the annotated data we assigned consensus PSPG motif sequence by BLASTX tool and run it

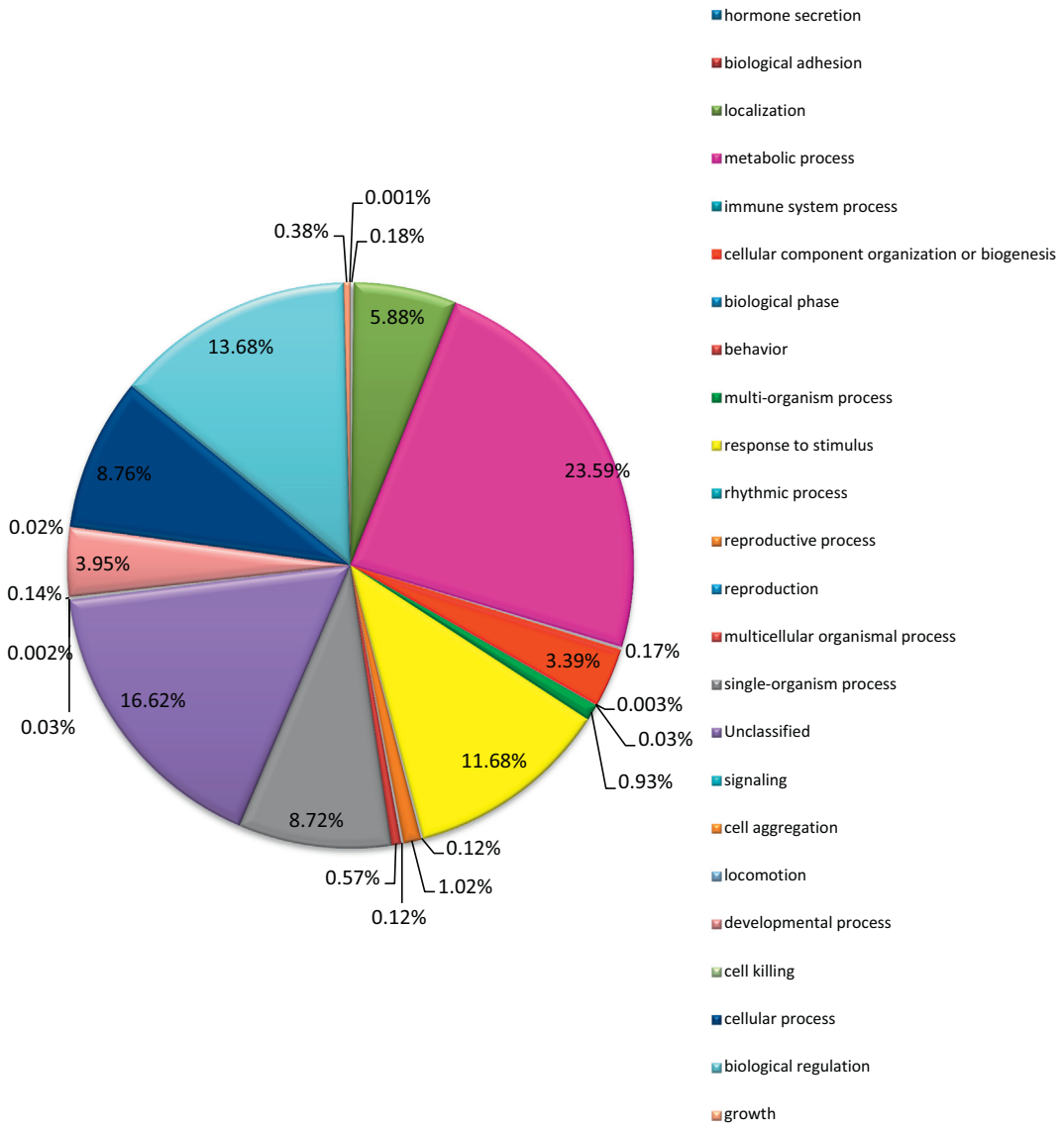

Fig. 3. Gene ontology (GO) classification of *Fritillaria imperialis* transcriptome data presented by percentages of contigs annotated to three main GO categories: (A) biological process, (B) cellular component, (C) molecular function.

against the glycosyltransferase database [39].

2.8. Transposons analysis

Transposons in *F. imperialis* transcriptome were detected using BLASTN search against Plants Transposable Elements Database v1.0 (http://genedenovoweb.ticp.net: 81/DPTEdb/) [40].

2.9. qPCR analysis and verification of primers specificity

Primers listed in Table 1 were designed by using Primer 3 online software (http://bioinfo.ut.ee/primer3-0.4.0/) [41] on the alignement of the crown imperial transcriptome data with conserved region using ClustalW v1.83 with default threshold settings and checked in NCBI Primer-BLAST tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) [42]. The primers were designed under the following settings: the optimal length of PCR primers 18–22 bp, Primer Melting Temperature (Tm) within the range of 58–60 °C, The content of GC about 45–55%, the target length closer to 100–200 bp and for standard PCR two primers of a primer pair had closely matched melting temperatures for maximizing PCR product yield. Designed primers specificity was further examined in silico using Primer BLAST (Last modified: Tue Jul 27 14:39:17 EDT 2010, default parameters against flowering plants (and verified as highly similar to the sequences contained in the same ten genes of other plant species. Finally, all primers were tested by standard PCR and 3% agarose gel electrophoresis, verifying primers' specificity for particular genes within gene families (Fig. S1).

Real-time PCR analysis was performed using glyceraldehyde-3-phosphate dehydrogenase (GADPH) as the internal reference gene (because it is stably expressed throughout different samples and tissues, [43] and leaf stage 2 as tissue calibrator. The relative gene expression was calculated by using the comparative quantitation method (Rotor-Gene software, Corbett) [44]. Three technical replicates and two biological replicates were used for each sample. Each sample contained three to four plants pooled together for each tissue.

Approximately 2 µg of DNase I-treated total RNA was subjected to cDNA synthesis (Thermos reverse transcriptase). The PCR reactions were performed on a Rotor-Gene Q Real-Time PCR Detection System (Qiagen). The reaction mixture (12 µl) contained 2× Power SYBR Green PCR Master mix, 1 µl of template cDNA and 1 µM of each primer (forward and reverse). The amplification program was set as following: 95 °C for 30 s, 95 °C for 5 s 60 °C for 1 min and repeated for 40 cycles.

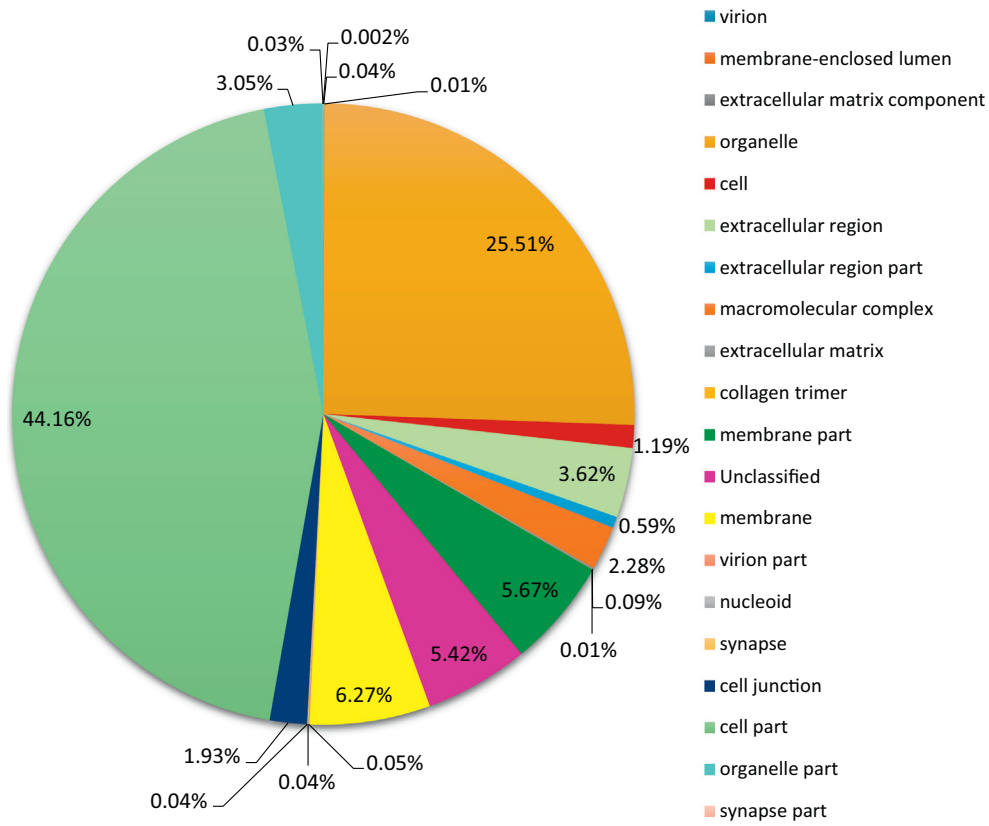

Fig. 3. (*continued*)

Melting curve for products was obtained by running a cycle of 62–95 °C for 5 s.

