De novo transcriptome analysis deciphered polyoxypregnane glycoside biosynthesis pathway in *Gymnema sylvestre*

Kuldeepsingh A. Kalariya¹ · Dipal B. Minipara¹ · Ponnuchamy Manivel¹

Received: 20 March 2018 / Accepted: 6 August 2018 / Published online: 21 August 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Gymnema sylvestre is an important medicinal plant containing antidiabetic activity. Through de novo transcriptomic study, the pathways of polyoxypregnane glycosides were explored and candidate genes of these pathways were identified in *G. sylvestre*. High-quality raw reads were assembled into transcripts which resulted in 193,615 unigenes. These unigenes further decoded 58,274 coding DNA sequences (CDSs). Functional annotation of predicted CDSs was carried out using the protein databases, i.e., NCBI's non-redundant, Uniprot and Pfam. Eukaryotic orthologous group (KOG) classification and transcription factor analysis has revealed most CDS-enriched categories as "Signal transduction mechanism" and "Basic Helix loop helix" (bHLH) transcription factor family, respectively. A total of 16,569 CDSs were assigned minimum one Gene Ontology (GO) term. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis disclosed 235 CDSs which represented total 27 genes of pregnane glycoside pathways and 19 CDSs represented 10 important enzymes of polyoxypregnane glycoside biosynthesis, i.e., sterol 24-C-methyltransferase, cycloeculanol cycloisomerase, Δ^{14} -sterol reductase, C-8,7 sterol isomerase, sterol methyltransferase 2, C-5 sterol desaturase, sterol Δ^7 reductase, Δ^{24} sterol reductase, 3 β -hydroxysteroid dehydrogenase and progesterone 5 β reductase (5 β POR). This transcriptome analysis provided an important resource for future functional genomic studies in *G. sylvestre*.

Keywords *Gymnema sylvestre* · Next generation sequencing · Pregnane glycosides · RNA sequencing

Introduction

Gymnema sylvestre (Madhunashini) is a member of the family Asclepiadaceae in which milky latex is a special characteristic. Leaves of these plants are traditionally used as an anti-inflammatory and palliative herbal medicine. Some triterpene saponins from *G. sylvestre* are reported to attenuate hyperglycemia (Gupta et al. 1962; Anupam and Malay 1994). Though it also contains steroidal glycosides, it is popular mainly for its antidiabetic properties due to the presence of triterpenoid gymnemic acids and thus, the presence of pregnanes and its derivatives like cardiac glycosides seems to be underestimated.

Pregnanes are a group of diverse steroids derived from triterpenoid with 21 carbon-fused rings evolved as a means of primary defense in plants (Agrawal et al. 2012). They

appear to modulate food intake by possibly affecting the hypothalamic feeding circuits (Komarnytsky et al. 2013). The pregnane is a precursor of cardenolides which is used in the therapy of cardiac insufficiency in humans. Hydroxysteroid dehydrogenase (HSD) and progesterone 5 β -reductase (5 β POR) are both supposed to be important enzymes in the biosynthesis of cardenolides (Ernst et al. 2010). In plants, the natural steroids are formed from acetic acid to cycloartenol through the formation of intermediate stable products like mevalonic acid and squalene. Based on a partial pathway derived in *Digitalis lanata*, it is postulated that steroidal framework of pregnane is derivative of terpenoids (Pandey et al. 2016).

Some researchers have shown the presence of cardiac glycosides in *G. sylvestre* extracts and NMR-based structural depiction of four new pregnane glycosides in hypoglycemic constituents named as gymmysylvestrosides A–D (Xu et al. 2015). Despite the immense medicinal importance of pregnane glycosides in *G. sylvestre* plant, putative genes involve in pregnane glycosides have not been studied till date. Hence, present study was undertaken to identify steroidal

✉ Kuldeepsingh A. Kalariya
kuldeep_ka@yahoo.co.in; Kuldeep.Kalariya@icar.gov.in

¹ ICAR-Directorate of Medicinal and Aromatic Plants
Research, Boriavi, Anand, Gujarat, India

pathways for polyoxypregnane glycoside biosynthesis in *G. sylvestre* genotype using next generation sequencing platform (Illumina). The present study is a pioneering attempt to foster comprehensive molecular information of expressed transcripts of *G. sylvestre* genotype.

