Plant Cell Rep
DOI 10.1007/s00299-016-2021-z

ORIGINAL ARTICLE

High-throughput sequencing and de novo transcriptome assembly of *Swertia japonica* to identify genes involved in the biosynthesis of therapeutic metabolites

Amit Rai[1] · Michimi Nakamura[1] · Hiroki Takahashi[2] · Hideyuki Suzuki[3] · Kazuki Saito[1] · Mami Yamazaki[1]

Received: 26 April 2016 / Accepted: 29 June 2016

Abstract

Key message **Here, we report potential transcripts involved in the biosynthesis of therapeutic metabolites in *Swertia japonica*, the first report of transcriptome assembly, and characterization of the medicinal plant from *Swertia* genus.**

Abstract *Swertia* genus, representing over 170 plant species including herbs such as *S. chirata*, *S. hookeri*, *S. longifolia*, *S. japonica*, among others, have been used as the traditional medicine in China, India, Korea, and Japan for thousands of years. Due to the lack of genomic and transcriptomic resources, little is known about the molecular basis involved in the biosynthesis of characteristic key bioactive metabolites. Here, we performed deep-transcriptome sequencing for the aerial tissues and the roots of *S. japonica*, generating over 2 billion raw reads with an average length of 101 bps. Using a combined approach of three popular assemblers, de novo transcriptome assembly for *S. japonica* was obtained, yielding 81,729 unigenes having an average length of 884 bps and N50 value of 1452 bps, of which 46,963 unigenes were annotated based on the sequence similarity against NCBI-nr protein database. Annotation of transcriptome assembly resulted in the identification of putative genes encoding all enzymes from the key therapeutic metabolite biosynthesis pathways. Transcript abundance analysis, gene ontology enrichment analysis, and KEGG pathway enrichment analysis revealed metabolic processes being up-regulated in the aerial tissues with respect to the roots of *S. japonica*. We also identified 37 unigenes as potential candidates involved in the glycosylation of bioactive metabolites. Being the first report of transcriptome assembly and annotation for any of the *Swertia* species, this study will be a valuable resource for future investigations on the biosynthetic pathways of therapeutic metabolites and their regulations.

Keywords *Swertia* genus · *S. japonica* transcriptome assembly · Secoiridoid metabolic pathways · Swertiamarin · MEP biosynthesis pathways

Communicated by P. P. Kumar.

Electronic supplementary material The online version of this article (doi:10.1007/s00299-016-2021-z) contains supplementary material, which is available to authorized users.

✉ Amit Rai
amit.rai@chiba-u.jp

✉ Mami Yamazaki
mamiy@faculty.chiba-u.jp

[1] Graduate School of Pharmaceutical Sciences, Chiba University, Inohana 1-8-1, Chuo-ku, Chiba 260-8675, Japan

[2] Medical Mycology Research Center, Chiba University, Chiba, Japan

[3] Department of Research and Development, Kazusa DNA Research Institute, Chiba, Japan

Introduction

Plants are the rich source of bioactive metabolites, and have been used for thousands of years for the treatment of various diseases. According to the World Health Organization (WHO), the traditional medicines are used over 75 % of the world's population as a primary treatment regime, while up to 7000 chemical compounds used in the phytotherapy are derived from the plant sources (Eddouks et al. 2014; Licciardi and Underwood 2011; Pan et al. 2014; Veeresham 2012). Yet, a lack of genetic information of participating processes that lead to their biosynthesis

Published online: 04 July 2016

restricts our ability to increase their production by in vitro or in vivo approaches (Weeks and Chang 2011). Availability of genome sequences, and knowledge of transcriptome expression profile and accumulating phytochemicals across different tissues provide a broad understanding of different on-going metabolic processes, which may serve as a basis to formulate strategies for increased biosynthesis of metabolites of interest (Rai and Saito 2016). Since the whole genome sequencing, assembly, and annotation have its challenges, including the cost involved, complexity of plant genome with redundancies, and computational resources required, very few medicinal plant genome has been sequenced so far. Recent advancements in the next-generation sequencing (NGS) with reduced operational cost, and development of computational resources to perform the de novo transcriptome assembly, annotation, and subsequent analysis have revolutionized the field of phytochemistry and natural medicine. It has gained a lot of popularity for studies particularly involving non-model plants with no genomic information available (Muranaka and Saito 2013; Saito 2013). RNA-seq-based transcriptome and transcripts abundance studies across different tissues provide key insights on the gene structure, function, and their role in the regulation of different associated active metabolic processes (Rai and Saito 2016; Saito 2013).

Among medicinal plants, *Swertia* herbs, belonging to the *Gentianaceae* family, hold an important position for being a rich source of therapeutic metabolites (Brahmachari et al. 2004; Dutt et al. 1996; Neerja Pant and Jain 2000). *Swertia* genus includes about 170 species distributed in the mountainous regions of tropical Asia, Europe, America, and Africa, and has been used as traditional medicine for thousands of years in China, India, Japan, and Korea (Brahmachari et al. 2004; Dutt et al. 1996; Khanal et al. 2015; Liang 1984; Misra et al. 2010; Takei et al. 2001). Several members from this genus, such as *S. japonica*, *S. chirata*, *S. hookeri*, *S. macrosperma*, *S. petiolate*, and *S. calycina* have received considerable attention, as their extracts are known to possess numerous pharmacological properties (Brahmachari et al. 2004; Dutt et al. 1996; Negi et al. 2011; Kakiuchi et al. 2014; Peres et al. 2000). Extracts from these herbs have been described to possess a characteristic bitter taste, and have potential applications for febrifuge and anthelmintic, tonic, anti-periodic, cathartic, stomachic, asthma, leucorrhoea, analeptic, mitigate inflammation, and relaxing the pregnant uterus (Basnet et al. 1994; Cao et al. 2013; Jiang et al. 2015; Khanal et al. 2015; Menkovic et al. 2002; Neerja Pant and Jain 2000; Zheng et al. 2014).

Among *Swertia* species, *S. japonica*, also known as *Senburi* in Japanese, is a notable Japanese folk medicine that stands with "Dokudami" and "Gennoshoko" as one of the three major folk medicine (Kakiuchi et al. 2014). It is a biennial plant with a characteristic straight dark violet-tinged cuboidal cell morphology, reaching a height of 20-25 cm, and is native to Japan, Korean peninsula, and China. Metabolite extracts from *S. japonica* have exhibited various pharmacological activities, e.g., anti-depressant, anti-tumour, anti-malaria, febrifuge, anthelmintic, choleretic, diuretic, anti-microbial, anti-fungal, anti-inflammatory, hypoglycaemic, among others (Brahmachari et al. 2004; Dutt et al. 1996; Hiramatsu et al. 2004; Ikeshiro and Tomita 1985; Kikuchi and Kikuchi 2005; Kimura and Sumiyoshi 2011; Komatsu et al. 1969; Neerja Pant and Jain 2000; Yamahara et al. 1991).

The broad-spectrum biological activities of metabolite extracts from *S. japonica* are attributed to the presence of diverse pharmacologically active secondary metabolites belonging to different classes, such as xanthones and their derivative, iridoids, secoiridoids and their glycosides, flavonoids, and terpenoids, among others (Basnet et al. 1994; Hase et al. 1997; Ikeshiro and Tomita 1985, 1987; Kakiuchi et al. 2014; Kelly et al. 2010; Kikuchi and Kikuchi 2004, 2005; Komatsu et al. 1968; Phoboo et al. 2013; Tomimori and Komatsu 1969; Yamahara et al. 1978). Among secoiridoid class of metabolites, swertiamarin and amarogentin are the two most prominent metabolites derived from a common intermediate, sweroside, which upon glycosylation provides rich diversity of bioactive metabolites (Kimura and Sumiyoshi 2011; Kshirsagar et al. 2016; Padhan et al. 2015; Yamahara et al. 1978). Swertiamarin and amarogentin have been reported to exhibit anti-cancer (Kavimani and Manisenthlkumar 2000; Pal et al. 2012), anti-diabetic (Phoboo et al. 2013; Vaidya et al. 2013), anti-arthritic (Saravanan et al. 2014), and antileishmanial (Medda et al. 1999) activities, with swertiamarin being an effective agent against hepatitis (Wang et al. 2011). Xanthones derivatives, such as bellidifolin, swertianin, among others, are one of the major constituents of their metabolite extracts, and have been described to have unique medicinal properties (Basnet et al. 1994).

Increased demand, over exploitation, and unregulated cultivation of *S. japonica* as traditional medicine have made it an endangered species. Despite being an important medicinal plant, the genome and transcriptome data are very limited, with only 904 sequences being reported for *Swertia* genus, with only six sequences for *S. japonica* are available in the NCBI database. A lack of genetic information and molecular understanding of the biosynthesis routes for the production of major chemical constituents serve as an impediment for efforts to increase the production of metabolites of interest and sustainable cultivation of the *S. japonica*.

Here, we present the first report on the de novo transcriptome assembly and characterization of *S. japonica*. RNA sequencing for aerial tissues (plant's part above the ground) and roots (Fig. 1a), and the pre-processing of the

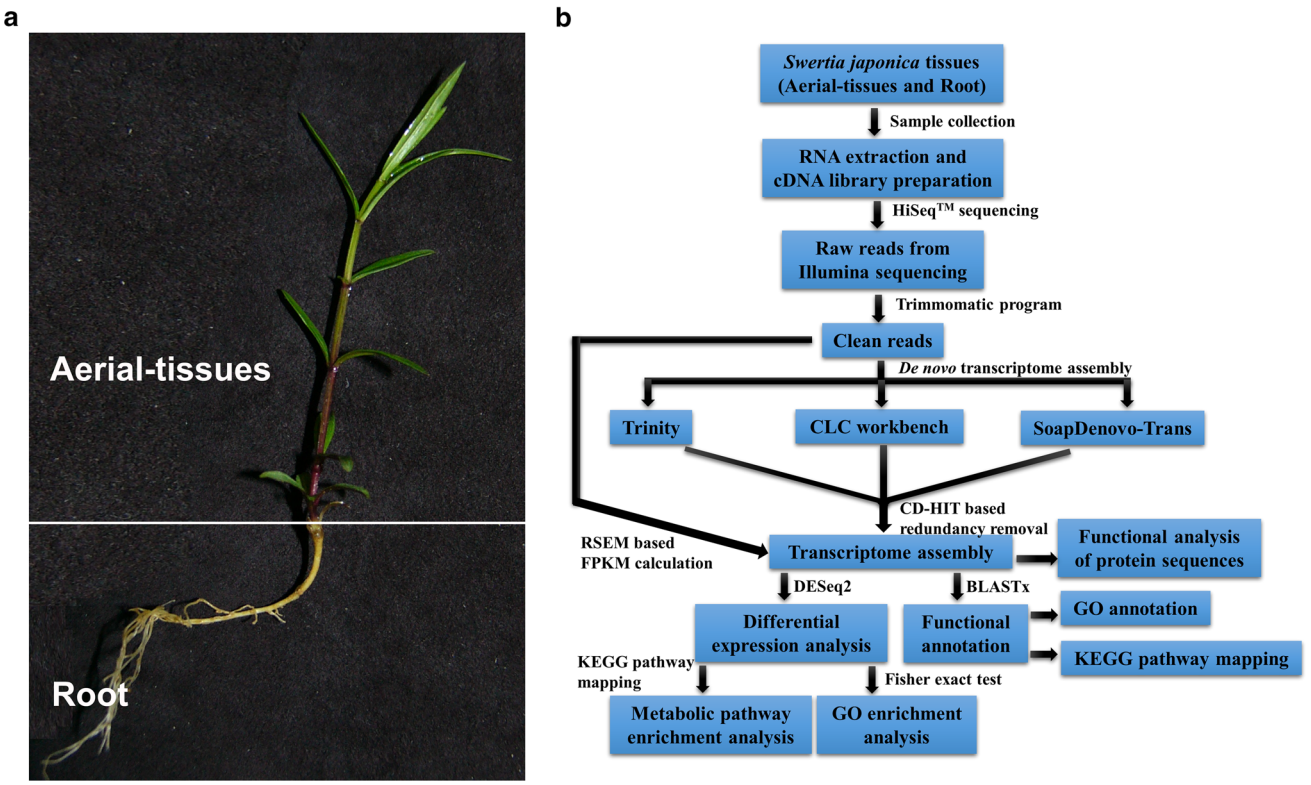


Fig. 1 *Swertia japonica* tissues and schematics of the de novo transcriptome assembly and characterization. **a** Six-month-old *S. japonica* plants were harvested, and dissected in two parts as roots and aerial tissues, which were used for the RNA extraction, cDNA library preparation, and RNA sequencing. **b** Experimental workflow for the de novo transcriptome assembly and characterization

raw reads resulted in a total of over 2.2 Gbps high-quality clean reads, which assembled de novo in 81,729 unigenes, and were further annotated and characterized. Sequence-similarity-based annotation assisted in the identification of potential unigenes associated with all the known biochemical steps for the synthesis of key metabolites. Transcript abundance analysis for the unigenes associated with the key metabolic pathways suggested higher expression levels in the aerial tissues in comparison with the roots. We also identified potential glycosyltransferases, which may play an important role for generating characteristic metabolite diversity. To the best of our knowledge, this is the first RNA-seq-based transcriptome study on any *Swertia* species. Our findings serve as a basis for future studies on this species, and will assist in the functional characterization of potential candidate genes involved in the key metabolic pathways.

Materials and methods

Plant materials, RNA isolation, and cDNA library construction

Swertia japonica plants were grown in the natural conditions at the Chiba University medicinal garden, Graduate School of Medical and Pharmaceutical Sciences, Chiba University, Chiba, Japan (located at 35°36′17.7″ N; 140°08′06.9″ E). Six-month-old plants were harvested on ice (Fig. 1a), roots and aerial tissues were dissected, and snap frozen in liquid N_2 before storing samples at −80 °C prior to RNA extraction.