2.10. Phylogenetic analysis and orthologous/paralogous gene search

To discern phylogenetic relationships between molecular sequences of ten crown imperial candidate genes BLAST search against other flowering plants and Phylogeny.fr v3.1/3.0 with default threshold parameters settings (http://www.phylogeny.fr) [45] were used. Also, ClustalW v1.83 with default threshold settings and Pairwise Alignment algorithm was used for genes alignments.

For the identification of clusters of putative orthologues and paralogues with complete protein sequence the updated 2.0 version of the OMA algorithm was used (https://omabrowser.org/oma/home/) [46].

3. Results

Calculated GC content in all transcripts was 45.62%. Out of 57,150,222 obtained reads 99,621 contigs with N50 value of 1132 bp, maximum contig length of 15,627 bp and the average contig length of 666.42 bp were assembled (Table 2). Most of contigs were 200–500 bp in length (63,080). The number of contigs of 500–1000 bp in length was 17,127, the number of contigs of 1000–2000 bp in length was 13,650 and the number of contigs larger than 2000 bp was 5764. The FPKM values ranged from 0 to 20,100.87 with the median FPKM value of 1.65 and 1435 transcripts showed FPKM values > 1.000. Total of 43,845 transcripts (44%) were assigned with GO terms that explained genes' functions (Fig. 3). The terms 'cell' (GO: 0005623), 'cell parts' (GO: 0044464) and 'organelle' (GO: 0043226) were the most frequent ones in the cellular component category, while the terms 'metabolic process' (GO: 0008152), 'biological regulation' (GO: 0065007) and 'response to stimulus' (GO: 0050896) were the most highly represented ones in the biological process category. On the other hand, the terms 'binding' (GO: 0005488) and 'catalytic activity' (GO: 0003824) were remarkably enriched in the molecular function category. Further, out of total of 82,372 unigenes detected 41,132 (ca. 50%) were assigned to 378 KEGG pathways, mainly those included in carbohydrate/amino acid/energy/lipid metabolism, biosynthesis of secondary metabolites, genetic/environmental information processing, cellular processes and cell and organizational systems categories. The pathways with the highest number of assigned transcripts were biosynthesis of amino acids (99), carbon metabolism (96), purine metabolism (95), RNA transport (94), oxidative phosphorylation (88), protein processing in endoplasm (77) and steroid alkaloid biosynthesis (18) (which was the main pathway in the metabolic category) (Table S1).

Based on KEGG pathway analysis ten genes presented in Table 3, that were possibly involved in steroidal alkaloid biosynthesis pathway, namely 3-hydroxy-3-methylglutaryl-CoA reductase (HMG), mevalonate kinase (MK), pyrophosphomevalonate decarboxylase (MVD), isopentenyl pyrophosphate isomerase (IPP), farnesyldiphosphate synthase (FPS), squalene synthase (SQS), cycloartenol synthase (CAS), sterol-C24-methyltransferase (SMT), cytochrome P450/obtusifoliol 14α-demethylase (CYP51) and UDP-glucosyltransferase (UGT), were selected for further analysis (Fig. S2). Their expression profiles were tested in six different crown imperial tissues (leaf, bulb, anther, ovary, sepal and petal) during two developmental stages (stem elongation and flower development) and in two tissues (fruit and bulb) during the seed head stage (Fig. 4). Tissue expression analysis revealed that in overall time course genes encoding for FPS, SQS and MVD were maximally expressed during the stem development in ovary and minimally expressed during the flower development in leaf (FPS and SQS)/bulb (MVD). On the other hand, the genes encoding for HMG, MK and IPP were maximally expressed during the flower development in anthers and minimally expressed during the stem elongation in bulb (HMG)/in petals (MK and IPP) (in the case of IPP encoding gene the expression level was also minimal in ovary during the flower development stage). The genes

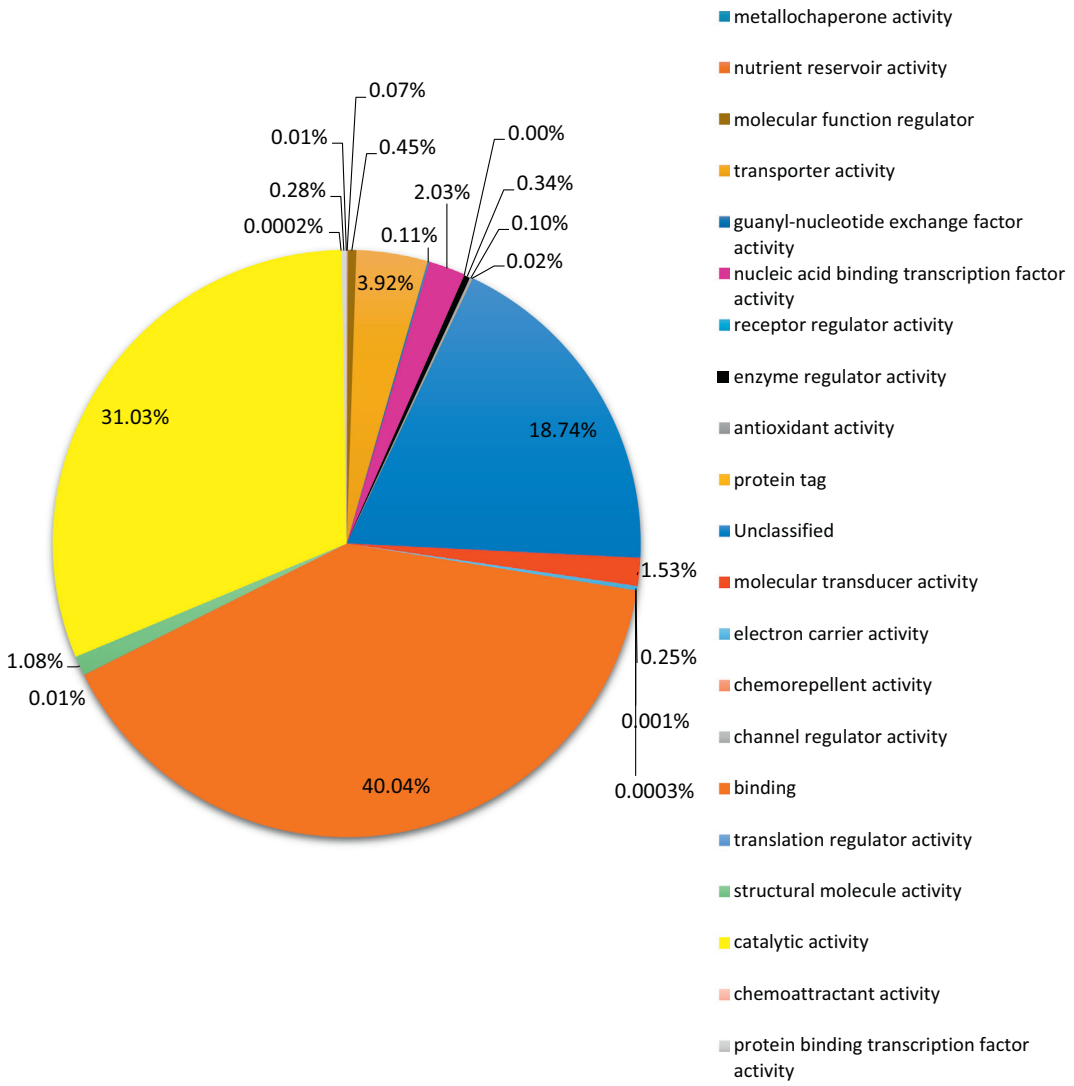

Fig. 3. (*continued*)

encoding for CAS, SMT and CYP51 displayed the highest expression level in petals during the stem development phase and the lowest expression level in bulb during the same phase, while the gene encoding for UGT showed the maximal expression in sepals during the flower development and the minimal expression in petals during the stem elongation phase.