Materials and methods

Plant material

The experimental plant of *G. sylvestre* genotype (DGS-3) was originally collected from Waghai, Dang district, Gujarat, India in 2009 and maintained with standard package of practices of cultivation at the experimental field of ICAR-Directorate of Medicinal and Aromatic Plants Research, Boriavi, Anand, Gujarat which falls under a sub-tropical area of India (22.556°N, 72.951°E) at an altitude of 40.63 m above mean sea level with annual average rainfall of 800 mm. Healthy leaves prior to flowering were collected from field grown *G. sylvestre* plants.

Total RNA extraction

Total RNA of leaf was extracted using Norgen total RNA isolation kit according to manufacturer's instruction. The total RNA was quantified using Qubit fluorometer. Purity of RNA was confirmed using 1% denaturing agarose gel electrophoresis and using Agilent RNA 6000 nanochip on bioanalyzer to check integrity of RNA.

cDNA library construction and sequencing

Total RNA sample was treated with Truseq stranded total RNA ribo zero kit (Illumina make) to remove the ribosomal RNA followed by fragmentation. First strand cDNA followed by second strand generation was synthesized from fragmented m-RNA. Adapter ligated libraries were prepared as per standard protocol of Truseq stranded total RNA ribo zero library preparation kit (Illumina make). Adapters were ligated to 5'–3' of cDNA. The PCR amplified libraries were analyzed in bioanalyzer 2100 (Agilent Technologies) using high sensitivity DNA chip. The resulting libraries were sequenced using Illumina Hiseq 2500 platform (Illumina Inc., San Diego, CA, USA) (2 × 150 PE) using standard protocol.

Sequence analysis

Image data acquired from cDNA sequencing were converted to the analogous nucleotide sequence data. Raw reads were first processed through in-house perl scripts. Raw reads were percolated to exclude reads containing adaptors and

low-quality reads with Trimmomatic software-0.36 at Min Phred Score (QV) 20 (Bolger et al. 2014) to generate high-quality data. All the downstream analyses were carried out using clean data with high quality. All the data have been deposited in the Short Read Archive (SRA) at the NCBI database with bio-sample ID: SAMN07528737, sample name: GSLFDGS03 and accession ID: SRR5965322.

Screening, clustering and functional annotation of assembled transcript

De novo assembly of high quality reads was achieved using Trinity software-2.1.1 (Haas et al. 2013) at default parameter (kmer 25) and no reference sequence. Further assessment of transcript length distribution was carried out according to its length. Functional and comparative analyses of assembled unitranscripts were also performed with the most closely related species. Unigenes were predicted using CD-HIT-4.6.1 (Fu et al. 2012) package while CD-HIT-EST executable was used to remove the shorter redundant transcripts. Transdecoder-2.0.1 (Haas et al. 2013) predicted coding region (CDS) from unigenes sequence at default parameters with the encoded protein length set to a minimum of 100 amino acids and homology search with Swiss-prot and Pfam databases. The predicted CDSs were subjected to similarity search against NCBI's non-redundant (nr) database using the BLASTP algorithm. Protein sequence similarity search against uniprot, KOG and Pfam databases was carried out for functional annotation. CDSs were searched against all the transcription factor protein sequences at plant transcription factor database PlantTFDB (Guo et al. 2008) using BLASTP with an *E* value cut-off of < 1E–10. The GO mapping (Young et al. 2010) provides ontology of defined terms representing gene products.

Results

Total RNA quality testing

The quality of total RNA extracted from the collected leaves of DGS-3 was tested. Sample showed intact 28S, 18S and 5S RNA bands and RNA integrity number (RIN) was 6.8 which met the requirements for library construction and sequencing.

Sequencing data and de novo assembly

The mean of the library fragment size distributions was 552 bp. The libraries were sequenced using 2 × 150 PE chemistry on Illumina platform which generated 4.28 Gb data. After transcriptome sequencing, a total of 28,501,582 raw reads were obtained from cDNA libraries.

Trimmomatic-0.36 was used for filtration of data at QV 20 and for removal of adapter contamination. From these raw reads, 27,580,204 high-quality reads were obtained after filtering impurities. Sequencing analysis of cDNA libraries produced 230,510 transcripts with maximum transcript length of 51,747 bp (Table 1). Unigenes with length > 300 bp were higher as compared to the unigenes falling in the range of 1000–5000 bp. The unigene length distribution is shown in Fig. 1. Total of 193,615 unigenes with maximum length of 51,747 bps and N50 of 806 bps were predicted using CD-HIT package from transcripts.