The frozen samples were powdered using a Multi-beads shocker (Yasui Kikai, Japan), and were used for the subsequent RNA extraction. Total RNA for each tissue type were isolated using RNeasy Plant Mini Kit (Qiagen, USA) according to the manufacturer's instructions. The extracted RNA samples were treated with RNase-free DNase (Qiagen, USA), purified by column binding, and further purified through ethanol precipitation. RNA quality for each sample was assessed by the Agilent Bioanalyzer 2100 (Agilent Technology, USA), and samples with RNA integrity number (RIN) value above 8.0 were selected for the cDNA library preparation.

For the cDNA library construction, mRNA of each samples was isolated from the total RNA using oligo (dT) beads. The obtained mRNA was sheared into small fragments by adding fragmentation buffer, and then used as templates for the first strand cDNA synthesis using random hexamer primers. cDNA library was then prepared using the SureSelect Strand Specific RNA library kit (Agilent

Technology, USA) according to the manufacturer's instruction.

Illumina sequencing, and raw data pre-processing

Constructed cDNA libraries were sequenced using Illumina HiSeq^{TM} 2000 sequencer (Illumina Inc., USA), resulting in the paired end reads for each library with an average length of 101 bps. cDNA library preparation and sequencing were performed at the Kazusa DNA Research Institute, Chiba, Japan. The raw reads, transcriptome assembly sequences, and annotation of the unigenes discussed in this study have been deposited in the NCBI Gene Expression Omnibus (GEO) (Edgar et al. 2002), and are accessible through GEO series accession number GSE80057 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE80057).

Raw read sequences thus obtained were pre-processed using Trimmomatic program (Bolger et al. 2014) for the removal of adaptor sequences, empty reads, reads with the ambiguous 'N' bases >5 %, low quality raw reads (phred score <30), and raw reads with an average length below 50 bps. Clean reads were obtained in the form of paired clean reads, or unpaired clean reads (forward and reverse), and were used to perform de novo transcriptome assembly.

De novo transcriptome assembly

De novo transcriptome assemblies were constructed using three popular transcriptome assemblers, namely, SOAPdenovo-Trans (Xie et al. 2014), Trinity v2.1 (Grabherr et al. 2011), and CLC genomics workbench v8.0.3 (https://www.qiagenbioinformatics.com/) (CLC bio, Aaarhus, Denmark), and resulting assemblies were merged together to give rise to the final transcriptome assembly. Using SOAPdenovo-Trans, we obtained de novo transcriptome assemblies using kmer sizes as 31, 41, 51, 63, 71, and 91, resulting in the six different transcriptome assemblies. Resulting assemblies were analysed using assemblathon_2.pl scripts (Bradnam et al. 2013), and corresponding N50-values and other assembly characteristics were obtained. On the basis of transcriptome assembly characteristics, such as N50 value, number of incorporated nucleotides, mean length, and median length, de novo transcriptome assembly by SOAPdenovo-Trans program, using kmer size as 31, emerged as the best among all the six assemblies. For the Trinity and CLC genomics workbench, default kmer size and other parameters were used, and resultant assemblies were analysed using assemblathon_2.pl scripts to obtain assembly characteristics. Transcriptome assemblies thus obtained using Trinity, CLC genomics workbench, and SOAPdenovo-Trans (kmer size-31) were pooled together into one merged assembly, and were processed using CD-HIT-EST v4.6 program (built on Mar 5, 2015) (Fu et al. 2012; Li and Godzik 2006) with the parameters used as "-c 0.95 –n 8" for the removal of sequence redundancy. Sequences with a length less than 200 bps were dropped, and de novo transcriptome assembly for *S. japonica* thus obtained was used for the subsequent annotation and characterization.

For the transcripts expression analysis, clean paired end reads for each tissues were aligned to the *S. japonica* transcriptome assembly using Bowtie2.0 program (Langmead et al. 2009), and transcripts abundance was estimated using RSEM program (Li and Dewey 2011). To calculate unigene expression, we used Fragments per Kilobase exon per Million mapped fragments (FPKM) method. Differential expression analysis was performed by DESeq2 program (Love et al. 2014), using input as count data obtained from RSEM program for unigenes from each tissues. In the differential transcripts expression analysis, we used a negative binomial test implemented in the DESeq2. GC content and basic statistics values for unigenes were calculated as described previously (Fukushima et al. 2015).

Functional annotation and classification of unigenes

We performed Blastx program (Altschul et al. 1990) based homology search for the *S. japonica* transcriptome assembly against the NCBI-non-redundant (nr) protein database (http://www.ncbi.nlm.nih.gov; formatted on Oct, 2015) using an *E* value cutoff as <10-5, and number of maximum allowed hit for each query was set as 20. The aligning sequences from NCBI with the smallest *E* value were used to annotate the unigenes. InterProScan (Jones et al. 2014) for the *S. japonica* transcriptome assembly was performed using Blast2GO program (Conesa and Gotz 2008) with the default parameters against the InterPro protein signature databases. Subsequently, using Blast2GO program, and on the basis of blastx and InterProScan results for the *S. japonica* transcriptome assembly, gene ontology (GO) (Gene Ontology 2015), EC number, and KEGG information (Kanehisa et al. 2014) to the unigenes were assigned. Visualization of the *E* value distribution, sequence similarity distribution, top-hit species distribution, and GO functional classification for *S. japonica* transcriptome assembly was performed by the Blast2GO program.

Simple sequence repeat (SSR) detection

Swertia japonica de novo transcriptome assembly was searched to determine the composition, frequency, and distribution of the Simple Sequence Repeats (SSRs) using the microsatellite identification tool (MISA) (http://pgrc.ipk-gatersleben.de/misa/) (Thiel et al. 2003). The search parameters for the maximum motif length group were set to

recognize hexamers for each SSRs length-based category to have at least ten repeats.

Gene ontology and KEGG pathway enrichment analysis

Differentially expressed unigenes (both up- and down-regulated in the aerial tissues with respect to the roots) were used as a test set, and compared against the whole transcriptome background of *S. japonica* as a reference set using Fisher exact test with a p value cutoff as 0.05 (hypergeometric test with Benjamini and Hochberg false discovery rate correction) as described previously (Rai et al. 2016). For the KEGG pathways enrichment analysis, differentially expressed unigenes (both up- and down-regulated) with assigned KEGG pathways were used as a test set against all unigenes assigned with KEGG pathways in the *S. japonica* transcriptome assembly as a reference set using Fisher exact test with p value cutoff as 0.05.

Phylogenetic analysis

Unigenes from the *S. japonica* transcriptome assembly with the sequence length above 500 bps, annotated as UDP-glycosyltransferases, and expression abundance >1 FPKM value in either of aerial tissues or of roots were used for the phylogenetic analysis together with 128 UDP-glycosyltransferase coding genes from the *Arabidopsis* genome (obtained and classified from http://www.p450.kvl.dk/UGT.shtml) (Paquette et al. 2009). Nucleotide sequences were aligned using MUSCLE program, evolutionary distance was computated using the Jones–Taylor–Thornton (JTT) method, and a Neighbor-Joining (NJ) tree was constructed with the bootstrap values obtained after 1000 replications using the MEGA6 software (Tamura et al. 2013).

Results and discussion

Sample preparation and Illumina sequencing

Traditionally, the whole plant or aerial tissues of *S. japonica* is used for the extraction of therapeutic metabolites. However, several studies have also identified biologically active metabolites such as bellidifolin, methylbellidifolin, swertianolin, amarogentin, amaroswerin, among others from the roots of *S. japonica* (Brahmachari et al. 2004; Negi et al. 2011).

To elucidate transcriptome landscape of *S. japonica,* and to identify candidate genes involved in the biosynthesis of therapeutic metabolites, total RNA from the aerial tissues and roots were isolated (Fig. 1a). Total RNA from individual tissues with RIN (RNA integrity number) value above 8.0 were used for mRNA preparation, fragmentation, and library preparation. cDNA libraries thus prepared were sequenced on the Illumina HiSeq^{TM} 2000, resulting in 1,193,840,800 and 950,603,200 bps raw reads for the aerial tissues and roots, respectively (Table S1). After the removal of adaptor sequences, low quality and ambiguous bases, and reads that were shorter than 50 base pairs (bps) using the Trimmomatic software (Bolger et al. 2014), about 2.2 Gbps clean reads were used for the de novo transcriptome assembly. The short reads with mean length of 101 bps showed mean quality scores of 36.5 and 38 % for the aerial tissues and roots, respectively, indicating that our RNA sequencing was adequate for the de novo transcriptome assembly. The study overview is shown in Fig. 1b.

De novo assembly optimization for the transcriptome of *S. japonica*

An important step for successful transcriptome-based studies requires building of a high-quality assembly. The completeness and quality of an assembled transcriptome are affected by the choice of the assembler program and different parameters in use, in particular the kmer size (Duan et al. 2012; Gruenheit et al. 2012). Comparative studies on the different assemblers using similar assembly parameters have shown quite different output set of reconstructed sequences, with no single assembler consistently outperforming others when tested using different input data sets (Steijger et al. 2013). Recent studies have proposed use of multiple assemblers to generate de novo transcriptome assembly (Kumar and Blaxter 2010; Nakasugi et al. 2014; Visser et al. 2015). To generate a more complete *S. japonica* transcriptome assembly by maximizing diversity of the de novo assembled transcripts, we used three popular assemblers, namely, SOAPdenovo-Trans (Xie et al. 2014), Trinity (Grabherr et al. 2011), and CLC genomic workbench (https://www.qiagenbioinformatics.com/).

Using SOAPdenovo-Trans program, we performed de novo transcriptome assemblies using kmer size as 31, 41, 51, 63, 71, and 91, resulting in six different sets of transcriptome assemblies. Several output parameters, such as total number of contigs, total bases assembled, N50 length, longest contig length, average contig length, and median contig length were analysed to compare transcriptome assemblies generated by SOAPdenovo-Trans using different kmer values (Table 1). Assembly from kmer size 31 emerged as the best among all six, with 38.76e6 nucleotides assembled in 46,384 contigs with N50 value as 1422 bps. De novo transcriptome assemblies using Trinity and CLC were performed using their default kmer sizes. Trinity-based transcriptome assembly incorporated

Table 1 Summary of sequence assembly

Assembler	Kmer	No. of contigs	Average length	Median length	Max length	N50	Total (bps)
Trinity	25	124,912	864.57	566	9347	1319	107.9e6
CLC-work bench	25	50,208	802	512	11,755	1300	40.25e6
SOAPdenovo-Trans	31	46,384	838	550	12,286	1422	38.76e6
	41	45,359	833	550	11,273	1423	37.68e6
	51	44,927	794	513	9888	1356	35.57e6
	63	43,082	734	451	7581	1240	31.5e6
	71	37,442	713	425	7310	1151	26.55e6
	91	4876	637	395	4609	770	3.08e6
Trinity + CLC + SOAPdenovo-Trans (kmer-31)	NA	81,648	884	617	12,286	1452	72.06e6

Summary of de novo transcriptome assembly statistics of *Swertia japonica* resulting from the three different assemblers and their combination

107.9e6 bases in 124,912 contigs with N50 value as 1319 bps, while CLC assembled 40.25e6 nucleotides into 50,208 contigs with N50 value as 1300 bps (Table 1). Transcriptome assemblies thus obtained using Trinity, CLC genomics workbench, and SOAPdenovo-Trans (kmer 31) were pooled together, and redundancies were removed using CD-HIT-EST program (Fu et al. 2012; Li and Godzik 2006), resulting in 72.06e6 nucleotides being assembled in 81,729 unigenes with N50 value as 1452 bps (Table 1), and length and %GC distribution, as shown in Fig. 2a and b, respectively. Resulting transcriptome assembly showed gain in the N50 value, average length, and median length of the sequences compared with individual assemblers, and was used for further annotation and characterization.

Functional annotation and classification of *S. japonica* transcriptome

Sequence homology-based functional annotation is usually the first step for studying the role and biological functions of gene products and its characterization. Unigenes from *S. japonica* transcriptome assembly were subjected to Blastx program (Altschul et al. 1990)-based homology search against NCBI non-redundant (nr) protein database (http://www.ncbi.nlm.nih.gov; formatted on October 2015) using a cut-off E value $<10^{-5}$ with maximum number of hits allowed for each query as 20, and the best aligning results were used to annotate the unigenes. Overall, 46,963 (57.5 %) unigenes transcriptome received at least one hit against NCBI-nr database (Table S2). Over 80 % of the aligned sequences showed significant homology against NCBI-nr database with 24,729 unigenes (52.6 % of unigenes with a blast hit) showing more than 85 % of sequence similarity against corresponding top hit (Fig. 3a, b).

InterPro is a database of protein functional sites, families, and domains, and InterProScan (IPS) allows functional characterization of a transcriptome assembly by its comparison with identifiable features found in known proteins (Jones et al. 2014; Mulder and Apweiler 2007). To perform functional characterization of unigenes based on protein functional sites and protein domains, we performed IPS using Blast2GO program (Conesa and Gotz 2008; Conesa et al. 2005), resulting in 46,332 unigenes with an IPS hit (Table S3). IPS family distribution analysis showed cytochrome P450 (IPR001128), UDP-glucosyltransferase (IPR002213), and glycoside hydrolase (IPR001360) being top three protein families being enriched (Fig. S1a). IPS domain distribution, IPS repeat distribution, and IPS IDs distribution (based on pfam and panther database) are summarized in Fig. S1b–d. Blast2GO program was used to perform annotation based on Blastx and IPS results, and is summarized in Fig. 3c, while the highest scoring pair (HSP) distributions over the corresponding hits and query sequences are shown in Fig. 3d and e, respectively. Top-hit species distribution analysis based on the Blastx results showed over 70 % of the unigenes with a blast hit sharing high sequence similarity with sequences from *Coffea canephora*, *Nicotiana sylvestris*, *Nicotiana tomentosiformis*, *Vitis vinifera*, *Erythranthe guttata*, and *Solanum tuberosum* (Fig. 3f).