Additionally, results for gene expression analyzed per each developmental phase and tissue discovered that during the stem elongation stage HMG, MK, MVD, IPP, FPS and SQS encoding genes showed the highest expression in ovary, while CAS, SMT and CYP51 encoding genes showed the highest expression in petals and UGT showed the highest expression encoding gene in sepals. During the flower development stage HMG, MK, IPP and FPS encoding genes showed the highest expression in anthers, while CYP51 encoding gene showed the highest expression in petals, SQS and UGT encoding genes showed the highest expression in sepals, MVD and CAS encoding genes showed the highest expression in ovary and SMT encoding gene showed the highest expression both in ovary and petals. During the seed head stage HMG, MK, MVD, FPS, CYP51, CAS and SQS encoding genes showed the highest expression in fruit, while IPP, SMT and UGT encoding genes showed the highest expression in bulb. Moreover, a comparison of expression levels of all ten genes across all developmental stages and tissues revealed that during the flower development in all analyzed tissues SQS encoding gene showed significantly higher expression than any of other analyzed genes, with the fold change ranging from 2 to almost 100,000 in different tissues/stages. Further, the expression profiles of these ten genes were compared to the expression profiles of their homologous in *A. thaliana* (http://www.bar.utoronto.ca) [47] revealing a high overall similarity in trends of gene expression in different tissues and developmental stages between these two species. For instance, SMT encoding gene expression was the highest in petal tissue in both Arabidopsis and *Fritillaria* and HMG encoding gene was maximally expressed in male reproductive organs during the flowering development phase in both species. The only difference was observed in SQS encoding gene expression, which was reported to be the highest in both male and female reproductive organs in Arabidopsis, while its *Fritillaria* homologue showed the highest expression in the ovary.

Further transcriptome data analysis led to annotation of 267 unique transcripts as CYP450s, including CYP90D1, CYP724B1 and CYP734 (Table S2), while 122 unique transcripts were identified as glycosyltransferases, including sterol glucosyltransferase (SGT) and UDP-glycosyltransferase (UGT) (Table S3). Also, based on the sequence comparison with the family of known transcription factors additional 1583 transcripts were assigned to at least 57 transcription factor gene families, like MYB, MADS, bHLH, bZIP, WRKY, ERF, AP2,zinc finger families etc. They accounted for 1.58% of total *F. imperialis* transcripts and among them the transcription factors of MYB, NAC, HB and bHLH families were identified as the most frequent ones (Fig. 5). Moreover,

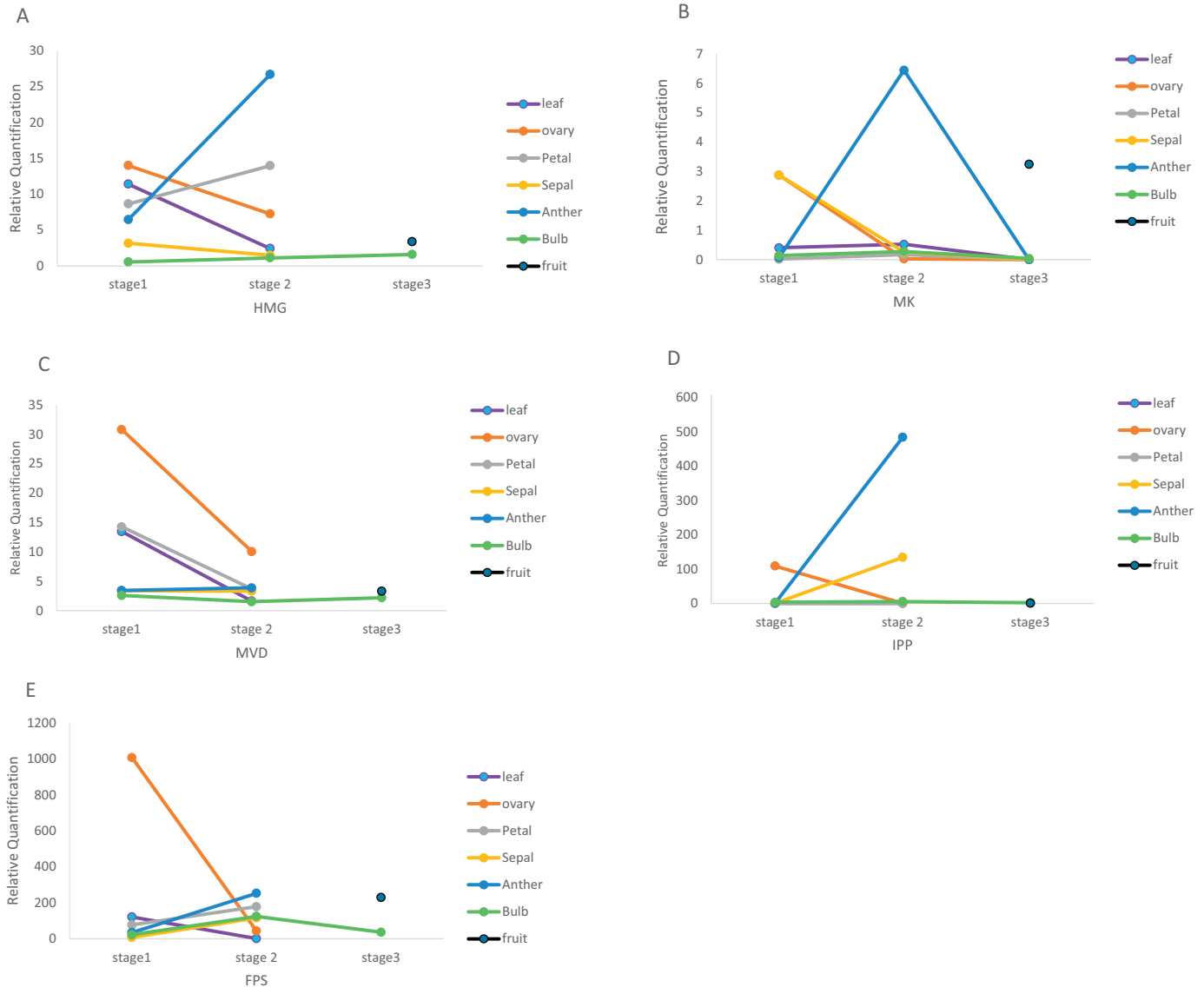

Fig. 4. Expression analysis of ten genes possibly involved in steroid alkaloid biosynthesis of *F. imperialis* in different tissues and developmental stages: Fim-HMG (A), Fim-MK(B), Fim-MVD(C), Fim-IPP(D), Fim-FPS(E), Fim-SQS(F), Fim-CAS(G), Fim-SMT(H), FiCYP51(I) and Fim-UGT(J).

proteins of WRKY, ERF, AP2 and zinc finger families were also represented in crown imperial transcriptome counting for the potential candidates that are worth of further investigation for their possible involvement in the biosynthesis of steroidal alkaloids in *F. imperialis* (Table S4). Further, total of 2077 Transposable Elements was identified in crow imperial transcriptome data (Table S5) with the Helitron Family (among DNA transposons) and Copia family (among retrotransposons) being the most frequent ones in *Fritillaria imperialis*. On the other hand, 55,776 transcripts (ca. 55%) could not be matched to any of known genes or annotation descriptions (i.e. unnamed, unknown, uncharacterized or hypothetical protein).

The phylogenetic analyses of 10 candidate genes showed sequence similarity that ranged up to 71% with these genes in plants of the same genus like *Fritillaria thunbergii* and *Fritillaria unibracteata*, than mainly of other monocots like *Phoenix dactylifera* and *Elaeis guineensis* of the family of the family Arecaceae, *Paris polyphylla* of the family Melanthiaceae, *Dioscorea composite* of the family Dioscoreaceae, *Musa acuminata* of the family Musaceae, *Ananas comosus* of the family Bromeliaceae, *Lilium longiflorum* of the family Liliaceae, *Phaleonopsis equestris* of the family Orchidaceae and even of some dicots like *Nelumbo nucifera* of the family Nelumbonaceae. Phylogenetic relationships of crown imperial HMG, MK, MVD, IPP, FPS, SQS, CAS, SMT, CYP51 and UGT encoding genes are given at the Fig. 6. In some of the analyzed genes, like *HMG, MVD, IPP, SQS, MK* and *CAS,* phylogenetic analysis revealed clustering in two separate clades for the sequences of monocotyledonous and dicotyledonous species, in which crown imperial clustered with other monocots (except in the case of MK encoding gene in which it was separated from both clades). In the rest of the analyzed genes, namely *FPS, SMT, UGT* and *CYP51*, there were no distinct separation of clustering of sequences of monocots and dicots and *F. imperialis* clustered the closest with other monocots (except in the case of CYP51 and SMT encoding genes in which it was separated from the main clade). Moreover, OMA analysis revealed orthologous of these 10 genes in monocots of different genera of the family Poaceae: *HMG* in *Oryza sativa*, *CAS* in *Oryza glaberrima*, *SMT* in *Oryza rufipogon*, *UGT* in *Oryza brachyantha, MK* in *Eragrostis tef*, *IPP* in *Setaria italic*, *SQS* in *Zea mays*, *CYP51* in *Triticum aestivum*, and *MVD* and *FPS* in *Aegilops tauschii*.