Similarity search and functional annotation

A total number of 58,274 CDSs were predicted from unigenes sequences with mean of 730 bp CDS length using Transdecoder. The CDSs were searched against NCBI's non-redundant (nr) database using the BLASTP algorithm. The maximum numbers of the protein hits were found to be against the species *Coffea canephora* (Fig. 2). Among 58,274 total proteins, 49,055 gave similarity to NCBI's non-redundant (nr) database while 9219 proteins did not show similarity. This may be due to novel genes, which

perform particular plant-specific function or highly diverse genes. Similarity search against non-redundant (nr) database and several other databases, namely, Uniprot, KOG and Pfam also provided an insight into the complex metabolic pathways and regulatory networks. These databases used BLASTP algorithm with an *e* value threshold of $1e-5$. NCBI's nt database was also used for similarity search of CDSs. Combining the similarity searches from all the protein databases, we identified total of 58,274 hits.

KOG classification and transcription factor analysis

KOG analysis showed that the most enriched KOG category was “Signal transduction mechanisms” followed by “General function prediction”, “Posttranslational modification”, “protein turnover” and chaperones (Fig. 3). Pfam analysis showed most plentiful domains were “Protein kinase domain” followed by “Protein tyrosine kinase”, “NB-ARC domain” and “Reverse transcriptase”. CDS annotation in different databases represented in the form of Venn diagram (Fig. 4). TF families classified by plant transcription factor database PlantTFDB and graphical distribution showed 58 transcription factor families (Fig. 5). The most abundant transcription factor families enriched were bHLH followed by NAC, MYB-related and ERF.

GO annotation

Gene ontology (GO) assignments were used to classify the functions of the predicted CDS. The GO mapping provides ontology of defined terms which are grouped into three main domains: biological process, molecular function and cellular component. Blast2GO software was used to predict CDSs for gene ontology (Conesa and Gotz 2007). The classifications of GO functions were carried out using the Web Gene Ontology Annotation Plot (WEGO) program (Ye et al. 2006). A total of 16,569 CDSs distributed into gene

Table 1 Statistics of reads and assembly data in *G. sylvestre* genotype

Features	
PE reads	13,790,102
Total reads	27,580,204
Total data	3,963,508,325
Total no of transcripts	230,510
Maximum transcript length (bp)	51,747
Total no. of unigenes	193,615
Unigene N50 (bp)	806
Maximum unigene length (bp)	51,747
Total no. of CDS	58,274

Fig. 1 Unigene length distribution in *G. sylvestre* genotype

Fig. 2 Top hit species distribution of *G. sylvestre* genotype

ontology among which 10,968 CDSs were categorized under biological process while 8014 and 13,264 CDSs were categorized into cellular components and molecular function, respectively (Fig. 6).

KEGG pathway analysis

The annotated CDSs were enriched into five major pathways in KEGG database. The pathways with most enriched CDSs was “Metabolism” (3232) followed by “Genetic information processing” (1949), “Cellular processes” (975), “Environmental information processing” (879) and “Organismal systems” (214). Metabolism of terpenoids and polyketides revealed 149 CDSs while 150 CDSs of other secondary metabolites were annotated from transcriptome data. There were 37 and 13 CDSs assigned to steroid biosynthesis pathway and steroid hormone biosynthesis pathway, respectively. Identification of transcripts related to enzymes of both terpenoid sub-pathways in *Calotropis procera* suggested that precursors from these pathways are utilized in synthesis of various intermediates (Pandey et al. 2016).

Identification of candidate genes and putative biosynthetic pathway of polyoxypregnane glycoside biosynthesis in *Gymnema sylvestre*

Formation of pregnane glycosides is supposed to be derived from mevalonic acid via triterpenoid and phytosterols intermediates. Incorporation of ¹⁴C-mevalonic acid into steroid

part of digitoxin suggests preferred route for genin unit formation (Ramstad and Beal 1960). Total 27 probable genes encoding enzymes in the pregnane biosynthesis pathway were identified using transcriptome data in *G. sylvestre* (Fig. 7). The largest group of enzymes having important role in steroidal biosynthesis in *G. sylvestre* was CYPs followed by the glucosyltransferases (GTs). A total number of 109 CYPs and 53 GTs were annotated against nr database. About 235 CDSs were represented 27 genes of pregnane glycoside pathways in *G. sylvestre*. Pregnane undergoes various modifications to yield polyoxypregnane glycosides. A total of 19 CDSs, representing transcripts of cycloartenol synthase, sterol 24-C-methyltransferase, cycloeucalenol cycloisomerase, Δ^{14} -sterol reductase, C-8,7 sterol isomerase, sterol methyltransferase 2, C-5 sterol desaturase, sterol Δ^7 reductase, Δ^{24} -sterol reductase, 3 β -hydroxysteroid dehydrogenase, progesterone 5 β -reductase (5 β POR) were identified as members of pregnane biosynthesis pathway enzymes in *G. sylvestre*. Details of sequence length of these CDSs and their hit accession and maximum similarities with NCBI's database repositories are presented in Table 2.