Using Blast2GO program, and based on the NCBI-nr database annotation and IPS results, gene ontology (GO) (Gene Ontology 2015) categories were assigned to 45,898 unigenes being grouped into 98 GO terms, classified into three major GO categories: biological process (45), molecular function (26), and cellular component (27). Top 20 GO terms (at GO level 3) and GO-level distribution for the biological process, molecular function, and cellular component categories are shown in Fig. 4a and b, respectively. Metabolic processes, such as organic substance metabolic process, cellular metabolic process, primary metabolic process, single-organism metabolic process, nitrogen compound metabolic process, and secondary metabolic process, were among top ten highly abundant

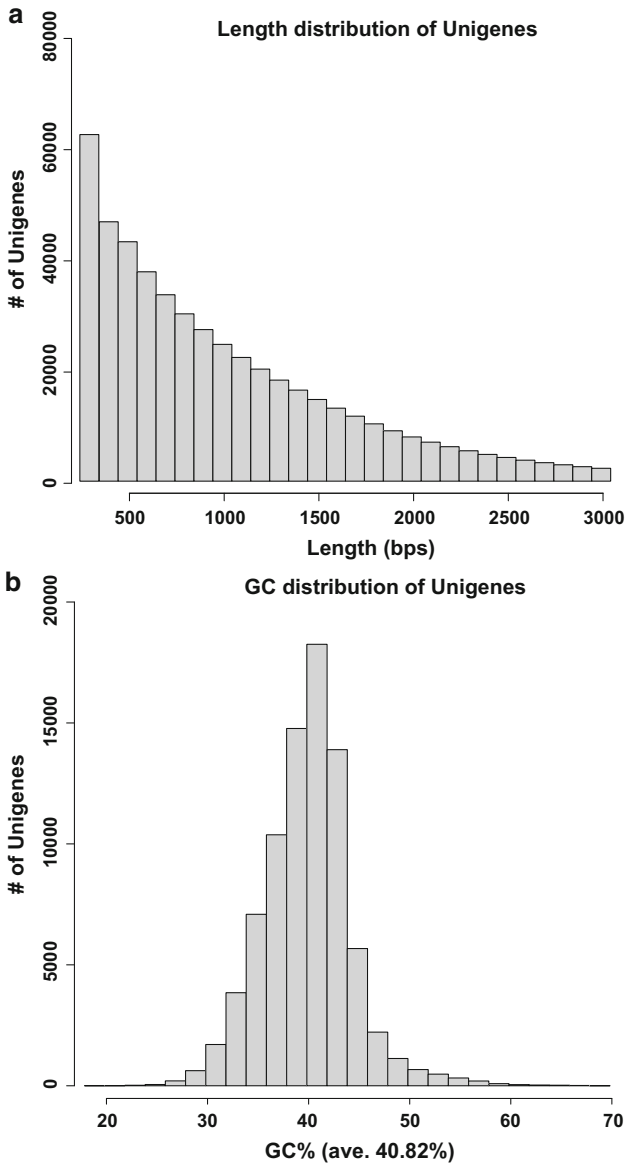


Fig. 2 Overview of the de novo transcriptome assembly of *S. japonica*. **a** Length distribution of the unigenes of the *S. japonica* transcriptome assembly. **b** Percentage GC distribution of the unigenes of the *S. japonica* transcriptome assembly

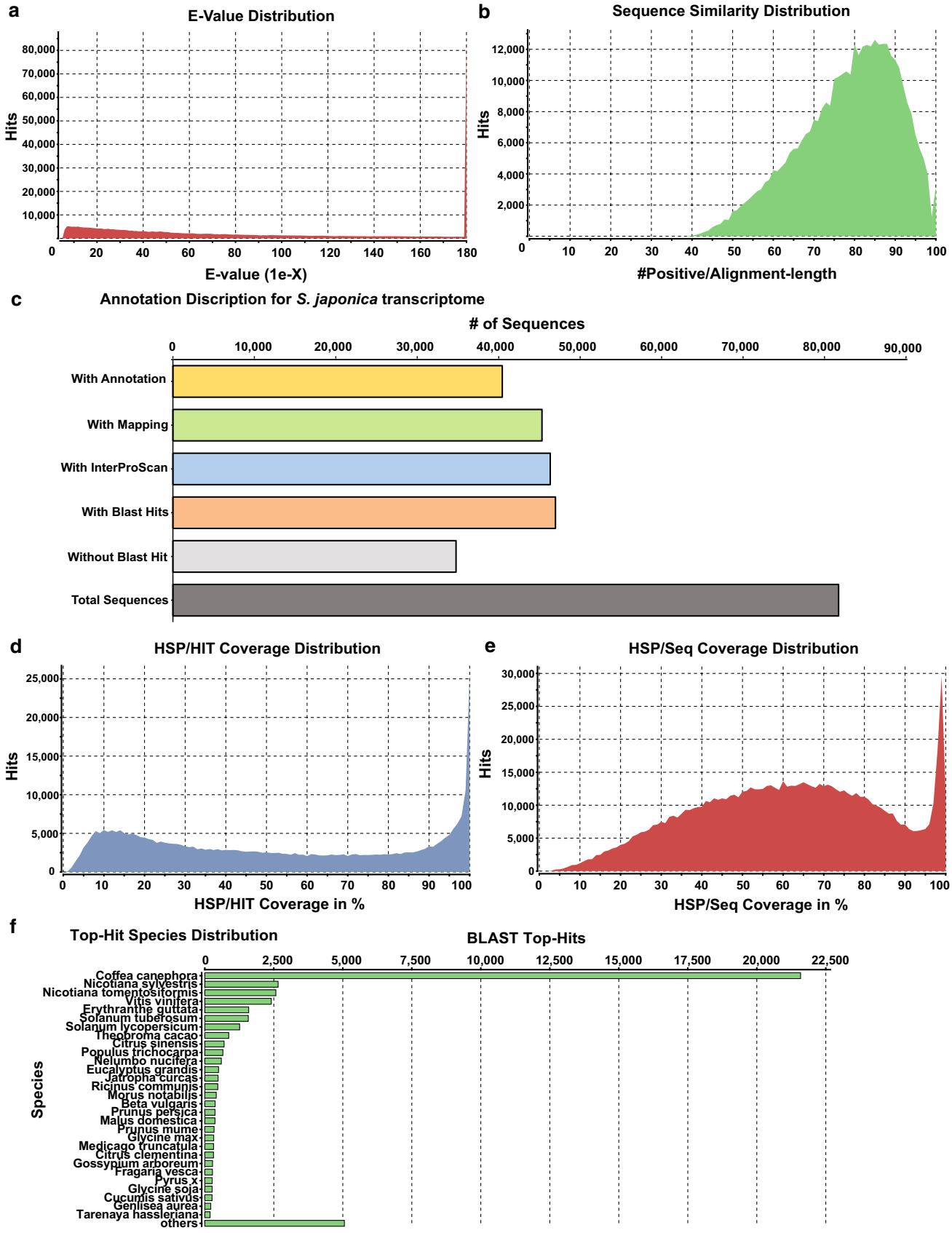
GO terms enriched under biological process category. For the molecular function category, GO terms corresponding to the organic cyclic compound binding, heterocyclic compound binding, transferase activity, hydrolase activity, oxidoreductase activity, and transmembrane transporter were assigned to majority of the unigenes. For the cellular component category, most unigenes were assigned to GO terms corresponding to the cell part, membrane-bounded organelle, organelle part, membrane part, and protein complex. First eight GO terms from biological process category being associated with various metabolic processes further indicates that *S. japonica* transcriptome assembly and annotation successfully captured the on-going metabolic processes associated with the key metabolites biosynthesis. Unigenes with assigned functional annotations, IPS annotation, top-hit accession id, description of annotation, GO terms, and EC numbers are provided in Table S4.

Functional classification of *S. japonica* transcriptome by KEGG pathways

The KEGG pathway database (Kanehisa et al. 2014) contains information on the intra-cellular components and their interactions within a metabolic network, and allows functional annotation of gene products and their assignment to various biological pathways. Such pathway-based annotation for a transcriptome assembly provides an overview of active metabolic processes within an organism and an additional layer of information to the unigenes. Blast2GO program was used for assigning KEGG pathway annotations to the unigenes, resulting in 14,550 unigenes being grouped in 144 KEGG pathways (Table S5). Top 50 KEGG pathways based on the number of assigned unigenes in *S. japonica* transcriptome assembly are shown in Fig. 5. Based on the number of unigenes assigned, purine metabolism, thiamine metabolism, aminobenzoate degradation, starch and sucrose metabolism, and phenylpropanoid biosynthesis were the first five metabolic pathways. Transcripts associated with metabolic pathways, such as flavonoid biosynthesis (137 Unigenes), flavone and flavonol biosynthesis (30 Unigenes), isoflavonoid biosynthesis (15 Unigenes), anthocyanin biosynthesis (27 Unigenes), terpenoid backbone biosynthesis (124 Unigenes), monoterpenoid biosynthesis (55 Unigenes), and diterpenoid biosynthesis (52 Unigenes), which are key for the biosynthesis of therapeutic metabolites, such as swertiamarin, swertianin, mangiferin, amarogentin, among others, were also identified (Table S5).

Identification of simple sequence repeats (SSRs) of *S. japonica* using its de novo assembled transcriptome

SSRs are simple motifs of 1–6 nucleotides, repeated over two to a few dozen times at a locus and ubiquitously distributed within genome of a species (Hancock 1996). Several studies have implicated SSRs in affecting gene expression, and have suggested polymorphism of SSRs tracts as key event in the evolution of genetic regulation (Jarvis et al. 2008; Varshney et al. 2002, 2005). SSRs are the indel mutational hotspots in the genome, and serve as an important marker for the determination of functional genetic variations, including paternity determination, population genetics studies, genetic diversity assessments, and for the development of genetic map (Senthilvel et al. 2008).


a
E-Value Distribution
Hits
E-value (1e-X)
b
Sequence Similarity Distribution
Hits
#Positive/Alignment-length
c
Annotation Discription for S. japonica transcriptome
# of Sequences
With Annotation
With Mapping
With InterProScan
With Blast Hits
Without Blast Hit
Total Sequences
d
HSP/HIT Coverage Distribution
Hits
HSP/HIT Coverage in %
e
HSP/Seq Coverage Distribution
Hits
HSP/Seq Coverage in %
f
Top-Hit Species Distribution
BLAST Top-Hits
Species
Coffea canephora
Nicotiana sylvestris
Nicotiana tomentosiformis
Vitis vinifera
Erythranthe guttata
Solanum tuberosum
Solanum lycopersicum
Theobroma cacao
Citrus sinensis
Populus trichocarpa
Nelumbo nucifera
Eucalyptus grandis
Jatropha curcas
Ricinus communis
Morus notabilis
Beta vulgaris
Prunus persica
Malus domestica
Prunus mume
Glycine max
Medicago truncatula
Citrus clementina
Gossypium arboreum
Fragaria vesca
Pyrus x
Glycine soja
Cucumis sativus
Genlisea aurea
Tarenaya hassleriana
others

◀**Fig. 3** Characterization of the *S. japonica* transcriptome assembly using blastx program against NCBI-non-redundant (nr) protein database search. **a** *E* value distribution plot associated with the blast hit unigenes of the *S. japonica* transcriptome assembly with *E* value cutoff $<10^{-5}$. **b** Sequence similarity distribution plot of the *S. japonica* assembled transcriptome using top-hits blastx search data. **c** Data distribution based on blastx and InterProScan results, and subsequent mapping and annotation using Blast2GO program. **d** Highest scoring pairs (HSP) distribution for the top-hit sequences of the unigenes. **e** Highest scoring pairs (HSP) distribution for the unigenes of the *S. japonica* transcriptome assembly. **f** Species distribution plot for the top-hit sequences associated with unigenes of *S. japonica* transcriptome assembly

To identify SSRs for *S. japonica*, we used de novo transcriptome assembly, and searched all unigenes for mono- to hexa-nucleotide motifs with a minimum of ten repetitions using the MISA software (Thiel et al. 2003). Our analysis detected a total of 17,992 SSRs in 14,702 unigenes, of which 2615 unigenes had more than one SSRs (Table 2). Among repeat type classes, mono-nucleotide represented the largest fraction (67 %) of all identified SSRs, followed by tri-nucleotide (18 %) and di-nucleotides (11.6 %). Although only a small fraction of tetra-nucleotide (219), penta-nucleotide (116), and hexa-nucleotide (166) were identified in the *S. japonica* transcriptome, their numbers were significant. All identified SSRs are shown in Table S6. SSRs identified in this study may provide potential genetic markers, which will be useful to conduct population genetics and comparative genomics studies across various *Swertia* species.