4. Discussion

The extensive metabolic activities in *F. imperialis* were suggested by matching a high percentage of annotated transcripts to various sub-

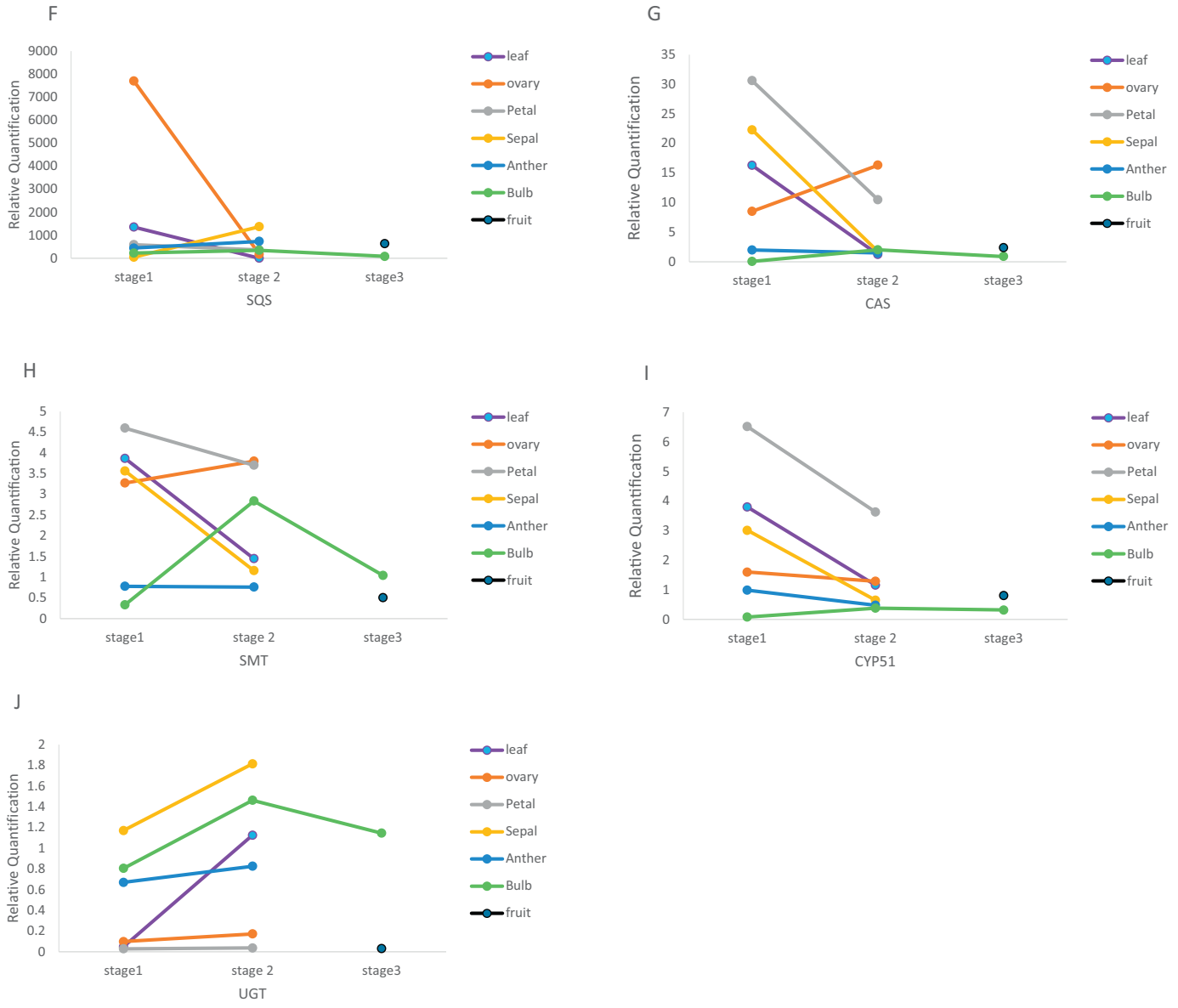

Fig. 4. (*continued*)

Table 2
The summary of *Fritillaria imperialis* transcriptome assembly main characteristics (nt is the abbreviation for nucleotide).

Total number of reads	57,150,222
Total Bases	5,566,522,103
Read length (nt)	101
Percent of GC content (%)	45.62
Total number of contigs	99,621
Maximum contig length (nt)	15,627
Median length of contigs (nt)	357
Minimum contig length (nt)	201
Total number of unigenes	82,372
Total assembled bases	66,389,167
N50	1132

categories of metabolic processes. Comparison of transcriptomes and their annotations between two *Fritillaria* species, namely *F. imperialis* and *F. cirrhosa*, showed that metabolic pathways presented the most abundant category (24.0% and 26.0%, respectively) with the steroid alkaloid biosynthesis as the major route in both species [48]. The value of GO categories were similar to GO annotations reported for *Fritillaria cirrhosa* and other medicinal plants [48,49]. In crown imperial among all annotated genes ten were strongly supported to be associated with steroidal alkaloid pathway and as such were used for the identification of putative orthologues/paralogues genes. This analysis showed that all genes displayed high similarity with the same genes involved in sterol and isoprenoid biosynthesis processes of other higher plants, especially with monocots of the family Poaceae. Further phylogenetic analysis of ten analyzed genes revealed clustering with the key genes in triterpene/steroid biosynthesis pathway in monocotyledonous plants from the families Liliaceae, Melanthiaceae, Bromeliaceae, Arecaceae, Musaceae, Poaceae, Orhidaceae, suggesting the possibility of sharing the similar role with their homologous in crown imperial.

For the crown imperial gene encoding for CAS the closest phylogenetic clustering was with similar genes from *F. thunbergii* and *Paris polyphylla* var. yunnanensis, while for the gene encoding for MVD phylogenetic clustering was the closest to the similar genes from *Lilium* hybrid cultivar Freya and then *Phalaenopsis equestris* and *Dendrobium catenatum* and for those encoding for FPS and HMG the closest phylogenetic clustering was with the similar genes from *F. unibracteata* (and then *Lilium longiflorum* in the case of FPS). In the case of genes encoding

Table 3
Genes identified as possibly involved in steroid alkaloid biosynthesis in *Fritillaria imperialis* by using de novo transcriptome approach during the plant flowering phase. Following the gene name are given gene abbreviation symbol, its accession number (https://www.ncbi.nlm.nih.gov/Genbank/update.html), transcript and gene identifier, enzyme commission (EC) number and FPKM value.

Gene name	Symbol	Gene Bank accession number	transcript_id	gene_id	EC number	FPKM
Hydroxymethylglutaryl-CoA reductase	HMG	MG255717	c27381_g3_i1	c27381_g3	1.1.1.34	50.57
Mevalonate kinase	MK	MG255772	c31974_g1_i1	c31974_g1	2.7.1.36	12.29
Diphosphomevalonate decarboxylase	MVD	MG255773	c21591_g1_i2	c21591_g1	4.1.1.33	7.14
Isopentenyl diphosphate isomerase	IPP	MG255774	c33471_g1_i3	c33471_g1	5.3.3.2	7.05
Farnesyl diphosphate synthase	FPS	MG255775	c27433_g1_i1	c27433_g1	2.5.1.10	53.26
Squalene synthase	SQS	MG255776	c27910_g1_i1	c27910_g1	2.5.1.21	38
Cycloartenol Synthase	CAS	MG255777	c31388_g1_i1	c31388_g1	5.4.99.8	6.29
24-methylenesterol *C*-methyltransferase	SMT	MG255778	c27392_g1_i1	c27392_g1	2.1.1.143	91.74
Sterol 14-demethylase	CYP51	MG255779	c25312_g1_i1	c25312_g1	1.14.13.70	50
UDP-glucosyltransferase	UGT	MG255780	c50670_g1_i1	c50670_g1	–	0.089