Discussion

The total RNA with intact 28S, 18S and 5S RNA bands and RIN of 6.8 showed good quality RNA which was used for library construction and sequencing. A large number of raw reads were obtained from cDNA libraries accounting

Fig. 3 KOG classification of predicted CDSs of *G. sylvestre* genotype

to 4.28 Gb data. The filtration of data performed at QV 20 ensured that the data from the given run had probability of incorrect base call of 1 in 100. Filtering of impurities eliminated only 3.2% reads and thus, the remaining number of reads were of high quality having 99% base accuracy and thus, suitable for assembly preparation followed by transcriptome analysis. From the de novo assembly, shorter redundant transcripts were removed and 84% of the transcripts were predicted as unigenes. The transcriptome had average 42% GC contents. Thus, all parameters like quality RNA, large number of high quality reads and a good conversion ratio of transcripts to unigenes all together indicated a high quality of de novo transcriptome used in this study.

Through the discovery of biosynthetic genes involved in phytosterol pathways in some plants, it is revealed that the biosynthesis of polyoxypregnane glycosides involves a complex enzymatic network producing several products. Formation of squalene through MEP or MVA pathways holds the beginning of steroid backbone through the formation of several intermediates of steroidal frame work and ends up in pregnenolone (Benveniste 2002). The de novo assembly

and CDSs prediction in *G. sylvestre* resulted in the identification of 58,274 CDSs with an average length of 730 bp of which 84.17% had high sequence similarities (> 60%) with the known protein from different databases.

Through transcriptomic data, we identified various enzymes involved in pregnane metabolism pathway in *G. sylvestre* (Fig. 7). The initial products isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) required for the formation of triterpenoids and pregnane glycosides are produced through two different pathways. The mevalonate (MVA) pathway is being operated in cytosol whereas the 2-C-Methyl-D-Erythritol 4-Phosphate (MEP) pathway is being operated in the plastid in plants. As per the transcriptome data analysis, the MVA and the MEP pathway both are operational in *G. sylvestre*. The MVA pathway also known as isoprenoid pathway is essential for the biosynthesis of sterols, sesquiterpenes and triterpenoids (Crowell and Huizinga 2009), but monoterpenes and diterpenes are synthesized through the MEP pathway. The MVA pathway starts with acetyl CoA and the MEP pathway starts with D-glyceraldehyde 3-phosphate and both pathways end up

Fig. 4 Venn diagram representing the annotation of predicted CDSs to different databases

Transcription Factor Families Distribution

Fig. 5 Graphical distribution of transcription factor families in *G. sylvestre* genotype

Fig. 6 Go distribution using WEGO in *G. sylvestre* genotype

with the production of IPP and its isomer DMAPP (Crowell and Huizinga 2009; Campbell et al. 1998). Two different operational modules to produce same isoprenoids are shown to provide a swift production of specific end products when required besides separate isoprenoid synthesis for specific functions like phytosterol versus chlorophyll synthesis in different plants (Kalariya and Minipara 2018).

Because exact structure of cardenolides in *G. sylvestre* is not reported and genomic data are also not available in this respect, we restricted this pathway up to pregnane formation only. This transcriptome analysis provided an important resource for future functional genomics studies in *G. sylvestre*. Using transcriptome data, putative biosynthesis pathways of pregnane glycosides have been identified from important medicinal plants, viz., *Calotropis procera* (Pandey et al. 2016), *Marsdenia tenacissima* (Zheng et al. 2014) and a flowering plant, viz., *Digitalis purpurea* (Wu et al. 2012) which could be correlated with polyoxypregnane glycoside pathway in *G. sylvestre*. Most components needed for progesterone biosynthesis are common in plants (Gershenzon and Kreis 1999; Kreis and Müller-Uri 2013); however, variation in progesterone content in different tissues and organs is reported (Iino et al. 2007). It shows a ubiquitous existence

of progesterone biosynthesis in various plant families like Ruscaceae, Hyacinthaceae, Malvaceae, Ranunculaceae, Moraceae, Fabaceae, Apocynaceae and Plantaginaceae (Luckner and Wichtl 2000; Bauer et al. 2010; Agrawal et al. 2012).