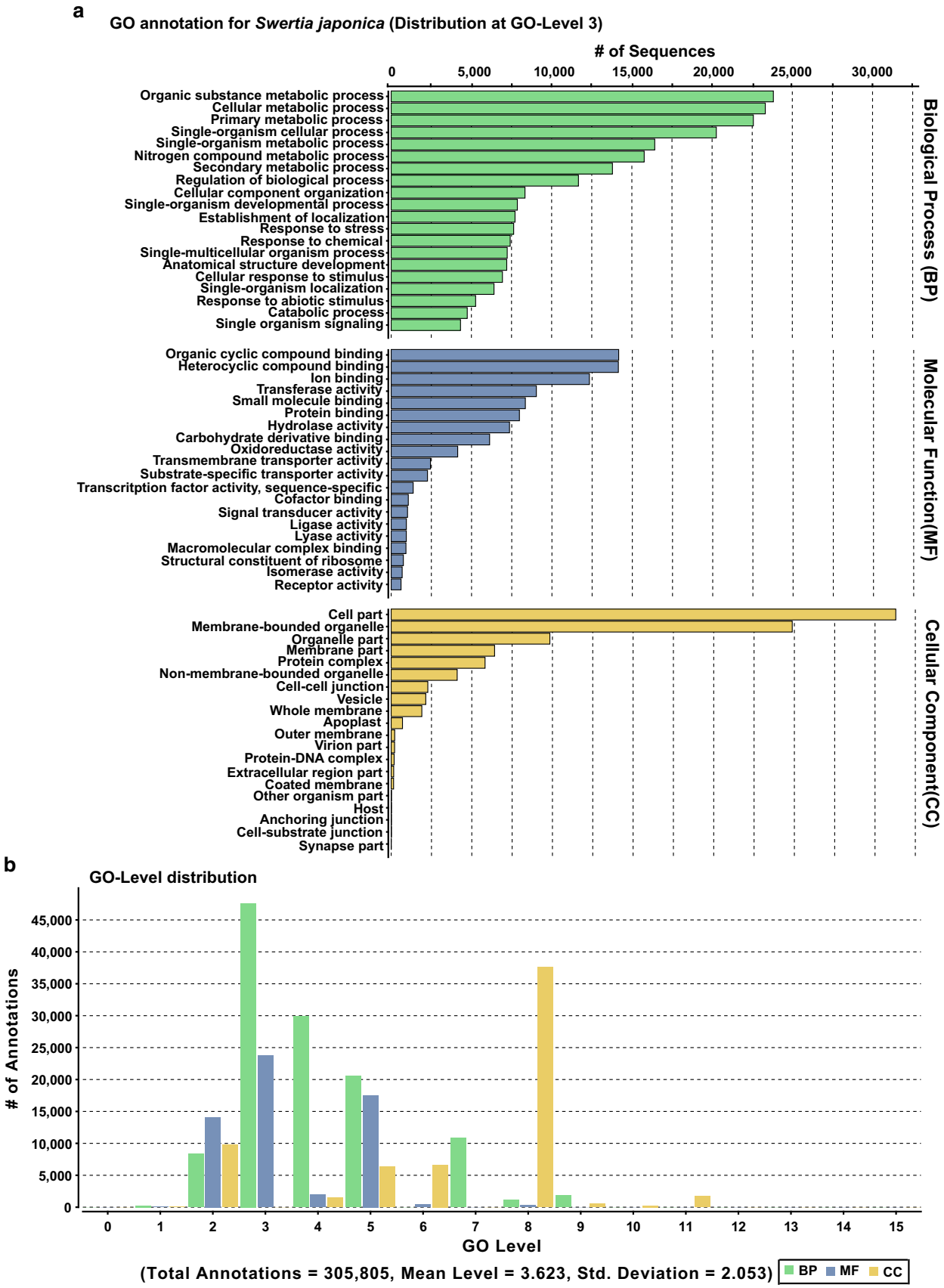
Functional classification and comparison of differentially expressed unigenes between aerial tissues and roots of *S. japonica*

Transcriptome profiling and differential expression analysis provide a global picture of different active processes across various conditions. To analyse transcript expression levels, clean paired short reads from aerial tissues and roots were aligned to the de novo transcriptome assembly using Bowtie2.0 program (Langmead et al. 2009), and FPKM values and count-read data for transcripts expression for each tissue type were obtained using RSEM program (Li and Dewey 2011). Distribution of number of sequences based on FPKM values for the aerial tissues and roots is shown in Fig. S2. Overall, 71,025 (87 %) unigenes showed FPKM values over 1 for at least one sample, while 42,349 unigenes had FPKM value over 1 in both aerial tissues and root tissues (Table S7). We used count-read data for the tissues, and performed differential expression analysis using DESeq2 program (Love et al. 2014), resulting into 1695 unigenes being up-regulated and 730 unigenes being down-regulated [false discovery rate (FDR) <0.05] in the aerial tissues with respect to the roots of *S. japonica*.

To identify functional groups enriched within these differentially expressed unigenes, we performed GO enrichment analysis through Fisher exact test with a *p* value cutoff <0.05. Up-regulated or down-regulated set of unigenes in the aerial tissues were used with respect to the roots as test set, while transcriptome assembly was used as a reference set (Fig. 6a, b). GO enrichment analysis identified 59 GO categories being significantly enriched in the up-regulated unigenes list, with mostly being associated with different metabolic processes (17 processes) or developmental processes (12 processes). GO terms corresponding to plastid, thylakoid, photosynthesis, generation of precursor metabolites and energy, cytoplasmic part, lipid metabolic process, and single-organism metabolic processes were within the top ten over-enriched processes in the up-regulated unigenes. Moreover, metabolic process-associated GO terms, including biosynthetic process, metabolic process, secondary metabolic process, cellular aromatic compound metabolic process, carbohydrate metabolic process, nitrogen compound metabolic process, among others, were significantly enriched. Several developmental process-associated GO terms, including flower development, shoot system development, developmental process involved in the reproduction, reproductive shoot system development, reproductive structure development, among others, were also significantly enriched in up-regulated unigene list. For the down-regulated unigenes, we identified 11 GO categories associated with processes, such as catalytic activity, extracellular space, oxygen binding, extracellular region, among others, to be significantly enriched. GO enrichment analysis for differentially expressed unigenes revealed active metabolic processes being enriched mainly in the aerial tissues, which corroborates with the traditional use of aerial tissues in different medicinal applications.

KEGG pathway enrichment analysis for the differentially expressed unigenes in the aerial tissues with respect to the roots of *S. japonica*

KEGG pathway enrichment analysis using differentially expressed unigenes provides an overview of active metabolic processes under test conditions or tissues as compared with the entire background of all KEGG pathway-associated unigenes within a transcriptome assembly. Differentially expressed unigenes in aerial tissues with respect to roots as a test set and transcriptome as a reference set were used to perform KEGG pathway enrichment analysis by Fisher exact test with a *p* value cutoff <0.05. We identified a total of 40 KEGG pathways to be statistically significant and enriched in either of the aerial tissues or of the roots of *S. japonica* (Fig. 7). Anthocyanin biosynthesis, lipopolysaccharides biosynthesis, isoflavonoid


a
GO annotation for Swertia japonica (Distribution at GO-Level 3)
# of Sequences
0
5,000
10,000
15,000
20,000
25,000
30,000
Organic substance metabolic process
Cellular metabolic process
Primary metabolic process
Single-organism cellular process
Single-organism metabolic process
Nitrogen compound metabolic process
Secondary metabolic process
Regulation of biological process
Cellular component organization
Single-organism developmental process
Establishment of localization
Response to stress
Response to chemical
Single-multicellular organism process
Anatomical structure development
Cellular response to stimulus
Single-organism localization
Response to abiotic stimulus
Catabolic process
Single organism signaling
Biological Process (BP)
Organic cyclic compound binding
Heterocyclic compound binding
Ion binding
Transferase activity
Small molecule binding
Protein binding
Hydrolase activity
Carbohydrate derivative binding
Oxidoreductase activity
Transmembrane transporter activity
Substrate-specific transporter activity
Transcritption factor activity, sequence-specific
Cofactor binding
Signal transducer activity
Ligase activity
Lyase activity
Macromolecular complex binding
Structural constituent of ribosome
Isomerase activity
Receptor activity
Molecular Function(MF)
Cell part
Membrane-bounded organelle
Organelle part
Membrane part
Protein complex
Non-membrane-bounded organelle
Cell-cell junction
Vesicle
Whole membrane
Apoplast
Outer membrane
Virion part
Protein-DNA complex
Extracellular region part
Coated membrane
Other organism part
Host
Anchoring junction
Cell-substrate junction
Synapse part
Cellular Component(CC)
b
GO-Level distribution
# of Annotations
0
5,000
10,000
15,000
20,000
25,000
30,000
35,000
40,000
45,000
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
GO Level
(Total Annotations = 305,805, Mean Level = 3.623, Std. Deviation = 2.053)
BP
MF
CC

◄**Fig. 4** Gene ontology (GO) distribution for the assembled unigenes of *S. japonica*. **a** Top 20 GO terms based on the number of unigenes assigned for three major GO categories, namely, biological process, molecular function, and cellular component. GO terms for all unigenes were assigned based on the blastx and InterProScan results using Blast2GO program at GO annotation level 3. **b** GO-level distribution in the *S. japonica* assembly for the three major GO categories

biosynthesis, and indole alkaloid biosynthesis were the top four most statistically significantly enriched KEGG pathways in the aerial tissues and roots. The majority of unigenes associated with the enriched KEGG pathways were up-regulated in the aerial tissues with respect to the roots of *S. japonica* (Fig. S3). Key metabolic pathways, such as anthocyanin biosynthesis, phenylpropanoid biosynthesis, isoflavonoid biosynthesis, terpenoid backbone biosynthesis, indole alkaloid biosynthesis, and monoterpenoid biosynthesis, which provide precursors for the biosynthesis of swertiamarin, its sugar conjugation, and other therapeutic metabolites, were all up-regulated in the aerial tissues of *S. japonica*. KEGG pathway enrichment analysis and GO enrichment analysis for differentially expressed unigenes showed metabolic pathways involved in the synthesis of therapeutic metabolites being enriched in the aerial tissues of *S. japonica*.

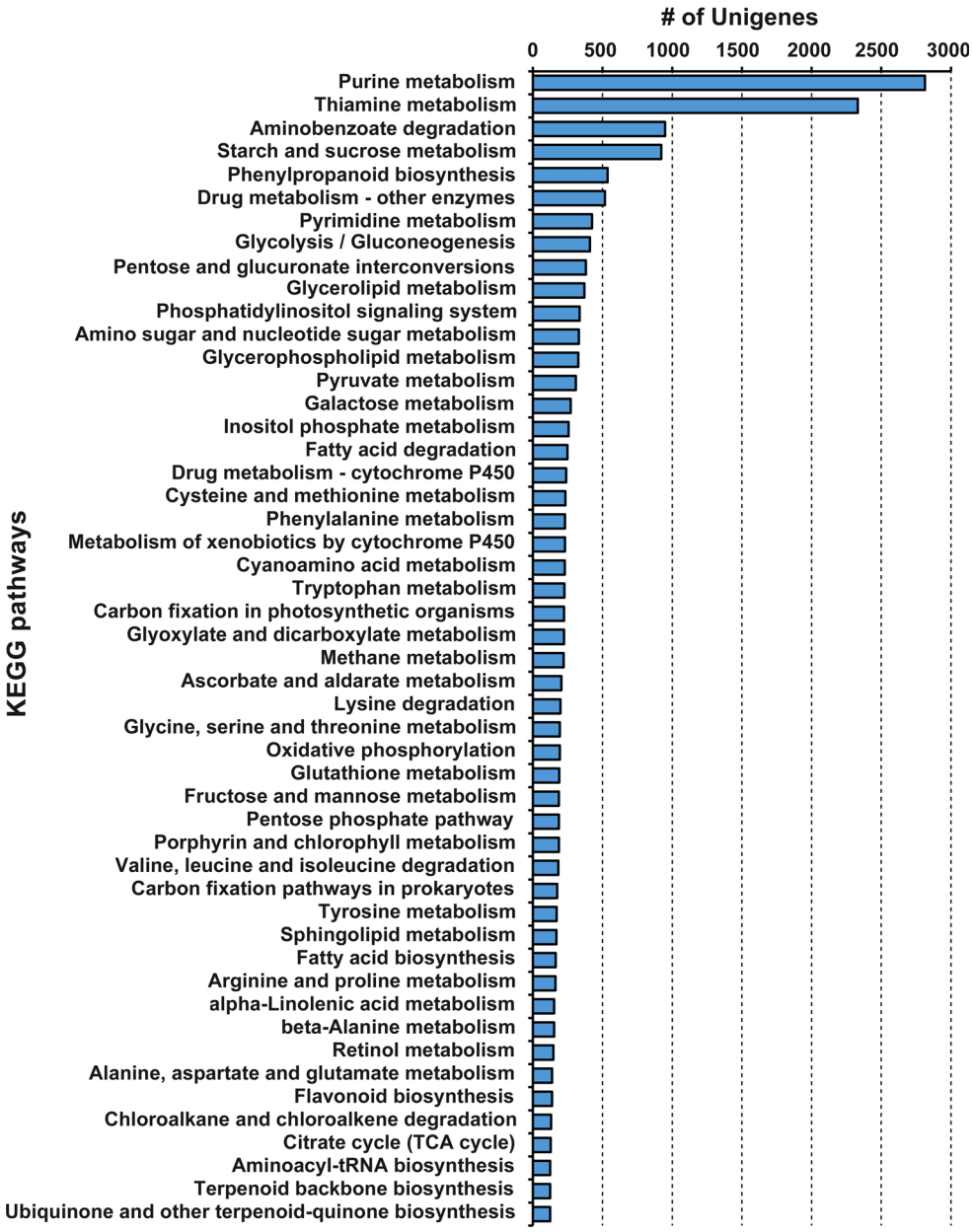


Fig. 5 Functional annotation of the *S. japonica* transcriptome assembly by KEGG pathways database. First 50 KEGG pathways on the basis of number of assigned unigenes of the *S. japonica* are represented here