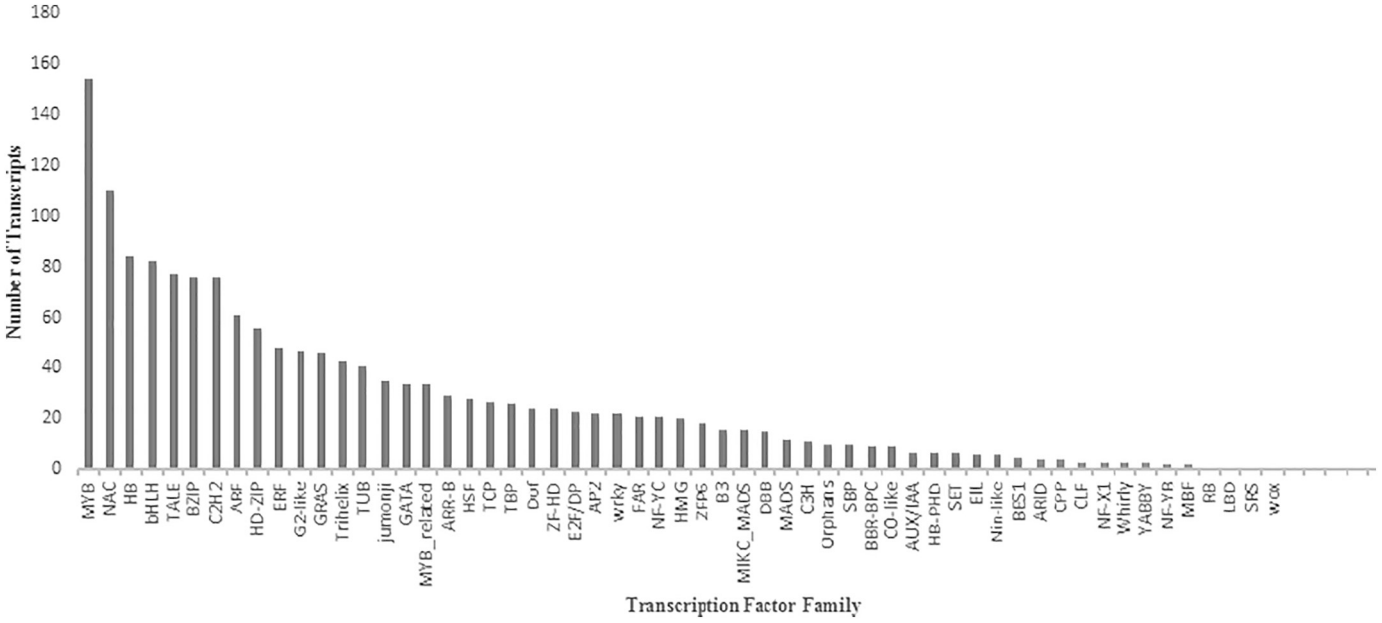

Fig. 5. The distribution of different classes of transcription factors detected in *F. imperialis* transcriptome data.

for CYP51 and MK phylogenetic clustering was significantly different with the gene from crown imperial being in a separate clade from the similar genes from *F. cirrhosa* and either from walnut, wild strawberry, Chinese rose, potato, etc. (*CYP51*) or from garden asparagus, Chinese sage, lettuce, potato, etc. (*MK*). Likewise, the gene encoding for SMT was in a separate clade from the family Poaceae and Solanaceae, which clustered in the joint clade. In the case of crown imperial gene encoding for IPP phylogenetic clustering was closer to the similar genes from *Ananas comosus*, *Phoenix dactylifera, Musa acuminata* subsp. *malaccensis* and other monocots than to the similar genes from *Nelumbo nucifera*, *Glycine* max*, Solanum lycopersicum* and other dicots. For the crown imperial gene encoding for SQS the closest phylogenetic clustering was with the similar genes from *F. thunbergii* and *F. unibracteata* and then *Paris polyphylla* var. yunnanensis, while in the case of UGT encoding gene the closest phylogenetic clustering was with the similar genes with *Phalaenopsis equestris, Dendrobium catenatum* and *Eucalyptus grandis*.

Finding the orthologous genes of ten analyzed genes mainly within species of the family Poaceae and phylogenetic clustering of ten genes with the similar genes in other monocots (mostly in the separate clade from the similar genes in dicots), supports the independent origin of steroidal alkaloid pathway in monocots and dicots. Interestingly, there were significant differences in the level of similarity of analyzed genes within the genus *Fritillaria* itself. In the case of CAS, FPS and HMG encoding genes in crown imperial, they clustered closely to *F. unibracteata* or *F. thunbergii,* while SQS encoding gene clustered closely to both *F. unibracteata* and *F. thunbergii*. On the other hand, CYP51 and MK encoding genes clustered separately from *F. cirrhosa* (as well as from all other species in the phylogenetic tree), suggesting the existence of at least some differences in the gene network that underlies the steroidal alkaloid biosynthesis in the species of the *Fritillaria* genus. Nevertheless, there are examples of the convergent evolution of some of the genes involved in the steroidal alkaloid biosynthesis (in some cases by the engagement genes from different gene subfamilies, in other cases by employing genes form the other existing pathways) [11], which could be a possibility worth further investigation concerning these genes in monocots and dicots, but also in different species of the genus *Fritillaria*.

Expression profiles of these 10 candidate genes were analyzed further in different tissues and developmental stages. In order to discuss correlations between expression patterns of ten analyzed genes we ought to briefly overview their involvement in plant sterol alkaloid biosynthesis according to the literature [8–10,48,50–59]. Mevalonate synthesis pathway that converts HMG-CoA to isoprenoids involves several enzymes, among which are HMG (as the main overall control point of the mevalonate pathway), MK, MVD and IPP. Produced isoprenoids are further processed by FPS and SQS, which present two main control points in terpenoid and sterol biosynthesis. The following steps include conversion of produced intermediates by CAS into cycloartenol, the first polycyclic sterol intermediate which is further converted into dominant sterols, as well as CYP51 - obtusifoliol14α-demethylase and SMT responsible for 24-methyl sterols biosynthesis. Downstream of

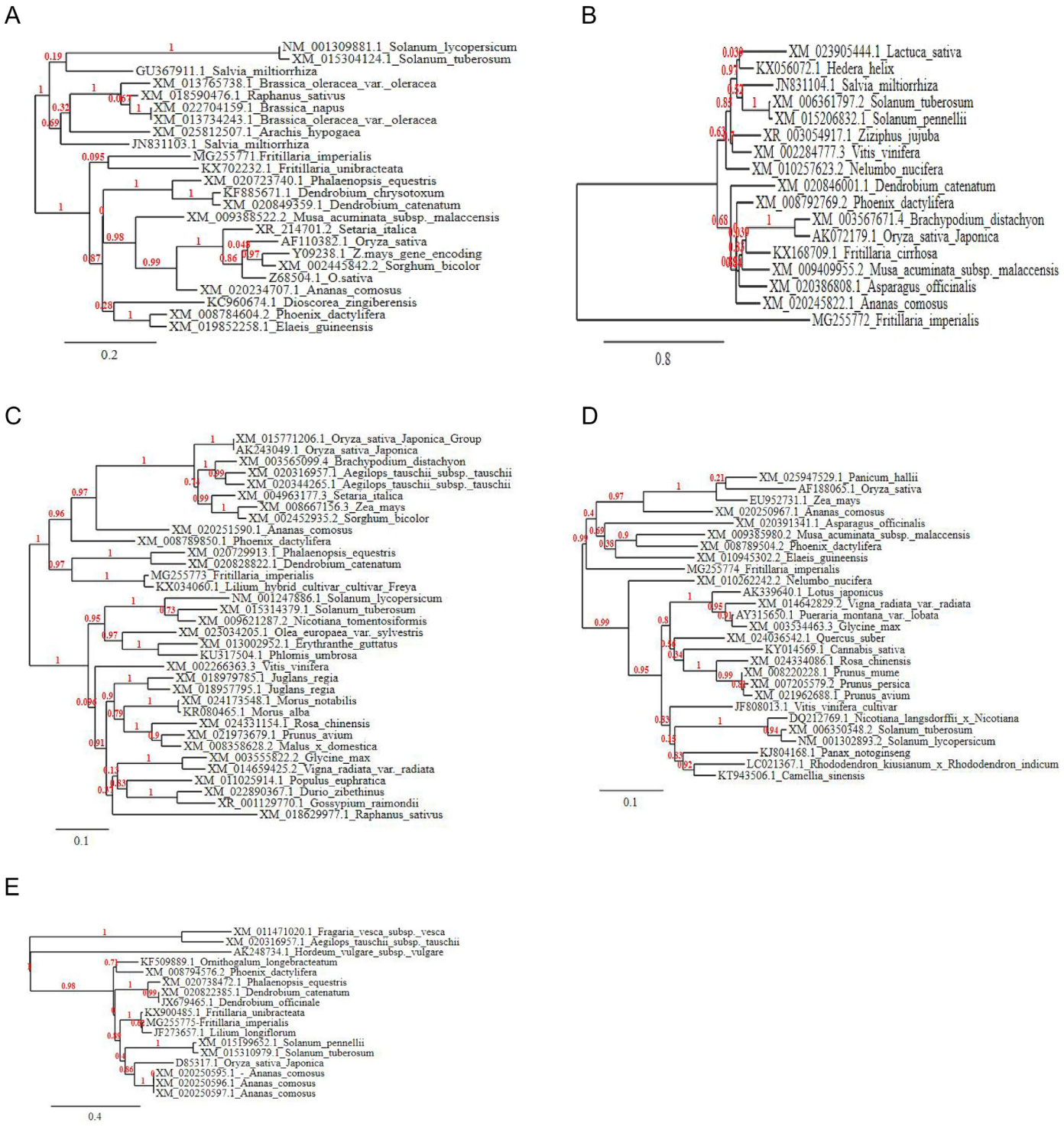

Fig. 6. Phylogenetic trees presenting relationships of ten candidate genes with the similar genes from other plant species using Phylogeny.fr v3.1/3.0: Fim-HMG (A), Fim-MK(B), Fim-MVD(C), Fim-IPP(D), Fim-FPS(E), Fim-SQS(F), Fim-CAS(G), Fim-SMT(H), FiCYP51(I) and Fim-UGT(J).