The substrates cholesterol and phytosterols like sitosterol are products of different oxidosqualene cyclases, the lanosterol synthases and cycloartenol synthases, respectively; however, lanosterol is not synthesized in plants (Furuya et al. 1971; Bode et al. 2003). The cycloartenol synthase is the first committed enzyme in steroid biosynthetic pathway in plants. Initial steps involved in the removal of the methyl groups at C-4 and C-14 are crucial and leads to conversion of cycloartenol to sterols. Formation of cycloeucalenol from cycloartenol is a multi-step NADPH-dependent process. Alkylation of the side chain is catalyzed by S-adenosylmethionine (AdoMet)-sterol-C-methyltransferases (SMTs). A plant sterol C-methyltransferase was able to perform second methylation step during plant sterol biosynthesis pathway leading to high production of sitosterol at the cost of campesterol (Benveniste 2004; Husselstein et al. 1996). Thus, the SMT2 is functionally different than the SMT1 which is responsible for conversion of cycloartenol to cycloeucalenol

Table 2 Details of sequence length hit accession and maximum similarities with NCBI's database repositories of CDSs of important genes of polyoxypregnane glycosides pathway in *G. sylvestre*

Pathway enzymes		Sequence name	Seq. length	KEGG ID	Hit ACC	Similarity
SMT1	Sterol 24-C-methyltransferase [EC:2.1.1.41]	GS_39893	348	K00559	EYU27036	92.5
SMO	Squalene monooxygenase [EC:1.14.14.17]	GS_37468	145	K00511	KYP54510	95.5
		GS_37469	530	K00511	XP_018851103	89.5
		GS_37470	136	K00511	XP_015956180	97.9
		GS_39309	278	K08246	XP_008239448	93.0
CPI 1	Cycloeucalenol cycloisomerase [EC:5.5.1.9]	GS_39310	168	K08246	XP_008239448	93.0
		GS_2294	490	K05917	CAL69911	94.9
CYP51G1	Sterol 14-demethylase [EC:1.14.13.70]	GS_2298	490	K05917	CAL69911	95.1
		GS_2301	490	K05917	CAL69911	95.1
		GS_2305	490	K05917	CAL69911	94.9
		GS_37643	250	K00222	XP_010278239	92.4
FK	Delta14-sterol reductase [EC:1.3.1.70]	GS_34408	223	K01824	XP_011080546.1	73.0
HYD1	C-8,7 sterol isomerase [EC:5.3.3.5]	GS_45418	236	K08242	KZV57513	95.3
SMT2	24-Methylenesterol-C-methyltransferase 2 [EC:2.1.1.143]	GS_45420	360	K08242	XP_011088645	93.9
		GS_28296	293	K00227	XP_016560770	91.9
STE	Delta7-sterol 5-desaturase [EC:1.14.19.20]	GS_51488	271	K00213	XP_008342642	90.4
DWF5	7-Dehydrocholesterol reductase [EC:1.3.1.21]	GS_51489	100	K00213	XP_011072766	74.4
		GS_51491	435	K00213	XP_010089172	94.0
		GS_51493	174	K00213	EYU21205	91.7
		GS_51495	100	K00213	XP_011072766	74.4
		GS_51496	174	K00213	EYU21205	91.7
DWF1	Delta24-sterol reductase [EC:1.3.1.72 1.3.1.-]	GS_39796	564	K09828	XP_011096752.1	91.0
HSD	3beta-Hydroxysteroid dehydrogenase decarboxylase [EC:1.1.1.145 5.3.3.1]	GS_28539	284	K00070	XP_019244829	91.2
		GS_34760	244	K00070	AIK35198	88.5
PBR	Progesterone 5-beta-reductase [EC:1.3.1.3]	GS_397	389	K22419	ADG56540	91.5
		GS_398	268	K22419	ADG56540	91.0