Table 2 Statistics of SSR detected in *Swertia japonica*

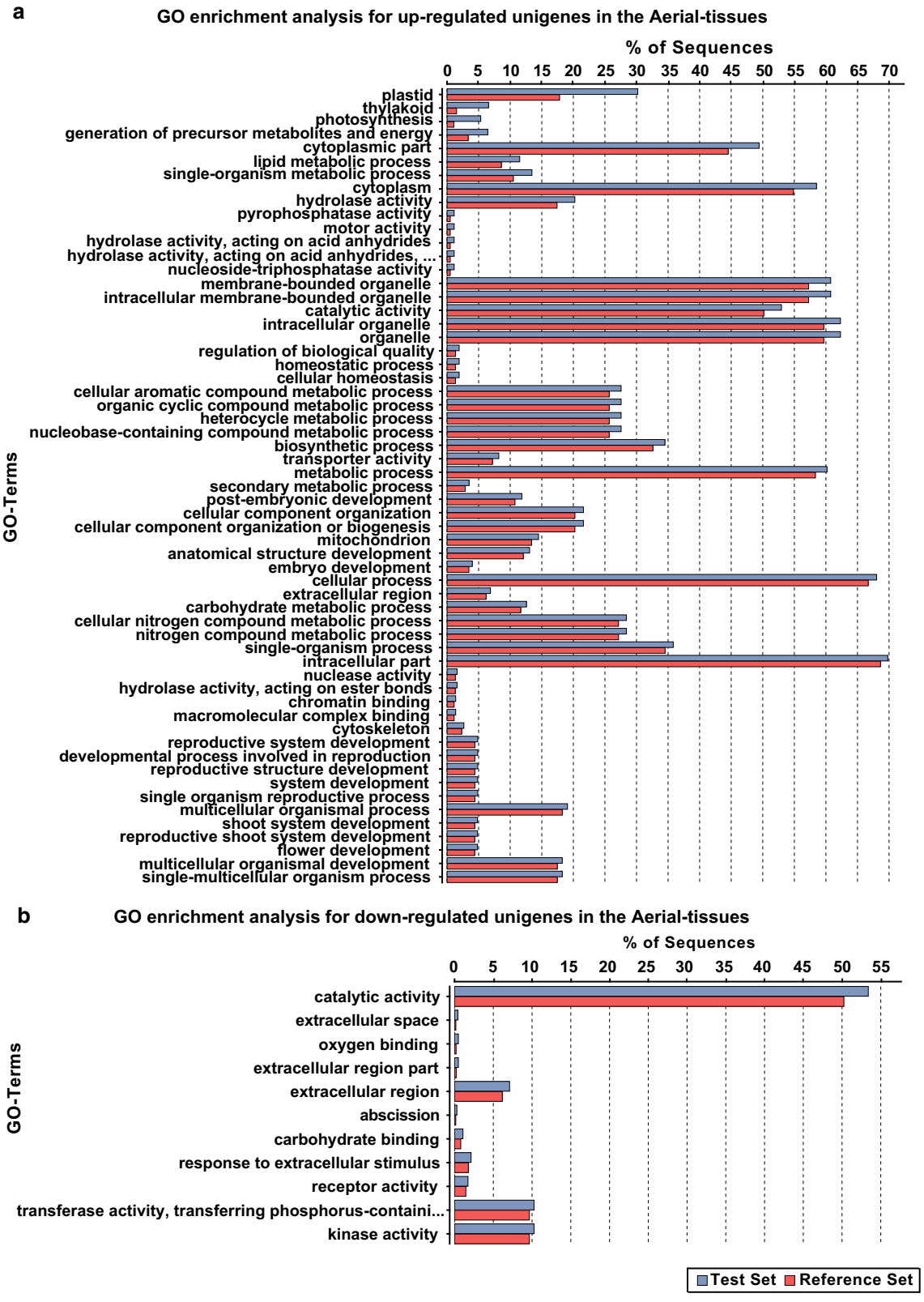
Results of SSR searches	
Total number of sequences examined	81,729
Total size of examined sequences (bp)	72,197,224
Total number of identified SSRs	17,992
Number of SSR containing sequences	14,702
Number of sequences containing more than 1 SSR	2615
Number of SSRs present in compound formation	1003
Distribution to different repeat type classes	
Mono-nucleotides	12,166
Di-nucleotides	2095
Tri-nucleotides	3230
Tetra-nucleotides	219
Penta-nucleotides	116
Hexa-nucleotides	166

Gene expression analysis of unigenes associated with swertiamarin, amarogentin, and mangiferin biosynthetic pathways

The major phytochemicals of the bitter tasting plant *S. japonica* includes swertiamarin, amarogentin, and their derivative along with other bioactive metabolite classes, such as xanthones (Brahmachari et al. 2004; Jagmohan S. Negi et al. 2011). Swertiamarin and amarogentin biosynthesis take the classical mevalonate (MVA) pathway or methylerythritol phosphate (MEP) pathway route of terpene biosynthesis that leads to the synthesis of isopentenyl diphosphate (IPPI) and dimethylallyl diphosphate (DMAPP). These intermediates serve as substrates for geranyl diphosphate synthase leading to the synthesis of geranyl diphosphate (Fig. 8). Geranyl diphosphate then undergoes a cascade of chemical conversions along with the formation of metabolite intermediates, such as geraniol, iridodial, iridotrial, loganic acid, secologanin, among others, through secoiridoid biosynthesis pathway (Fig. 9). Sweroside, a common intermediate for the biosynthesis of swertiamarin and amarogentin, is synthesized from secologanin through uncharacterized set of biochemical reactions. Hydroxylation of sweroside leads to the biosynthesis of swertiamarin, while amarogentin is the biphenylcarboxylic acid derivative of sweroside. In *S. japonica*, biphenylcarboxylic acid precursor is synthesized from m-hydroxybenzoyl-CoA by a polyketide-type pathway from phenylalanine via cinnamic acid and benzoic acid.

Based on our Blastx results and annotation of transcriptome assembly, we identified 15 and 17 unigenes with a minimum length of 800 bps, and over 85 % sequence similarity with the top-hit sequence, annotated as different enzyme coding genes from the MEP and MVA pathways, respectively. Among unigenes annotated in the enzyme coding genes associated with MEP pathway, one unigene each was annotated as isopentenyl diphosphate isomerase (IPPI), (E)-4-hydroxy-3-methylbut-2-enyl diphosphate reductase (ISPH), (E)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase (ISPG), 4-(cytidine-5′-diphospho)-2-C-methyl-D-erythritol kinase (ISPE), and 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR), while two or more candidate unigenes were identified for 1-deoxy-D-xylulose 5-phosphate synthase (DXS), 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (ISPD), and 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (ISPF) (Data set. S1). Moreover, their expression levels were significantly up-regulated in the aerial tissues (Fig. 8). For the MVA pathways, one unigene each was annotated as mevalonate kinase (MVK), phosphomevalonate kinase (PMK), and isopentenyl diphosphate isomerase (IPPI), while two or more unigenes were annotated as acetoacetyl-CoA-thiolase (AACT), 3-hydroxy-3-methylglutaryl-CoA synthase (HMGS), 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR), and mevalonate diphosphate decarboxylase (MVDD) (Data set. S2). Expression levels of unigenes annotated as MVK, MVDD, and PMK were down-regulated, while unigenes annotated as IPPI, AACT, HMGS, and HMGR were up-regulated in the aerial tissues (Fig. 8). Therefore, unlike MEP pathways associated unigenes, the expression of unigenes associated with MVA pathways showed mix expression trend between the aerial tissues and the roots.

All the ten enzyme coding genes known to be involved and reported in the secoiridoid biosynthesis pathways (Miettinen et al. 2014) were identified based on the sequence similarity search. In total, 32 unigenes were annotated as different enzyme coding genes with each unigenes having a minimum sequence length of 800 bps and above 85 % of sequence similarity against its top-hit sequence. Among these, one unigene each were annotated as iridoid oxidase (IO) and loganic acid O-methyltransferase (LAMT), while more than one unigenes were assigned to the remaining enzyme coding genes from secoiridoid biosynthesis pathways (Data set. S3). Overall, expression of unigenes from the secoiridoid pathways was significantly up-regulated in the aerial tissues (Fig. 9). Unigenes, such as unigene 63350 [geranyl diphosphate synthase (GPPS)], unigene 51905 [geraniol-10-hydroxylase (G10H)], unigene 57164 [8-hydroxygeraniol oxidoreductase (8HGO)], and unigene 07368 [iridoid synthase (IS)] were the exceptions, showing higher expression in the roots, but were not statistically significant (Fig. 9). These unigenes may represent isoforms of corresponding genes within *S. japonica*. Expression level of unigenes associated with the secoiridoid biosynthesis pathways showed similar expression trends as that of unigenes associated with MEP


a
GO enrichment analysis for up-regulated unigenes in the Aerial-tissues
% of Sequences
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70
GO-Terms
plastid
thylakoid
photosynthesis
generation of precursor metabolites and energy
cytoplasmic part
lipid metabolic process
single-organism metabolic process
cytoplasm
hydrolase activity
pyrophosphatase activity
motor activity
hydrolase activity, acting on acid anhydrides
hydrolase activity, acting on acid anhydrides, ...
nucleoside-triphosphatase activity
membrane-bounded organelle
intracellular membrane-bounded organelle
catalytic activity
intracellular organelle
organelle
regulation of biological quality
homeostatic process
cellular homeostasis
cellular aromatic compound metabolic process
organic cyclic compound metabolic process
heterocycle metabolic process
nucleobase-containing compound metabolic process
biosynthetic process
transporter activity
metabolic process
secondary metabolic process
post-embryonic development
cellular component organization
cellular component organization or biogenesis
mitochondrion
anatomical structure development
embryo development
cellular process
extracellular region
carbohydrate metabolic process
cellular nitrogen compound metabolic process
nitrogen compound metabolic process
single-organism process
intracellular part
nuclease activity
hydrolase activity, acting on ester bonds
chromatin binding
macromolecular complex binding
cytoskeleton
reproductive system development
developmental process involved in reproduction
reproductive structure development
system development
single organism reproductive process
multicellular organismal process
shoot system development
reproductive shoot system development
flower development
multicellular organismal development
single-multicellular organism process
b
GO enrichment analysis for down-regulated unigenes in the Aerial-tissues
% of Sequences
0 5 10 15 20 25 30 35 40 45 50 55
GO-Terms
catalytic activity
extracellular space
oxygen binding
extracellular region part
extracellular region
abscission
carbohydrate binding
response to extracellular stimulus
receptor activity
transferase activity, transferring phosphorus-containi...
kinase activity
Test Set
Reference Set

◀**Fig. 6** Gene ontology (GO) enrichment analysis for the differentially expressed unigenes in the aerial tissues compared to roots of the *S. japonica*. Unigenes up- or down-regulated by twofold in the aerial tissues of the *S. japonica* were used as a test set, *S. japonica* transcriptome assembly as a reference set, and gene ontology enrichment analysis was performed using Fisher exact test with *p* value cut-off set as <0.05. GO terms enriched and in the differentially expressed unigene list, and statistically significant with respect to the reference transcriptome assembly is shown here

biosynthesis pathways (Fig. 8), and were all up-regulated in the aerial tissues of *S. japonica*. As MEP pathways provide precursor for the secoiridoid pathway, similar expression trends for unigenes associate with MEP and secoiridoid metabolic pathways, and their high expression in the aerial parts compared to the roots suggests that aerial parts are the major contributor of therapeutic metabolites. This may serve as an ecological advantage to the plant, as these metabolites are bitter in taste, and therefore, higher expression in the aerial tissues may serve as a line of defence against the predators. Furthermore, the previous report using an in silico and in vivo tracer study from ^{13}C glucose in the hairy roots of *O. pumila* showed that the monoterpene-secologanin moiety was synthesized via the MEP pathways, and not via MVA pathways (Yamazaki

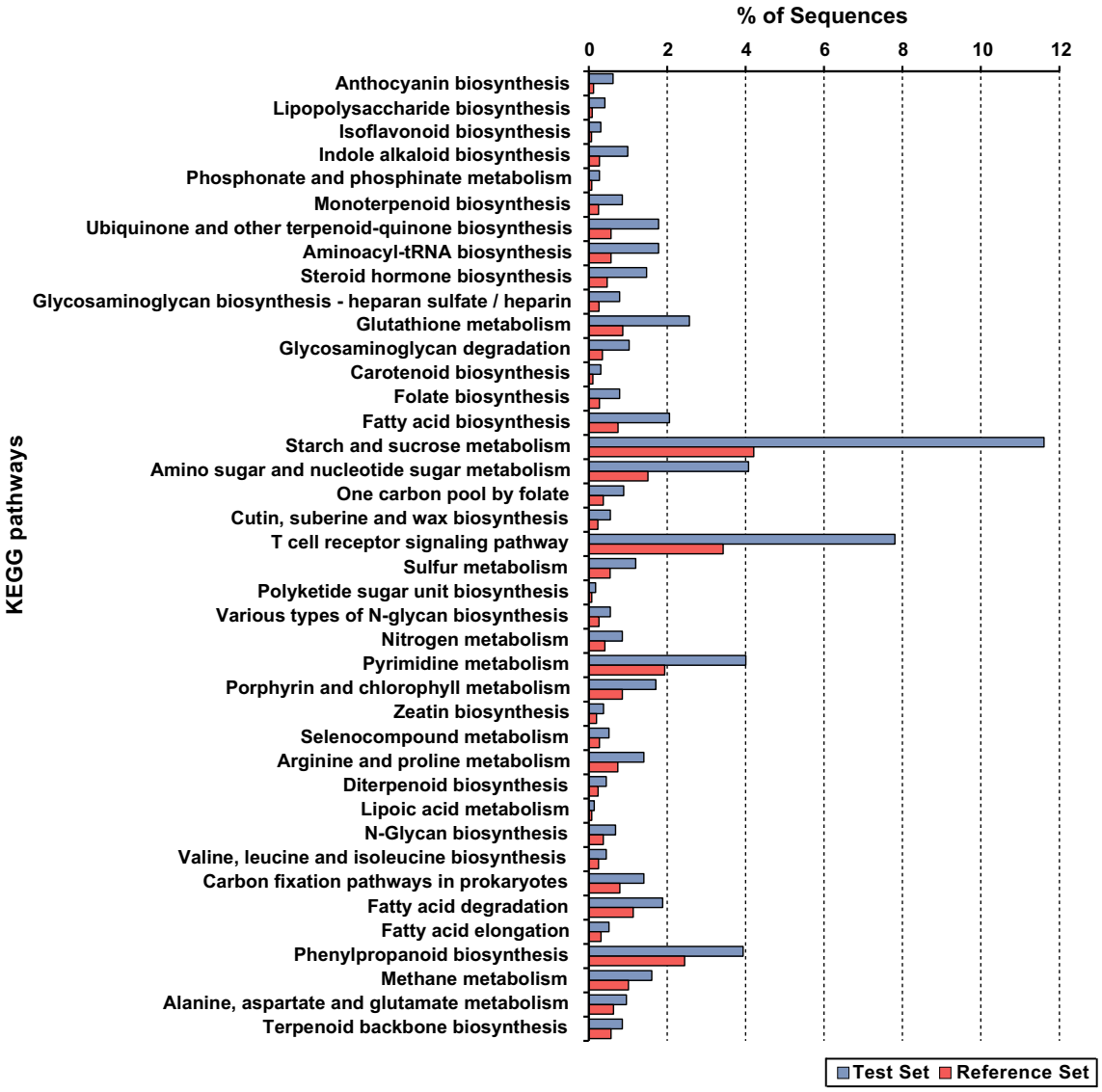
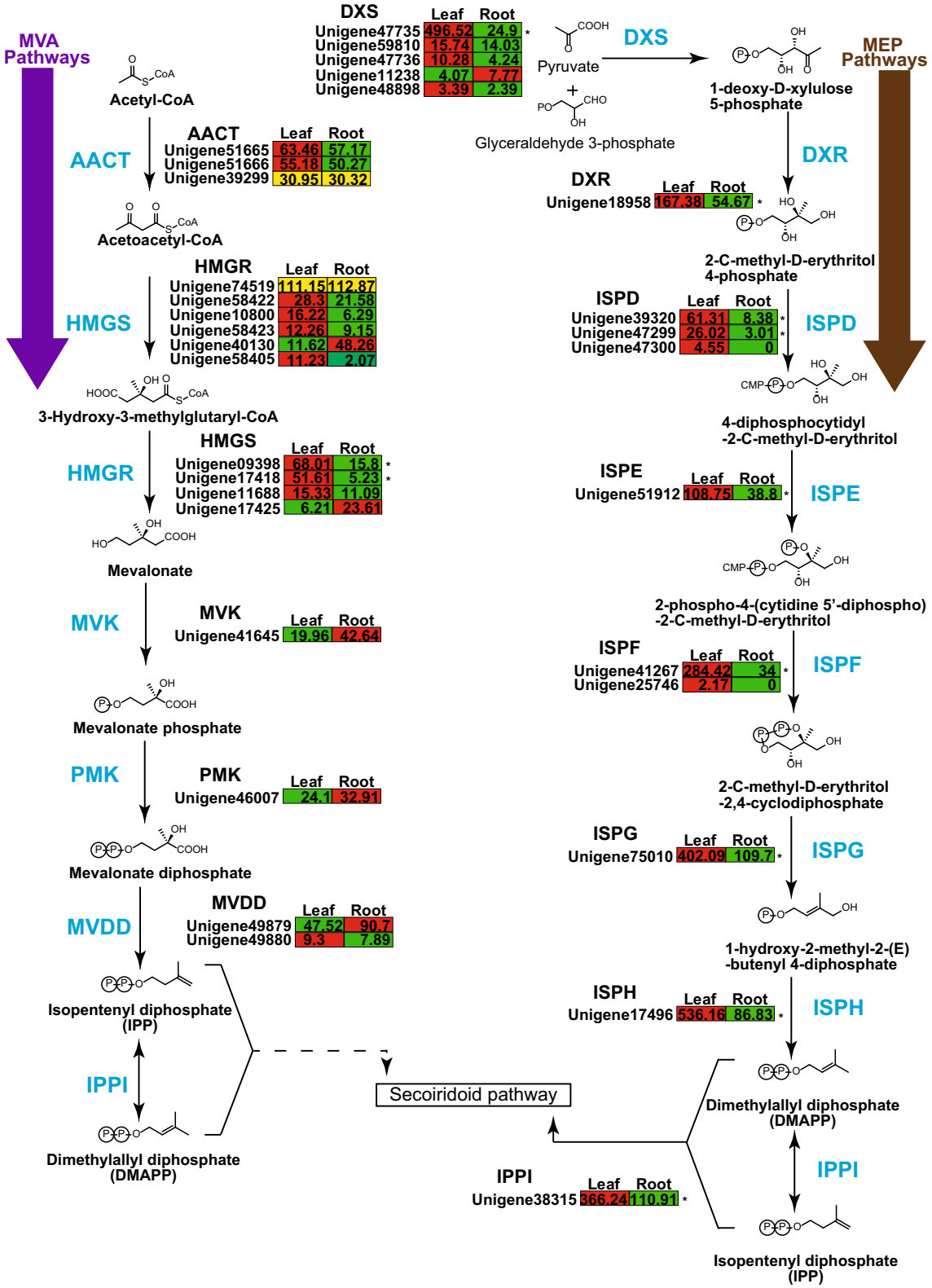