these enzymes acts UGT, which is responsible for the glycosylation of different types of sterols, enabling steroidal glycoalkaloids biosynthesis.

Considering the expression of genes encoding for enzymes that control three major regulation points in plant sterol biosynthesis, namely HMG, FPS and SQS, results showed that during the transition from the stem elongation phase to the flower development phase the expression level of all three genes decreased in leaf and ovary, while it increased in anthers and bulb. The only difference in the expression profiles of these three genes was observed in sepal and petal tissues: HMG encoding gene expression decreased in sepals and increased in petals, SQS encoding gene expression increased in sepals and decreased in petals, while FPS encoding gene expression increased in both sepals and petals. The expression level of SQS was overall high in all tissues and stages with the only exception found in leaf during the flower development, which may suggest that SQS is the key enzyme in *F. imperialis* sterol biosynthesis. Since the expression of HMG, FPS and SQS encoding genes in other plant species was directed by numerous developmental and environmental signals resulting in the existence of

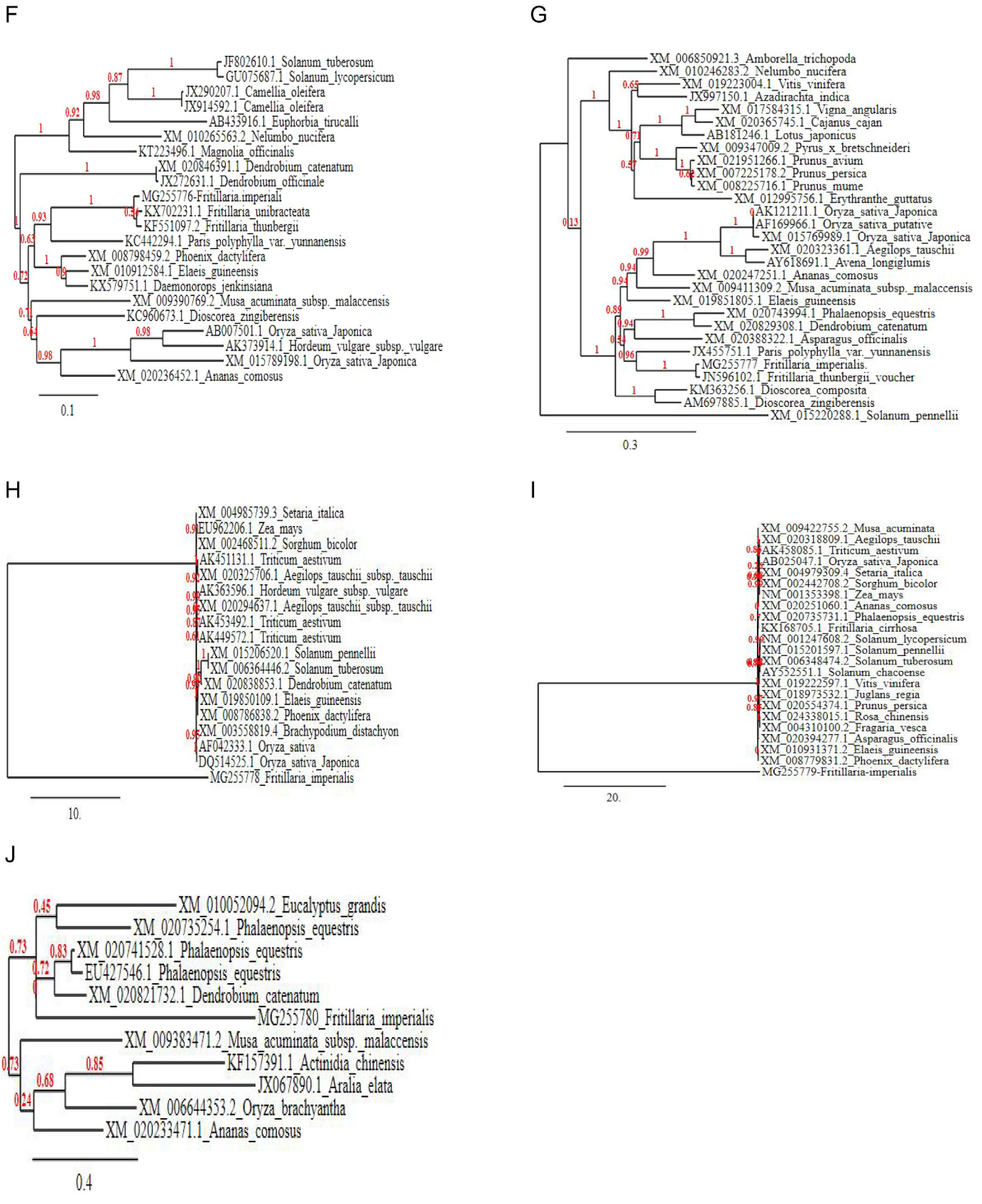

Fig. 6. (*continued*)

their different isoforms, characterized by different tissue and end-product specificity [52–55], most likely these three genes display different isoforms with variable distribution in tissues during different developmental phases in crown imperial as well (i.e. for HMG encoding gene three isoforms were detected among crown imperial transcripts - results not shown). Nevertheless, trends in the expression of genes

encoding for HMG, FPS and SQS were highly similar, which may imply the possibility of co-regulation of expression of these genes and/or their isoforms.

Further, tissue expression profiles of genes encoding for MK and IPP displayed highly similar overall trend during the stem elongation and the flower development phase. Also, the transition from the stem elongation to the flower development phase was marked by the increase of the expression level of both genes in leaf, petal, anther and bulb tissues and by the decrease of the expression level of both genes in ovary tissue. The only difference during this phase transition was observed in sepal tissue in which MK encoding gene's expression level decreased, while IPP encoding gene's expression level increased. However, during the seed head stage IPP encoding gene displayed a higher expression level in bulb than in fruit, while MK encoding gene expression profile showed the opposite trend. On the other hand, tissue expression profiles of genes encoding for CYP51, CAS and SMT shared the overall trend during the stem elongation phase, but differences were observed during the flowering development and the seed head stage. During the transition from the stem elongation to the flower development phase all three genes displayed the increase of the expression level in bulb and the decrease of the expression level in leaf, petal, sepal and anther tissue, with the only difference observed in ovary tissue in which the expression of CAS and SMT encoding genes increased, while that of CYP51 encoding gene decreased. Correspondingly to the trend observed in HMG, FAS and SQS encoding genes expression, the overall trend in the expression of CAS, SMT and CYP51 encoding genes displayed high similarity, which may imply co-regulation of these genes too. Implied co-regulation of expression of these candidate genes may also suggest their potential organization into appropriate metabolic gene clusters, like the ones recorded in the Solanaceae. This is an assumption worth of further investigation, especially when one takes into consideration records like the one on crassulacean acid metabolism (CAM) convergent evolution in flowering plants reporting that “convergent evolution of protein sequence and temporal gene expression underpins the independent emergences of (CAM) from C3 photosynthesis” [60].