through the first methylation step leading to sterol biosynthesis at the cost of triterpenoid saponins. The opening of cyclopropane ring of cycloeucalenol leading to production of obtusifoliol in *Rubus fruticosus* tissue is reported as early as in 1974 (Heintz and Benveniste 1974). Interestingly, cyclopropyl sterol isomerase (CPI)-mediated isomerization of cycloeucalenol to obtusifoliol is reported in plant kingdom only. In higher plants, methyl group removal is sequential in which C-4 methyl is removed first followed by C-14 methyl whereas in animals and fungi, the 14- α methyl group is the first among three methyls to be removed. Obtusifoliol 14 α -demethylase (CYP51) catalyzes 14- α demethylation of obtusifoliol to 4- α -methyl-5 α -ergosta-8,14,24(28)-trien-3 β -ol known as $\Delta^{8,14}$ sterols. Sterol C-14 reductase (FK) reduces $\Delta^{8,14}$ sterols in to 4 α -methyl fecosterol. Fecosterol is known as 24-methylene-5 α -cholest-8-en-3 β -ol or $\Delta^{8(24),28}$ -ergostadienol. C-8,7 sterol isomerase (HYD-1) catalyzes isomerization reaction between 5 α -cholest-8-en-3 β -ol and 5 α -cholest-7-en-3 β -ol. 4 α -Methyl fecosterol is converted to 24-methylene lophenol by isomerization. The biosynthetic pathway from cycloartenol to sterols and

brassinosteroids diverges at 24-methylenelophenol, in which Δ^7 -sterol 5-desaturase (STE) leads to brassinosteroids formation via intermediate campesterol and 24-methylenesterol C-methyltransferase (SMT2) leads to pregnane formation via intermediate sitosterol (Schaeffer et al. 2000). In higher plants, Sterol Δ^7 reductase (DWF5) reduces Δ^7 -double bond leading to conversion of $\Delta^{5,7}$ -sterols into Δ^5 -sterols. Desaturase removed hydrogen from avenosterol which is Δ^7 sterols to produce 5-dehydroavenosterol. DWF5 converted 5-dehydroavenosterol to Δ^5 -avenosterol known as isofucosterol. 24-Methylene cholesterol and isofucosterol are the substrates for Δ^5 -Sterol Δ^{24} -reductase/isomerase (DWF1). In higher plants, DWF1 isomerized $\Delta^{24(25)}$ sterols prior to reduction. Through our transcriptomic study, the putative enzyme DWF1 responsible to reduce isofucosterol to sitosterol is identified. Sitosterol is a precursor of stigmasterol and ratio of these two sterols is important for plant adaptation to environment.

The conversion of pregnenolone into progesterone is completed in two steps. In the first reaction, the Δ^5 -pregnane-3-one is produced through oxidation of the 3 β -hydroxy group

by the Δ^5 -3 β -hydroxysteroid dehydrogenase (3 β -HSD) which is NAD-dependent. In the second step, the double bond is shifted from position 5 to position 4. To connect rings A and B of the steroid structure in *cis* configuration, it requires progesterone 5 β -reductase (5 β -POR) which leads to the transformation of progesterone into 5 β -pregnane-3,20-dione and thus, this is one of the most important enzyme imparting structural modification in the pregnane biosynthesis pathway. Thorn et al. (2008) fully characterized the crystal structure and found that the progesterone reductase from *D. lanata* defines a novel class of short-chain dehydrogenases/reductases.

CYPs are mediators of reactions involved in the synthesis of large variety of secondary metabolites in plants (Pandey et al. 2016). Glycosylation of free sterols into their respective glycosides is catalyzed by nucleotide-dependent sterol glycosyltransferases (SGTs) that are classified as family one of the 94 families of SGTs. Thus, the process of esterification indirectly regulates the pool of free sterols in plants (Bouvier-Navé and Benveniste 1995). Annotated CDSs of putative enzymes in *G. sylvestre* outlined the pregnane biosynthesis with the presence of monooxygenases, cytochrome P450, and glycosyltransferase essentially required for the formation of polyoxypregnane glycosides.

Structurally, pregnanes are C21 steroidal compounds with sugar moiety linked to an alcoholic hydroxyl group of the aglycon part. They are reported to be found in nature either in the free state or as glycosides. The alcoholic hydroxyl group of the pregnane aglycon most frequently linked at C-3, C-20 or both through an acetal linkage, but the sugar moiety also reported to be linked to hydroxyl functions at C-2, C-4 or C-21. Yoshikawa et al. (1999) reported 11 pregnane glycosides, that are gymnepregosides G–Q from the root of *Gymnema alternifolium*. Ten novel C21 pregnane glycosides, extensumside C–L (1–10), were isolated from the edible pericarps of *Myriopterion extensum* have aglycone 3 β ,16 α -dihydroxy-pregan-5-en-20-one and contained the deoxysugar chain and the glucose chain which were linked to C-3 and C-16 of the aglycone, respectively (Sun et al. 2016).