Fig. 7 KEGG pathways enrichment analysis for the differentially expressed unigenes in the aerial tissues compared to the roots of *S. japonica*. Differentially expressed unigenes (up- or down-regulated) in the aerial tissues compared to the roots of *S. japonica* assigned with a KEGG pathway annotation as a test set against entire transcriptome assembly of *S. japonica* with assigned KEGG pathway as a reference set, KEGG pathway enrichment analysis was performed using Fisher exact test with a *p* value cutoff as <0.05, and are arranged in the increasing order of *p* value


MVA Pathways
MEP Pathways
Acetyl-CoA
AACT
Acetoacetyl-CoA
HMGS
3-Hydroxy-3-methylglutaryl-CoA
HMGR
Mevalonate
MVK
Mevalonate phosphate
PMK
Mevalonate diphosphate
MVDD
Isopentenyl diphosphate (IPP)
IPPI
Dimethylallyl diphosphate (DMAPP)
Pyruvate
Glyceraldehyde 3-phosphate
DXS
1-deoxy-D-xylulose 5-phosphate
DXR
2-C-methyl-D-erythritol 4-phosphate
ISPD
4-diphosphocytidyl -2-C-methyl-D-erythritol
ISPE
2-phospho-4-(cytidine 5'-diphospho) -2-C-methyl-D-erythritol
ISPF
2-C-methyl-D-erythritol -2,4-cyclodiphosphate
ISPG
1-hydroxy-2-methyl-2-(E) -butenyl 4-diphosphate
ISPH
Dimethylallyl diphosphate (DMAPP)
IPPI
Isopentenyl diphosphate (IPP)
Secoiridoid pathway
DXS Leaf Root
Unigene47735 496.52 24.9 *
Unigene59810 15.74 14.03
Unigene47736 10.28 4.24
Unigene11238 4.07 7.77
Unigene48898 3.39 2.39
AACT Leaf Root
Unigene51665 63.46 57.17
Unigene51666 55.18 50.27
Unigene39299 30.95 30.32
HMGR Leaf Root
Unigene74519 111.15 112.87
Unigene58422 28.3 21.58
Unigene10800 16.22 6.29
Unigene58423 12.26 9.15
Unigene40130 11.62 48.26
Unigene58405 11.23 2.07
HMGS Leaf Root
Unigene09398 68.01 15.8 *
Unigene17418 51.61 5.23 *
Unigene11688 15.33 11.09
Unigene17425 6.21 23.61
MVK Leaf Root
Unigene41645 19.96 42.64
PMK Leaf Root
Unigene46007 24.1 32.91
MVDD Leaf Root
Unigene49879 47.52 90.7
Unigene49880 9.3 7.89
DXR Leaf Root
Unigene18958 167.38 54.67 *
ISPD Leaf Root
Unigene39320 61.31 8.38 *
Unigene47299 26.02 3.01 *
Unigene47300 4.55 0
ISPE Leaf Root
Unigene51912 108.75 38.8 *
ISPF Leaf Root
Unigene41267 284.42 34 *
Unigene25746 2.17 0
ISPG Leaf Root
Unigene75010 402.09 109.7 *
ISPH Leaf Root
Unigene17496 536.16 86.83 *
IPPI Leaf Root
Unigene38315 366.24 110.91 *

◀**Fig. 8** Detailed pathways and candidate unigenes involved in the mevalonate (MVA) pathway or methylerythritol phosphate (MEP) pathway route of terpene biosynthesis. Biochemical reactions and enzyme coding genes involved in the MVA or MEP pathways, and the candidate unigenes of *S. japonica* for the corresponding enzyme coding genes at different biochemical reactions and their expression values are shown. *Red* or *green colour* represents high or low expression levels, respectively, of a specific unigenes across aerial tissues and roots of *S. japonica*, while *asterisks* identify the false discovery rate, (FDR) <0.05, with the DESeq2 package. *AACT* acetoacetyl-CoA thiolase, *HMGR* 3-hydroxy-3-methylglutaryl-CoA reductase, *HMGS* 3-hydroxy-3-methylglutaryl-CoA synthase, *MVK* mevalonate kinase, *PMK* phosphomevalonate kinase, *MVDD* mevalonate diphosphate decarboxylase, *DXS* 1-deoxy-D-xylulose 5-phosphate synthase, *DXR* 1-deoxy-D-xylulose 5-phosphate reductoisomerase, *ISPD* 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase, *ISPE* 4-(cytidine-5′-diphospho)-2-C-methyl-d-erythritol kinase, *ISPF* 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, *ISPG* (E)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase, *ISPH* (E)-4-hydroxy-3-methylbut-2-enyl diphosphate reductase, and *IPPI* isopentenyl diphosphate isomerase

et al. 2004). While our results are certainly not enough and conclusive to make the similar statement for *S. japonica*, we are tempted to propose that MEP pathways could potentially be the major pathway contributing to the biosynthesis of secoiridoids and other associated metabolites. It will be interesting to test this hypothesis in future studies.

Mangiferin, a xanthone-C-glucoside, is an important constituent of metabolic extracts from several *Swertia* species. Mangiferin is biosynthesized through coumaric acid derived from the phenylalanine, which then converts into caffeic acid or to the benzophenone intermediates, such as iriflophenone, and combined with malonyl-CoA results to the final product (Fig. S4). While there are no reports on the extraction or identification of mangiferin or its derivatives from *S. japonica*, we investigated transcripts expression of unigenes annotated as different biosynthetic enzymes from mangiferin biosynthesis pathways. Out of 16 enzyme coding genes known to be involved in the mangiferin biosynthesis, sequence-similarity-based annotation resulted in the identification of all genes with the exception of the 3-dehydroquinate dehydratase (DHQD) annotated unigene (Data set. S4). Expression levels of unigenes showed no specific trend, with unigenes annotated as 3-deoxy-D-arabinoheptulosonate-7-phosphate synthase (DAHPS), 3-dehydroquinate synthase (DHQS), shikimate kinase (SAK), tyrosine ammonia lyase (TAL), and arogenate dehydrogenase (ADH) being effectively up-regulated, while unigenes annotated as chorismate mutase (CM), p-coumarate-3-hydroxylase (C3H), trans-cinnamate-4-hydroxylase (C4H), 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), and phenylalanine ammonia lyase (PAL) being effectively down-regulated in the aerial tissues with respect to the roots (Fig. S4). Unigenes annotated as prephenate dehydratase (ADT), chorismate synthase (CS), and aspartate–prephenate aminotransferase (PAT) showed mixed expression trend, with equal number of assigned unigenes showing higher or reduced expression levels in the aerial tissues with respect to the roots. Unlike metabolic pathways leading to the biosynthesis of swertiamarin and amarogentin, showing significant higher transcript abundance in the aerial tissues when compared with the roots, no such trends were observed for the mangiferin biosynthetic pathways. Furthermore, unigenes annotated as 3-dehydroquinate dehydratase using 3-dehydroquinate as a substrate for the synthesis of 3-hydroshikimate, an important step for the biosynthesis of mangiferin, were not identified in the *S. japonica* transcriptome assembly. While mangiferin is considered as an important constituent of extracts derived from several *Swertia* species, such as *S. chirata*, *S. punicea*, *S. perfoliata*, *S. mussotii*, *S. macrosperma*, among others, it has not been identified in *S. japonica*. It is not clear if mangiferin is synthesized in *S. japonica* at all or at a very low level. Furthermore, inability to identify DHQD coding unigenes could be due to low expression levels of this transcript, which, therefore, were not captured in the de novo transcriptome assembly. It would be interesting to investigate if at all mangiferin is synthesized in *S. japonica*, and if not, what caused this deviation from other mangiferin producing *Swertia* species.

Phylogenetic analysis of unigenes annotated as glycosyl transferase and up-regulated in the aerial tissues of the *Swertia japonica*

Swertiamarin, amarogentin, along with other xanthones, and secoiridoids, synthesized in the *S. japonica*, undergo metabolic conjugation processes, such as glycosylation, resulting in the increased metabolic diversity with novel bioactivity. Several glycosylated forms of swertiamarin, swertianin, bellidifolia, isoswertianin, norswertianin among others have been extracted and characterized in the *S. japonica*, yet none of the involved UDP-glucosyltransferases in these reactions are known. In the *S. japonica* transcriptome assembly, we identified 103 unigenes annotated as UDP-glycosyltransferases with a minimum length of 500 bps, and transcripts abundance over 1 FPKM in either of the root or of aerial tissues. Among these, 37 of the unigenes were up-regulated (Table S8, Data set. S5), while 66 unigenes were down-regulated (Table S9, Data set. S6) in the aerial tissues with respect to the roots. Since MEP and secoiridoid biosynthesis pathways were up-regulated in the aerial tissues, we considered up-regulated UDP-glycosyltransferase-annotated unigenes in the aerial tissues as potential candidates to be involved in the glycosylation of key bioactive metabolites. Phylogenetic