Interestingly, in all tissues except sepals and bulb, MK and IPP encoding genes showed directly opposite expression trend to that of CAS, SMT and CYP51 encoding genes, suggesting that with the higher level of sterols being synthesized the suppression of the activity of isoprenoid pathway occurs (although not entirely because intermediates of this pathway are needed for non-sterol products synthesis as well). Also, UGT encoding gene displayed a distinct expression profile: the expression level increased in all tissues during the transit from the stem elongation to the flower development phase, but in the comparison to the expression levels of other nine genes its expression was kept at the generally low level in all tissues during all stages, illustrating its role in steroidal glycoalkaloids biosynthesis. Among other detected CYP450s there were at least three more CYPs interesting to steroidal alkaloid biosynthesis, like CYP90D1 (responsible for C22-hydroxilation in *A. thaliana*), CYP724B1 (responsible for C-23 hydroxilation in *A. thaliana*) and CYP734 (responsible for C-26 hydroxilation in *A. thaliana*), which most likely serve the same function in crown imperial as their homologues in Arabidopsis do. Presented results may also provide the basis to explore the side-route pathways in the biosynthesis of secondary metabolites in *F. imperialis*.

Considering transcription factors (TFs) detected in crown imperial transcriptome, there are many TFs that belong to the protein families recorded to be involved in the biosynthesis of steroidal alkaloids in other plants, like an AP2/ERF transcription factor (GAME9) involved in the steroidal alkaloids biosynthesis in the Solanaceae and others [10,61]. Further, many transposon elements (TEs) were detected in crown imperial transcriptome, as expected, which is in concordance with important roles of TEs in the functional and evolutionary dynamics of the large plants genomes, especially in *Fritillaria* species [62,63]. Moreover, since out of 82,372 unigenes detected almost 50% could not be annotated, there is a reasonable possibility that some of these genes may have a function important in main or side-route pathway of steroidal alkaloid biosynthesis or some other pathways/networks, which is why the further progress in crown imperial genomics would be highly appreciated in biomedical, bioengineering and biocatalyst research.

5. Conclusion

Fritillaria imperialis transcriptome analysis revealed numerous uncharacterized genes, but it also revealed the presence of at least ten genes known to be associated with plants' steroidal alkaloid biosynthesis. Further orthologues/paralogues, phylogenetic and expression analysis of these ten genes suggested that they are the most likely candidates for biosynthesis of steroidal alkaloids in crown imperial, with a strong implication that SQS may act as the key enzyme of this process in this plant species. Additionally, expression profiles of analyzed genes revealed the possibility of their co-regulation, which may easily be achieved by the arrangement of these genes in jointly regulated clusters, like it was the case in the Solanaceae species. Having in mind convergent evolution of steroidal biosynthetic pathways in flowering plants and cases of convergent protein evolution as well as convergent evolution of orchestration of temporal expression of such proteins, this opens a reasonable possibility of the existence of similar gene cluster organization in steroidal alkaloid pathway in the family Liliaceae or at least in some species of its genus *Fritillaria*. Moreover, numerous transcriptional factors that belong to protein families known to regulate steroidal alkaloid biosynthesis in other plants were revealed in crown imperial, which opens the door for additional computational and experimental analysis in elucidating the ones involved in the same process in *F. imperialis*. Even though all suggested implications need to be investigated further by genomics, proteomics, biochemical and computational analysis before reaching any conclusions, the obtained results support transcriptomics as a highly efficient and useful approach in elucidating genes underlying the complex biochemical pathways.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2018.09.008.

Acknowledgment

This research was supported by Shahrekord University. We would also like to acknowledge the Biotechnology Research Institute at the University of Shahrekord, which assisted in the realization of this project.

References

[1] G.A. Cordell, Chemistry and Biology, Elsevier Science, 2008, https://books.google.com/books?id=Y7kVy1iwdYkC.
[2] H.-J. Li, Y. Jiang, P. Li, Chemistry, bioactivity and geographical diversity of steroidal alkaloids from the Liliaceae family, Nat. Prod. Rep. 23 (2006) 735–752, https://doi.org/10.1039/B609306J.
[3] V. De Luca, B. St Pierre, The cell and developmental biology of alkaloid biosynthesis, Trends Plant Sci. 5 (2000) 168–173.
[4] M. Kushiro, T. Nakano, K. Sato, K. Yamagishi, T. Asami, A. Nakano, et al., Obtusifoliol 14α-demethylase (CYP51) antisense Arabidopsis shows slow growth and long life, Biochem. Biophys. Res. Commun. 285 (2001) 98–104.
[5] D. Bowles, E.-K. Lim, B. Poppenberger, F.E. Vaistij, Glycosyltransferases of lipophilic small molecules, Annu. Rev. Plant Biol. 57 (2006) 567–597.
[6] G.A. Cordell, T. Choi, Alkaloids and their Biosynthesis, in: M.J. Kato, J.M. Pezzuto (Eds.), Photochemistry and Pharmacognosy, UNESCO-EOLSS, USA, 2011, http://www.academicjournals.org/JMPR.
[7] M.N. Akhtar, M.I. Choudhary, B. Sener, I. Erdogan, Y. Tsuda, New class of steroidal alkaloids from *Fritillaria imperialis*, Phytochemistry 63 (2003) 115–122.
[8] H.-W. Nützmann, A. Osbourn, Gene clustering in plant specialized metabolism, Curr. Opin. Biotechnol. 26 (2014) 91–99.
[9] M. Itkin, I. Rogachev, N. Alkan, T. Rosenberg, S. Malitsky, L. Masini, et al., GLYCOALKALOID METABOLISM1 is required for steroidal alkaloid glycosylation and prevention of phytotoxicity in tomato, The Plant Cell (2011), https://doi.org/10.1126/science.1240230.
[10] P.D. Cárdenas, P.D. Sonawane, J. Pollier, R.V. Bossche, V. Dewangan, E. Weithorn, et al., GAME9 regulates the biosynthesis of steroidal alkaloids and upstream