Conclusion

A putative pathway of polyoxypregnane glycosides was identified using transcriptome data in *G. sylvestre*. Total 58,274 CDSs were predicted from 193,615 unigenes and KEGG pathway analysis revealed 235 CDSs representing 27 genes of pregnane glycosides pathways and thus, provided information on key enzymes of pregnane glycoside biosynthesis. Future research on over expression through genetic engineering or physiological interventions of these key enzymes will provide scope of higher production of pregnane glycosides having potential medicinal demands.

Acknowledgements We acknowledge the funding through FAP Scheme 2015–16, sanctioned dated 23.09.2016 from GSBTM, Govt. of Gujarat, Gujarat, India, the ICAR-DMAPR, Anand and the ICAR, New Delhi for providing the basic facilities for this research work, all germplasm explorers who collected this germplasm and all curators who maintained this genotype at our research farm.

Author contributions KK conceived the project, designed the experiment, analyzed the data and drafted the MS. DM analyzed the data and drafted the MS. PM explored the germplasm used, gave overall technical and administrative support in this study.

Compliance with ethical standards

Availability of data and materials The data generated or analyzed during this study are included in this published article, its supplementary information files, and publicly available repositories. The transcriptome raw data are deposited at NCBI under submission ID: SUB2977090, sample ID: GSLFDGS03 and SRA accession ID: SRR5965322.

Conflict of interest The authors declare no competing financial interests.

References

- Agrawal AA, Petschenka G, Bingham RA, Weber MG, Rasmann S (2012) Toxic cardenolides: chemical ecology and coevolution of specialized plant–herbivore interactions. *New Phytol* 194:28–45
- Anupam B, Malay CH (1994) Hypolipidaemic and antiatherosclerotic effects of oral *Gymnema sylvestre* R. Br. leaf extract in albino rats fed on a high fat diet. *Phytother Res* 8:118–120
- Bauer P, Munkert J, Brydziun M, Burda E, Müller-Uri F, Gröger H et al (2010) Highly conserved progesterone 5 beta-reductase genes (P5 beta R) from 5 beta-cardenolide-free and 5 beta-cardenolide-producing angiosperms. *Phytochemistry* 71:1495–1505
- Benveniste P (2002) Sterol metabolism. In: Somerville CR, Meyerowitz EM (eds) *The arabidopsis book*, vol 1. American Society of Plant Biologists, Rockville
- Benveniste P (2004) Biosynthesis and accumulation of sterols. *Annu Rev Plant Biol* 55:429–457
- Bode HB, Zeggel B, Silakowski B, Wenzel SC, Reichenbach H, Müller R (2003) Steroid biosynthesis in prokaryotes: identification of myxobacterial steroids and cloning of the first bacterial 2,3(S)-oxidosqualene cyclase from the myxobacterium *Stigmatella aurantiaca*. *Mol Microbiol* 47:471–481
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120
- Bouvier-Navé P, Benveniste P (1995) Sterol acyl transferase and sterol ester hydrolase activities in a tobacco mutant which overproduces sterols. *Plant Sci* 110:11–19
- Campbell M, Hahn FM, Poulter CD, Leustek T (1998) Analysis of the isopentenyl diphosphate isomerase gene family from *Arabidopsis thaliana*. *Plant Mol Biol* 36:323–328
- Conesa A, Gotz S (2007) Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int J Plant Genom* 2008:1–12
- Crowell DN, Huizinga DH (2009) Protein isoprenylation: the fat of the matter. *Trends Plant Sci* 14:163–170
- Ernst M, Padua RM, Herl V, Müller-Uri F, Kreis W (2010) Expression of 3 β HSD and P5 β R, genes respectively coding for Δ^5 -3 β -hydroxysteroid dehydrogenase and progesterone 5 β -reductase, in leaves and cell cultures of *Digitalis lanata* EHRH. *Planta Med* 76(9):923–927