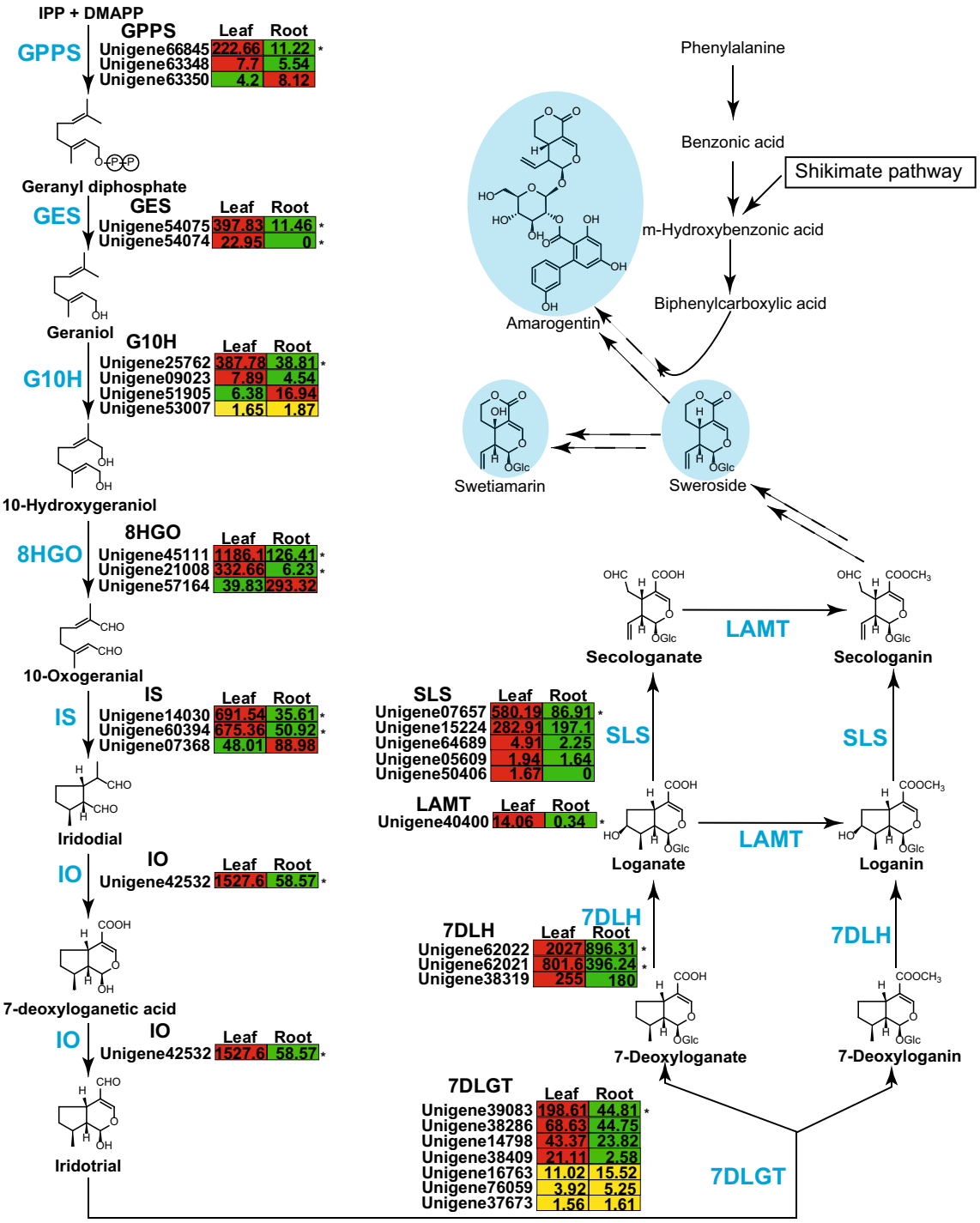


Fig. 9 Proposed swertiamarin and amarogentin biosynthetic pathways, and expression level of associated unigenes across the aerial tissues and the roots of *S. japonica*. Biosynthetic pathways and candidate unigenes of *S. japonica* annotated as the enzyme coding genes at different biochemical reactions, and their expression levels are shown here. *Red* or *green colour* represents high or low expression levels, respectively, for a specific unigenes across the aerial tissues and the roots of *S. japonica*, while *asterisks* identify the false discovery rate (FDR) <0.05 with the DESeq2 package. Biochemical reactions with unknown enzyme coding genes involved or sequences of uncharacterized reactions are shown through *discontinued arrows*. *GPPS* geranyl diphosphate synthase, *GES* geraniol synthase, *G10H* geraniol 10-hydroxylase, *8HGO* 8-hydroxygeraniol oxidoreductase, *IS* iridoid synthase, *IO* iridoid oxidase, *7DLGT* 7-deoxyloganetic acid glucosyl transferase, *7DLH* 7-deoxyloganic acid hydroxylase, *LAMT* loganic acid *O*-methyltransferase, and *SLS* secologanin synthase

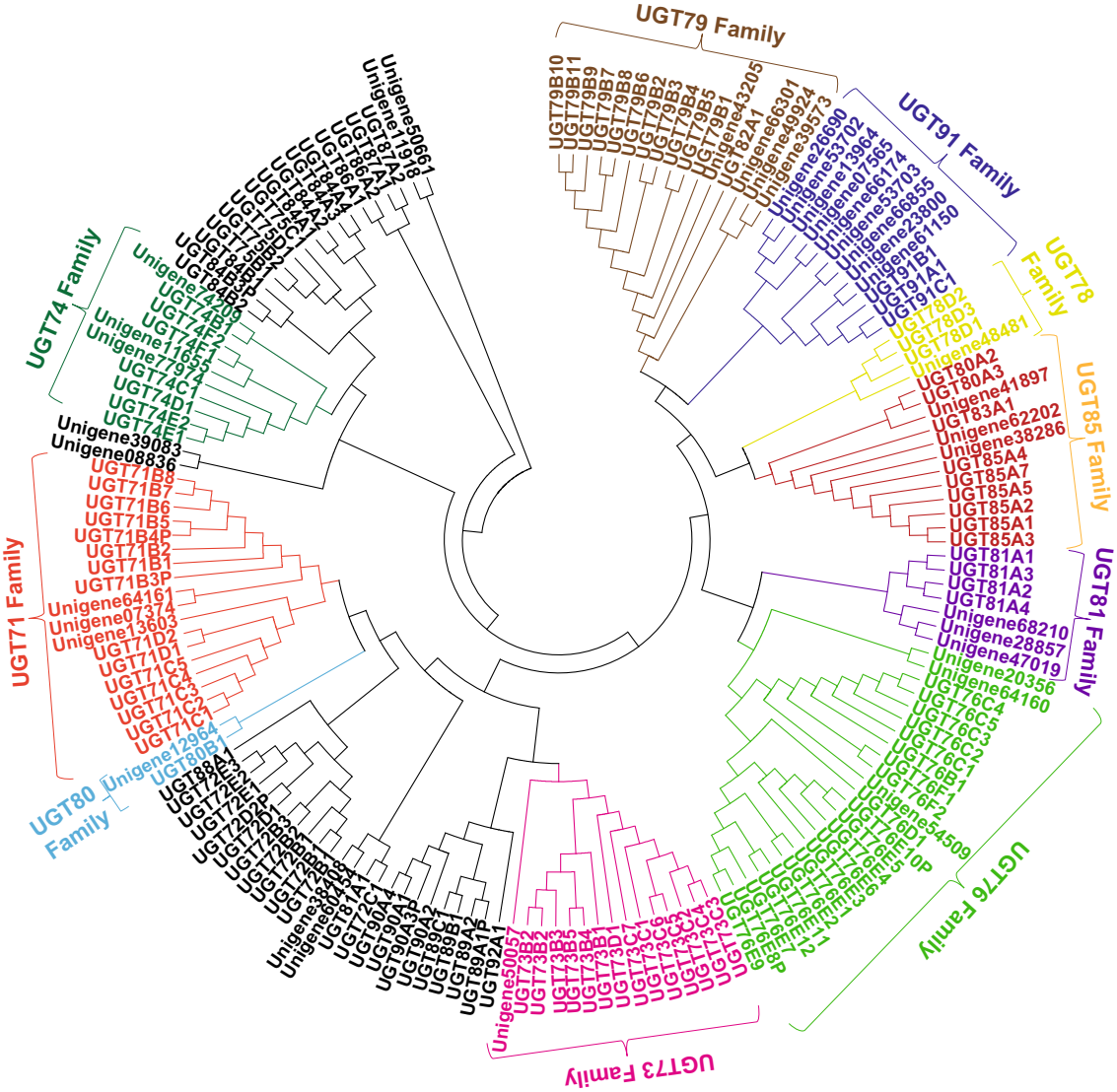


Fig. 10 Phylogenetic analysis of UDP-glycosyltransferase-annotated unigenes with genes annotated as UDP-glycosyltransferase from *Arabidopsis* genome. Nucleotide sequences of the unigenes, annotated as UDP-glycosyltransferase and up-regulated in the aerial tissues with respect to the roots of *S. japonica*, 37 in total, were aligned with sequences of 128 UDP-glycosyltransferase-annotated genes from the *Arabidopsis* genome using MUSCLE program, and evolutionary distance was computated using the Jones–Taylor–Thornton (JTT) method. Neighbor-Joining (NJ) tree was then constructed with the bootstrap values obtained with applying 10,000 replications using MEGA6 program

analysis of these 37 unigenes was performed together with all UDP-glycosyltransferase encoding genes from *Arabidopsis* genome, 128 genes in total, obtained and classified from http://www.p450.kvl.dk/UGT.shtml (Bertrand 1926). The majority of the unigenes were grouped in the UGT91 family (nine unigenes), followed by UGT79 family (four unigenes), while three unigenes each were grouped in the UGT85, UGT81, UGT76, UGT74, and UGT71 family (Fig. 10). Several the *Arabidopsis* genes from these GT families, such as UGT85A1 (Hou et al. 2004), UGT81A1 (Benning and Ohta 2005), UGT71B2 (Zhao et al. 2007), UGT71B6 (Priest et al. 2005), UGT71C1 (Lim et al. 2008), UGT78D1 (Jones et al. 2003), UGT79B1 (Bogs et al. 2007), UGT76C1 (Hou et al. 2004), UGT76C2, among others, have been characterized to add sugar moieties to the different classes of metabolites with varying specificity. These UGTs' families have been discussed to be important for greater metabolite diversity (Yonekura-Sakakibara and Hanada 2011), and therefore, unigenes up-regulated in the aerial tissues and grouped within these UDP-glycosyltransferase families most likely play an important role in the biosynthesis of diverse bioactive metabolites, and, therefore, represent important candidates for the further functional characterization.

Conclusion

Our study is the first attempt of de novo transcriptome assembly and annotation of any of the plants from the *Swertia* genus, and represents comprehensive resources for *S. japonica* plant. The de novo transcriptome assembly was obtained by merging assemblies from three individual assemblers. Our results showed an advantage to the use of multi-assembler approach over individual assemblers to obtain de novo transcriptome assembly. The transcriptome assembly annotation and characterization identified putative candidate unigenes involved in the major secondary metabolic pathways, including secoiridoids, iridoid glycosides, xanthones, and flavonoids biosynthesis pathways. Transcript expression analysis showed key metabolic pathways leading to the biosynthesis of therapeutic metabolites being enriched to the aerial tissues. Our datasets are an integrated genomic resource for molecular cloning and further functional characterization of genes of interest associated with key bioactive metabolite biosynthesis pathways. Identified sequences for enzyme coding genes potentially involved in the biosynthesis of key therapeutic metabolites may serve as a basis to understand medicinal properties of *S. japonica*, and may be extended to other *Swertia* species. Given the incomplete knowledge on the molecular basis for the biosynthesis of key therapeutic metabolites in the *Swertia* species, our transcriptome analysis provides useful information regarding the specialized metabolism of *Swertia* species in general, and *S. japonica* in particular.

Author contribution statement KS MY: conceived and designed the experiments. MN AR: performed the experiments. HS: performed deep-transcriptome sequencing. KS MY HS: contributed reagents/materials/analysis tools. AR MY HT: conducted the data interpretation and laboratory analysis. AR MY KS: contributed to the manuscript preparation. All authors read and approved the final manuscript.

Acknowledgments This study, in part, was supported by the "Health and Labour Sciences Research Grant" on the enhancement of "Comprehensive Medicinal Plant Database" by the Grants-in-Aid for the Scientific Research of the Japan Society for the Promotion of Science (JSPS), and by the "Strategic Priority Research Promotion Program" of the Chiba University. HT was partially supported by MEXT KAKENHI (no. 221S0002). The super-computing resources were provided by the National Institute of Genetics (NIG), and Research Organization of Information and Systems, Japan. The computing resources were provided by the Medical Mycology Research Center, Chiba University, Japan. We also thank Mr. Tsutomu Hosouchi, and Ms. Sayaka Shinpo from the Kazusa DNA Research Institute for technical support in Illumina sequencing. We would also like to thank Dr. Ashfaq Mahmood, Megha from Chiba University, Dr. Gourvendu Saxena from the National University of Singapore, and two anonymous reviewers for their helpful and critical comments to improve this manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Basnet P, Kadota S, Shimizu M, Namba T (1994) Bellidifolin: a potent hypoglycemic agent in streptozotocin (STZ)-induced diabetic rats from Swertia japonica. Planta Med 60:507–511

Benning C, Ohta H (2005) Three enzyme systems for galactoglycerolipid biosynthesis are coordinately regulated in plants. J Biol Chem 280:2397–2400

Bertrand G (1926) The importance of minute chemical constituents (infiniment petits chimiques) of biological products: nickel, cobalt and insulin. Science 64:629–630

Bogs J, Jaffe FW, Takos AM, Walker AR, Robinson SP (2007) The grapevine transcription factor VvMYBPA1 regulates proanthocyanidin synthesis during fruit development. Plant Physiol 143:1347–1361

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

Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I, Boisvert S, Chapman JA, Chapuis G, Chikhi R, Chitsaz H, Chou WC, Corbeil J, Del Fabbro C, Docking TR, Durbin R, Earl D, Emrich S, Fedotov P, Fonseca NA, Ganapathy G, Gibbs RA, Gnerre S, Godzaridis E, Goldstein S, Haimel M, Hall G, Haussler D, Hiatt JB, Ho IY, Howard J, Hunt M, Jackman SD, Jaffe DB, Jarvis ED, Jiang H, Kazakov S, Kersey PJ, Kitzman JO, Knight JR, Koren S, Lam TW, Lavenier D, Laviolette F, Li Y, Li Z, Liu B, Liu Y, Luo R, Maccallum I, Macmanes MD, Maillet N, Melnikov S, Naquin D, Ning Z, Otto TD, Paten B, Paulo OS, Phillippy AM, Pina-Martins F, Place M, Przybylski D, Qin X, Qu C, Ribeiro FJ, Richards S, Rokhsar DS, Ruby JG, Scalabrin S, Schatz MC, Schwartz DC, Sergushichev A, Sharpe T, Shaw TI, Shendure J, Shi Y, Simpson JT, Song H, Tsarev F, Vezzi F, Vicedomini R, Vieira BM, Wang J, Worley KC, Yin S, Yiu SM, Yuan J, Zhang G, Zhang H, Zhou S, Korf IF (2013) Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. GigaScience 2:10

Brahmachari G, Mondal S, Gangopadhyay A, Gorai D, Mukhopadhyay B, Saha S, Brahmachari AK (2004) Swertia (Gentianaceae): chemical and pharmacological aspects. Chem Biodivers 1:1627–1651