isoprenoids in the plant mevalonate pathway, Nat. Commun 7 (2016) 10654.
[11] M.M. Augustin, D.R. Ruzicka, A.K. Shukla, J.M. Augustin, C.M. Starks, M. O'Neil-Johnson, et al., Elucidating steroid alkaloid biosynthesis in *Veratrum californicum*: production of verazine in Sf9 cells, Plant J. 82 (2015) 991–1003.
[12] C.O. Nnadi, N.J. Nwodo, M. Kaiser, R. Brun, T.J. Schmidt, Steroid Alkaloids from *Holarrhena africana* with strong activity against *Trypanosoma brucei rhodesiense*, Molecules 22 (2017) 1129.
[13] P.D. Day, M. Berger, L. Hill, M.F. Fay, A.R. Leitch, I.J. Leitch, et al., Evolutionary relationships in the medicinally important genus *Fritillaria* L.(Liliaceae), Mol. Phylogenet. Evol. 80 (2014) 11–19.
[14] Bailey LH.Manual of Cultivated Plants. MacMillan Company. New York 218–219
[15] C. Sun, Y. Sun, J. Song, C. Li, X. Li, X. Zhang, et al., Discovery of genes related to steroidal alkaloid biosynthesis in *Fritillaria cirrhosa* by generating and mining a dataset of expressed sequence tags (ESTs), J. Med. Plants Res. 5 (2011) 5307–5314 http://www.academicjournals.org/JMPR.
[16] L. van der Fits, J. Memelink, ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism, Science 289 (2000) 295–297.
[17] N. Kato, E. Dubouzet, Y. Kokabu, S. Yoshida, Y. Taniguchi, J.G. Dubouzet, et al., Identification of a WRKY protein as a transcriptional regulator of benzylisoquinoline alkaloid biosynthesis in *Coptis japonica*, Plant Cell Physiol. 48 (2007) 8–18.
[18] B. Pauw, F.A. Hilliou, V.S. Martin, G. Chatel, C.J. de Wolf, A. Champion, et al., Zinc finger proteins act as transcriptional repressors of alkaloid biosynthesis genes in *Catharanthus roseus*, J. Biol. Chem. 279 (2004) 52940–52948, https://doi.org/10.1074/jbc.M404391200.
[19] K. Tanase, C. Nishitani, H. Hirakawa, S. Isobe, S. Tabata, A. Ohmiya, et al., Transcriptome analysis of carnation (*Dianthus caryophyllus* L.) based on next-generation sequencing technology, BMC Genomics 13 (2012) 1.
[20] Q. Tang, X. Ma, C. Mo, I.W. Wilson, C. Song, H. Zhao, et al., An efficient approach to finding *Siraitia grosvenorii* triterpene biosynthetic genes by RNA-seq and digital gene expression analysis, BMC Genomics 12 (2011) 1.
[21] C. Sui, J. Zhang, J. Wei, S. Chen, Y. Li, J. Xu, et al., Transcriptome analysis of *Bupleurum chinense* focusing on genes involved in the biosynthesis of saikosaponins, BMC Genomics 12 (2011) 1.
[22] S. Kalra, B.L. Puniya, D. Kulshreshtha, S. Kumar, J. Kaur, S. Ramachandran, et al., *De novo* transcriptome sequencing reveals important molecular networks and metabolic pathways of the plant, *Chlorophytum borivilianum*, PLoS One 8 (2013) e83336.
[23] N.R.R. Reddy, R.H. Mehta, P.H. Soni, J. Makasana, N.A. Gajbhiye, M. Ponnuchamy, et al., Next generation sequencing and transcriptome analysis predicts biosynthetic pathway of sennosides from Senna (*Cassia angustifolia* Vahl.), a non-model plant with potent laxative properties, PLoS One 10 (2015) e0129422.
[24] M.N. Akhtar, M.I. Choudhary, Y. Tsuda, B. Sener, A. Khalid, M. Parvez, New steroidal alkaloids from *Fritillaria imperialis* and their cholinesterase inhibiting activities, Chem. Pharm. Bull. 50 (2002) 1013–1016.
[25] E. Ebrahimie, A. Habashy, M. Mohammadie-Dehcheshmeh, M. Ghannadha, B. Ghareyazie, B. Yazdi-Amadi, Direct shoot regeneration from mature embryo as a rapid and genotype-independent pathway in tissue culture of heterogeneous diverse sets of cumin (*Cuminum cyminum* L.) genotypes, In Vitro Cell. Develop. Biol. Plant 42 (2006) 455–460.
[26] M. Mohammadi-Dehcheshmeh, A. Khalighi, R. Naderi, M. Sardari, E. Ebrahimie, Petal: a reliable explant for direct bulblet regeneration of endangered wild populations of *Fritillaria imperialis* L, Acta Physiol. Plant. 30 (2008) 395–399.
[27] J. Marienhagen, M. Bott, Metabolic engineering of microorganisms for the synthesis of plant natural products, J. Biotechnol. 163 (2013) 166–178.
[28] H. Kries, S.E. O'Connor, Biocatalysts from alkaloid producing plants, Curr. Opin. Chem. Biol. 31 (2016) 22–30.
[29] P. Chomczynski, A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples, BioTechniques 15 (1993) 532–534.
[30] O. Mueller, S. Lightfoot, A. Schroeder, RNA Integrity Number (RIN)–Standardization of RNA Quality Control, Agilent Application Note, Publication, 2004, pp. 1–8.
[31] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114–2120.
[32] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, et al., Full-length transcriptome assembly from RNA-Seq data without a reference genome, Nat. Biotechnol. 29 (2011) 644–652, https://doi.org/10.1038/nbt.1883.
[33] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, BMC Bioinformatics. 12 (2011) 323.
[34] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, et al., STAR: ultrafast universal RNA-seq aligner, Bioinformatics 29 (2013) 15–21.
[35] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[36] M. Kanehisa, S. Goto, KEGG: Kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (2000) 27–30.
[37] Y. Moriya, M. Itoh, S. Okuda, A.C. Yoshizawa, M. Kanehisa, KAAS: an automatic genome annotation and pathway reconstruction server, Nucleic Acids Res. 35 (2007) W182–W185.
[38] D.R. Nelson, The cytochrome p450 homepage, Hum. Genom. 4 (2009) 59.
[39] C.M. Gachon, M. Langlois-Meurinne, P. Saindrenan, Plant secondary metabolism glycosyltransferases: the emerging functional analysis, Trends Plant Sci. 10 (2005) 542–549.
[40] S.-F. Li, G.-J. Zhang, X.-J. Zhang, J.-H. Yuan, C.-L. Deng, L.-F. Gu, et al., DPTEdb, an integrative database of transposable elements in dioecious plants, Database 2016 (2016).
[41] A. Untergasser, I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, et al., Primer3—new capabilities and interfaces, Nucleic Acids Res. 40 (2012) e115.
[42] J. Ye, G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, T.L. Madden, Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction, BMC Bioinformatics. 13 (2012) 134.
[43] J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, et al., Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes, Genome Biol. 3 (2002) 1.
[44] R.D. McCurdy, J.J. McGrath, A. MacKay-Sim, Validation of the comparative quantification method of real-time PCR analysis and a cautionary tale of housekeeping gene selection, Gene Ther. Mol. Biol. 12 (2008) 15–24.
[45] A. Dereeper, S. Audic, J.-M. Claverie, G. Blanc, BLAST-EXPLORER helps you building datasets for phylogenetic analysis, BMC Evol. Biol. 10 (2010) 8.
[46] A.M. Altenhoff, N.M. Glover, C.-M. Train, K. Kaleb, A. Warwick Vesztrocy, D. Dylus, et al., The OMA orthology database in 2018: retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces, Nucleic Acids Res. 46 (2017) D477–D485.
[47] D. Winter, B. Vinegar, H. Nahal, R. Ammar, G.V. Wilson, N.J. Provart, An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets, PLoS One 2 (2007) e718.
[48] Q. Zhao, R. Li, Y. Zhang, K. Huang, W. Wang, J. Li, Transcriptome analysis reveals *in vitro*-cultured regeneration bulbs as a promising source for targeted *Fritillaria cirrhosa* steroidal alkaloid biosynthesis, 3 Biotech 8 (2018) 191.
[49] X.-J. Han, Y.-D. Wang, Y.-C. Chen, L.-Y. Lin, Q.-K. Wu, Transcriptome sequencing and expression analysis of terpenoid biosynthesis genes in *Litsea cubeba*, PLoS One 8 (2013) e76890.
[50] E.E. Johnson, R. Jetter, G. Wasteneys, Rapid induction of the triterpenoid pathway in *Arabidopsis thaliana* mesophyll protoplasts, Biotechnol. Lett. 36 (2014) 855–858.
[51] N. Nahar, E. Westerberg, U. Arif, A. Huchelmann, A.O. Guasca, L. Beste, et al., Transcript profiling of two potato cultivars during glycoalkaloid-inducing treatments shows differential expression of genes in sterol and glycoalkaloid metabolism, Sci. Rep. 7 (2017) 43268, https://doi.org/10.1038/srep43268.
[52] P. Jawahar, SNP Analysis of Farnesyl Diphosphate Synthase from Centella asiatica, https://books.google.com/books?id=VkuUDgAAQBAJ, (2017).
[53] R. Kribii, M. Arró, A. Arco, V. González, L. Balcells, D. Delourme, et al., Cloning and characterization of the *Arabidopsis thaliana* SQS1 gene encoding squalene synthase, FEBS J. 249 (1997) 61–69.
[54] K.L. Korth, B.A. Stermer, M.K. Bhattacharyya, R.A. Dixon, HMG-CoA reductase gene families that differentially accumulate transcripts in potato tubers are developmentally expressed in floral tissues, Plant Mol. Biol. 33 (1997) 545–551.
[55] B.A. Stermer, G.M. Bianchini, K.L. Korth, Regulation of HMG-CoA reductase activity in plants, J. Lipid Res. 35 (1994) 1133–1140.
[56] M. Itkin, U. Heinig, O. Tzfadia, A. Bhide, B. Shinde, P. Cardenas, et al., Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes, Science 1240230 (2013), https://doi.org/10.1126/science.1240230.
[57] S. Sawai, K. Ohyama, S. Yasumoto, H. Seki, T. Sakuma, T. Yamamoto, et al., Sterol side chain reductase 2 is a key enzyme in the biosynthesis of cholesterol, the common precursor of toxic steroidal glycoalkaloids in potato, The Plant Cell (2014).
[58] A.C. Diener, H. Li, W. Zhou, W.J. Whoriskey, W.D. Nes, G.R. Fink, Sterol methyltransferase 1 controls the level of cholesterol in plants, Plant Cell 12 (2000) 853–870.
[59] L. Arnqvist, P.C. Dutta, L. Jonsson, F. Sitbon, Reduction of cholesterol and glycoalkaloid levels in transgenic potato plants by overexpression of a type 1 sterol methyltransferase cDNA, Plant Physiol. 131 (2003) 1792–1799.
[60] X. Yang, R. Hu, H. Yin, J. Jenkins, S. Shu, H. Tang, et al., The Kalanchoë genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism, Nat. Commun. 8 (2017) 1899.
[61] Y. Yamada, F. Sato, Transcription factors in alkaloid biosynthesis, Int. Rev. Cell Mol. Biol. (2013) 339–382.
[62] S.-I. Lee, N.-S. Kim, Transposable elements and genome size variations in plants, Genom. Inform. 12 (2014) 87–97, https://doi.org/10.5808/GI.2014.12.3.87.
[63] C.M. Vicient, J.M. Casacuberta, Impact of transposable elements on polyploid plant genomes, Ann. Bot. 120 (2017) 195–207.