- Fu L, Niu B, Zhuy Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28(23):3150–3152
- Furuya T, Kawaguchi K, Hirotani M (1971) Biotransformation of progesterone and pregnenolone by plant suspension cultures. *Phytochemistry* 10:1013–1017
- Gershenzon J, Kreis W (1999) *Biochemistry of plant secondary metabolism: annual plant reviews*, vol 2. Sheffield Academic Press, Sheffield
- Guo AY, Chen X, Gap G, Zhang QH, Liu XC, Zhong YF, GU X, He K, Luo J (2008) PlantTFDB: a comprehensive plant transcription factor database. *Nucleic Acids Res* 36:966–969
- Gupta SS, Seth CB, Variyar MC (1962) Experimental studies on pituitary-diabetes. Part I. Inhibitory effect of a few Ayurvedic antidiabetic remedies on anterior pituitary extract induced hyperglycemia in albino rats. *J Med Res* 50:73–81
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Le Duc RD, Friedman N, Regev A (2013) De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat Protoc* 8(8):1494–1512
- Heintz R, Benveniste P (1974) Enzymatic cleavage of the 9,19-cyclopropane ring of cyclopropyl sterol in bramble tissue cultures. *J Biol Chem* 249:4267–4274
- Husselstein T, Gachotte D, Desprez T, Bard M, Benveniste P (1996) Transformation of *Saccharomyces cerevisiae* with a cDNA encoding a sterol methyltransferase from *Arabidopsis thaliana* results in the synthesis of 24-ethyl sterols. *FEBS Lett* 381:87–92
- Iino M, Nomura T, Tamaki Y, Yamada Y, Yoneyama K, Takeuchi Y et al (2007) Progesterone: its occurrence in plants and involvement in plant growth. *Phytochemistry* 68:1664–1673
- Kalariya KA, Minipara D (2018) An overview of triterpenoid biosynthesis in plants and structural depiction of gymnemasides and gymnemosides from *Gymnema sylvestre*. *J Plant Physiol Pathol* 5:6. <https://doi.org/10.4172/2329-955X.1000174>
- Komarnytsky S, Esposito D, Rathinasabapathy T, Poulev A, Raskin I (2013) Effects of pregnane glycosides on food intake depend on stimulation of the melanocortin pathway and bdnf in an animal model. *J Agric Food Chem* 61(8):1841–1849
- Kreis W, Müller-Uri F (2013) Cardenolide aglycone formation in *Digitalis*, isoprenoid synthesis in plants and microorganisms. Springer, New York, pp 425–438
- Luckner M, Wichtl M (2000) *Digitalis*. WVG, Stuttgart. ISBN 3804717055
- Pandey A, Swarnkar V, Pandey T, Srivastava P (2016) Transcriptome and Metabolite analysis reveal candidate genes of the cardiac glycoside biosynthetic pathway from *Calotropis procera*. *Sci Rep* 6(34464):1–14
- Ramstad E, Beal JL (1960) Mevalonic acid as a precursor in the biogenesis of digitoxigenin. *J Pharm Pharmacol* 12:552–556
- Schaeffer A, Bouvier-Navé P, Benveniste P, Schaller H (2000) Plant sterol-C24-methyl transferases: different profiles of tobacco transformed with SMT1 or SMT2. *Lipids* 35(3):263–269
- Sun G, Dai Q, Zhang H, Li Z, Du Z (2016) New sweet-tasting C21-pregnane glycosides from pericarps of *Myriopterion extensum*. *J Agric Food Chem* 64(49):9381–9389
- Thorn A, Egerer-Sieber C, Jäger CM, Herl V, Müller-Uri F, Kreis W, Müller YA (2008) The crystal structure of progesterone 5-reductase from *Digitalis lanata* defines a novel class of short chain dehydrogenases/reductases. *J Biol Chem* 283:17260–17269
- Wu B, Li Y, Yan H, Ma Y, Luo H, Yuan L, Chen S, Lu S (2012) Comprehensive transcriptome analysis reveals novel genes involved in cardiac glycoside biosynthesis and lncRNAs associated with secondary metabolism and stress response in *Digitalis purpurea*. *BMC Genom* 13:15
- Xu R, Yang Y, Zhang Y, Ren F, Xu J, Yu N, Zhao Y (2015) New pregnane glycosides from *Gymnema sylvestre*. *Molecules* 20:3050–3066
- Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J (2006) WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res* 34:293–297
- Yoshikawa K, Matsuchika K, Takahashi K, Tanaka M, Arihara S, Chang HC, Wang JD (1999) Pregnane glycosides, gymnepregosides G–Q from the roots of *Gymnema alternifolium*. *Chem Pharm Bull (Tokyo)* 47(6):798–804
- Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* 11:1–12
- Zheng K, Zhang G, Jiang N, Yang S, Li C, Meng Z, Guo Q, Long G (2014) Analysis of the transcriptome of *Marsdenia tenacissima* discovers putative polyoxypregnane glycoside biosynthetic genes and genetic markers. *Genomics* 104:186–193