Cao TW, Geng CA, Ma YB, He K, Wang HL, Zhou NJ, Zhang XM, Tao YD, Chen JJ (2013) Xanthones with anti-hepatitis B virus activity from Swertia mussotii. Planta Med 79:697–700

Conesa A, Gotz S (2008) Blast2GO: a comprehensive suite for functional analysis in plant genomics. Int J Plant Genom 2008:619832

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676

Duan J, Xia C, Zhao G, Jia J, Kong X (2012) Optimizing de novo common wheat transcriptome assembly using short-read RNA-Seq data. BMC Genom 13:392

Dutt B, Srivastava LJ, Singh JM (1996) Swertia spp: a source of bitter compounes for medicinal use. Anc Sci Life 15:226–229

Eddouks M, Chattopadhyay D, De Feo V, Cho WC (2014) Medicinal plants in the prevention and treatment of chronic diseases 2013. Evid Based Complement Alter Med eCAM 2014:180981

Edgar R, Domrachev M, Lash AE (2002) Gene expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30:207–210

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

Fukushima A, Nakamura M, Suzuki H, Saito K, Yamazaki M (2015) High-throughput sequencing and de novo assembly of red and green forms of the *Perilla frutescens* var. crispa transcriptome. PLoS One 10:e0129154

Gene Ontology C (2015) Gene ontology consortium: going forward. Nucleic Acids Res 43:D1049–D1056

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652

Gruenheit N, Deusch O, Esser C, Becker M, Voelckel C, Lockhart P (2012) Cutoffs and k-mers: implications from a transcriptome study in allopolyploid plants. BMC Genom 13:92

Hancock JM (1996) Simple sequences in a "minimal" genome. Nat Genet 14:14–15

Hase K, Li J, Basnet P, Xiong Q, Takamura S, Namba T, Kadota S (1997) Hepatoprotective principles of Swertia japonica Makino on D-galactosamine/lipopolysaccharide-induced liver injury in mice. Chem Pharm Bull 45:1823–1827

Hiramatsu N, Xiufen W, Takechi R, Itoh Y, Mamo J, Pal S (2004) Antimutagenicity of Japanese traditional herbs, gennoshoko, yomogi, senburi and iwa-tobacco. BioFactors 22:123–125

Hou B, Lim EK, Higgins GS, Bowles DJ (2004) *N*-glucosylation of cytokinins by glycosyltransferases of *Arabidopsis thaliana*. J Biol Chem 279:47822–47832

Ikeshiro Y, Tomita Y (1985) Iridoid glucoside of *Swertia japonica*. Planta Med 51:390–393

Ikeshiro Y, Tomita Y (1987) Senburiside II, a new iridoid glucoside from *Swertia japonica*. Planta Med 53:158–161

Jarvis DE, Kopp OR, Jellen EN, Mallory MA, Pattee J, Bonifacio A, Coleman CE, Stevens MR, Fairbanks DJ, Maughan PJ (2008) Simple sequence repeat marker development and genetic mapping in quinoa (Chenopodium quinoa Willd.). J Genet 87:39–51

Jiang W, Zhu DL, Wang MF, Yang QS, Zuo MY, Zeng L, Li GP (2015) Xanthones from the herb of Swertia elata and their anti-TMV activity. Nat Product Res 1–6

Jones P, Messner B, Nakajima J, Schaffner AR, Saito K (2003) UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in *Arabidopsis thaliana*. J Biol Chem 278:43910–43918

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240

Kakiuchi N, Iwaki N, Mikage M, Xiao H, Wang Z, Hattori M (2014) Phylogenetic examination of crude drugs derived from Yunnanese Swertia plants. J Nat Med 68:206–210

Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M (2014) Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res 42:D199–D205

Kavimani S, Manisenthlkumar KT (2000) Effect of methanolic extract of Enicostemma littorale on Dalton's ascitic lymphoma. J Ethnopharmacol 71:349–352

Kelly RJ, Vincent DE, Friedberg I (2010) IPRStats: visualization of the functional potential of an InterProScan run. BMC Bioinform 11(Suppl 12):S13

Khanal S, Shakya N, Thapa K, Pant DR (2015) Phytochemical investigation of crude methanol extracts of different species of Swertia from Nepal. BMC Res Notes 8:821

Kikuchi M, Kikuchi M (2004) Studies on the constituents of Swertia japonica MAKINO I. On the structures of new secoiridoid diglycosides. Chem Pharm Bull 52:1210–1214

Kikuchi M, Kikuchi M (2005) Studies on the constituents of *Swertia japonica* Makino II. On the structures of new glycosides. Chem Pharm Bull 53:48–51

Kimura Y, Sumiyoshi M (2011) Effects of Swertia japonica extract and its main compound swertiamarin on gastric emptying and gastrointestinal motility in mice. Fitoterapia 82:827–833

Komatsu M, Tomimori T, Makiguchi Y, Asano K (1968) Studies on the constituents of Swertia japonica. 3. On the flavonoid constituents of the plants of Swertia spp. Yakugaku zasshi J Pharm Soc Jpn 88:832–837

Komatsu M, Tomimori T, Mikuriya N (1969) Studies on the constituents of Swertia japonica. IV. Isolation and structure of xanthones. Chem Pharm Bull 17:155–162

Kshirsagar PR, Pai SR, Nimbalkar MS, Gaikwad NB (2016) RP-HPLC analysis of seco-iridoid glycoside swertiamarin from different Swertia species. Nat Prod Res 30:865–868

Kumar S, Blaxter ML (2010) Comparing de novo assemblers for 454 transcriptome data. BMC Genom 11:571

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform 12:323

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659

Liang JZ (1984) The active constituents of Swertia patens. Zhong yao tong bao 9:34–36

Licciardi PV, Underwood JR (2011) Plant-derived medicines: a novel class of immunological adjuvants. Int Immunopharmacol 11:390–398

Lim CE, Choi JN, Kim IA, Lee SA, Hwang YS, Lee CH, Lim J (2008) Improved resistance to oxidative stress by a loss-of-function mutation in the Arabidopsis UGT71C1 gene. Mol Cells 25:368–375

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550

Medda S, Mukhopadhyay S, Basu MK (1999) Evaluation of the in vivo activity and toxicity of amarogentin, an antileishmanial agent, in both liposomal and niosomal forms. J Antimicrob Chemother 44:791–794

Menkovic N, Savikin-Fodulovic K, Bulatovic V, Aljancic I, Juranic N, Macura S, Vajs V, Milosavljevic S (2002) Xanthones from Swertia punctata. Phytochemistry 61:415–420

Miettinen K, Dong L, Navrot N, Schneider T, Burlat V, Pollier J, Woittiez L, van der Krol S, Lugan R, Ilc T, Verpoorte R, Oksman-Caldentey KM, Martinoia E, Bouwmeester H, Goossens A, Memelink J, Werck-Reichhart D (2014) The seco-iridoid pathway from *Catharanthus roseus*. Nat Commun 5:3606

Misra A, Shasany AK, Shukla AK, Darokar MP, Singh SC, Sundaresan V, Singh J, Bagchi GD, Jain SP, Saikia D, Khanuja SP (2010) AFLP markers for identification of Swertia species (Gentianaceae). Genet Mol Res GMR 9:1535–1544

Mulder N, Apweiler R (2007) InterPro and InterProScan: tools for protein sequence classification and comparison. Methods Mol Biol 396:59–70

Muranaka T, Saito K (2013) Phytochemical genomics on the way. Plant Cell Physiol 54:645–646

Nakasugi K, Crowhurst R, Bally J, Waterhouse P (2014) Combining transcriptome assemblies from multiple de novo assemblers in the allo-tetraploid plant *Nicotiana benthamiana*. PLoS One 9:e91776

Neerja Pant DC, Jain Bhakuni RS (2000) Phytochemicals from genus Swertia and their biological activities. Indian J Chem 39B:565–586

Negi Jagmohan S, Singh Pramod, Rawat B (2011) Chemical constituents and biological importance of Swertia: a review. Curr Res Chem 3(1):1–15

Padhan JK, Kumar V, Sood H, Singh TR, Chauhan RS (2015) Contents of therapeutic metabolites in Swertia chirayita correlate with the expression profiles of multiple genes in corresponding biosynthesis pathways. Phytochemistry 116:38–47

Pal D, Sur S, Mandal S, Das A, Roy A, Das S, Panda CK (2012) Prevention of liver carcinogenesis by amarogentin through modulation of G1/S cell cycle check point and induction of apoptosis. Carcinogenesis 33:2424–2431

Pan SY, Litscher G, Gao SH, Zhou SF, Yu ZL, Chen HQ, Zhang SF, Tang MK, Sun JN, Ko KM (2014) Historical perspective of traditional indigenous medical practices: the current renaissance and conservation of herbal resources. Evid Based Complement Alter Med eCAM 2014:525340

Paquette SM, Jensen K, Bak S (2009) A web-based resource for the Arabidopsis P450, cytochromes b5, NADPH-cytochrome P450 reductases, and family 1 glycosyltransferases (http://www.P450.kvl.dk). Phytochemistry 70:1940–1947

Peres V, Nagem TJ, de Oliveira FF (2000) Tetraoxygenated naturally occurring xanthones. Phytochemistry 55:683–710

Phoboo S, Pinto Mda S, Barbosa AC, Sarkar D, Bhowmik PC, Jha PK, Shetty K (2013) Phenolic-linked biochemical rationale for the anti-diabetic properties of Swertia chirayita (Roxb. ex Flem.) Karst. Phytother Res PTR 27:227–235

Priest DM, Jackson RG, Ashford DA, Abrams SR, Bowles DJ (2005) The use of abscisic acid analogues to analyse the substrate selectivity of UGT71B6, a UDP-glycosyltransferase of *Arabidopsis thaliana*. FEBS Lett 579:4454–4458

Rai A, Saito K (2016) Omics data input for metabolic modeling. Curr Opin Biotechnol 37:127–134

Rai A, Yamazaki M, Takahashi H, Nakamura M, Kojoma M, Suzuki H, Saito K (2016) RNA-seq transcriptome analysis of *Panax japonicus*, and its comparison with other Panax species to identify potential genes involved in the saponins biosynthesis. Front Plant Sci 7

Saito K (2013) Phytochemical genomics—a new trend. Curr Opin Plant Biol 16:373–380

Saravanan S, Islam VI, Babu NP, Pandikumar P, Thirugnanasambantham K, Chellappandian M, Raj CS, Paulraj MG, Ignacimuthu S (2014) Swertiamarin attenuates inflammation mediators via modulating NF-kappaB/I kappaB and JAK2/STAT3 transcription factors in adjuvant induced arthritis. Eur J Pharm Sci Off J Eur Fed Pharm Sci 56:70–86

Senthilvel S, Jayashree B, Mahalakshmi V, Kumar PS, Nakka S, Nepolean T, Hash C (2008) Development and mapping of simple sequence repeat markers for pearl millet from data mining of expressed sequence tags. BMC Plant Biol 8:119

Steijger T, Abril JF, Engstrom PG, Kokocinski F, Hubbard TJ, Guigo R, Harrow J, Bertone P, Consortium R (2013) Assessment of transcript reconstruction methods for RNA-seq. Nat Methods 10:1177–1184

Takei H, Nakauchi K, Yoshizaki F (2001) Analysis of swertiamarin in Swertia herb and preparations containing this crude drug by capillary electrophoresis. Anal Sci Int J Jpn Soc Anal Chem 17:885–888

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). TAG Theor Appl Genet 106:411–422

Tomimori T, Komatsu M (1969) Studies on the constituents of *Swertia japonica*. V. On the xanthone constituents of the plants of Swertia ssp. Yakugaku zasshi J Pharm Soc Jpn 89:410–417

Vaidya H, Goyal RK, Cheema SK (2013) Anti-diabetic activity of swertiamarin is due to an active metabolite, gentianine, that upregulates PPAR-gamma gene expression in 3T3-L1 cells. Phytother Res PTR 27:624–627

Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002) In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 7:537–546

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55

Veeresham C (2012) Natural products derived from plants as a source of drugs. J Adv Pharm Technol Res 3:200–201

Visser EA, Wegrzyn JL, Steenkmap ET, Myburg AA, Naidoo S (2015) Combined de novo and genome guided assembly and annotation of the Pinus patula juvenile shoot transcriptome. BMC Genom 16:1057

Wang J, Zhao C, Liu C, Xia G, Xiang F (2011) Introgression of Swertia mussotii gene into Bupleurum scorzonerifolium via somatic hybridization. BMC Plant Biol 11:71

Weeks AM, Chang MC (2011) Constructing de novo biosynthetic pathways for chemical synthesis inside living cells. Biochemistry 50:5404–5418

Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, Zhou X, Lam TW, Li Y, Xu X, Wong GK, Wang J (2014) SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 30:1660–1666

Yamahara J, Konoshima T, Sawada T, Fujimura H (1978) Biologically active principles of crude drugs: pharmacological actions of *Swertia japonica* extracts, swertiamarin and gentianine (author's transl). Yakugaku zasshi J Pharm Soc Jpn 98:1446–1451

Yamahara J, Kobayashi M, Matsuda H, Aoki S (1991) Anticholinergic action of *Swertia japonica* and an active constituent. J Ethnopharmacol 33:31–35

Yamazaki Y, Kitajima M, Arita M, Takayama H, Sudo H, Yamazaki M, Aimi N, Saito K (2004) Biosynthesis of camptothecin. In silico and in vivo tracer study from [1-13C]glucose. Plant Physiol 134:161–170

Yonekura-Sakakibara K, Hanada K (2011) An evolutionary view of functional diversity in family 1 glycosyltransferases. Plant J Cell Mol Biol 66:182–193

Zhao Y, Chow TF, Puckrin RS, Alfred SE, Korir AK, Larive CK, Cutler SR (2007) Chemical genetic interrogation of natural variation uncovers a molecule that is glycoactivated. Nat Chem Biol 3:716–721

Zheng HH, Luo CT, Chen H, Lin JN, Ye CL, Mao SS, Li YL (2014) Xanthones from Swertia mussotii as multitarget-directed antidiabetic agents. ChemMedChem 9:1374–1377