Chapter 5: Discussion

Secondary metabolites, also known as natural products are organic compounds which do not have direct function in growth and development. These compounds do not have direct roles in respiration, photosynthesis, protein synthesis, translocation, solute transport, nutrient assimilation, differentiation, or the formation of carbohydrates, proteins and lipids (Croteau et al., 2000; Taiz and Zeiger, 2002). Unlike, primary metabolites which are found throughout the plant kingdom, a particular type of secondary metabolite is found only in a related group of species (Croteau et al., 2000; Taiz and Zeiger, 2002).

Plants produce a vast number of secondary metabolites which help them to interact with the environment and stay fit. These metabolites protect the host against herbivorous insects and pathogens, attract the organisms for pollination and seed dispersal and help them to adapt to various abiotic stresses (Hartmann, 2007; Zhao et al., 2013). Some of these metabolites have also been beneficial to humans and are being harnessed from nature for food ingredients, health supplements and for medicines to treat various ailments. These plant secondary metabolites are classified into three major groups: terpenes, phenolics and nitrogen-containing compounds and are biosynthesized through various biosynthetic pathways like shikimic acid pathway, malonic acid pathway, mevalonate (MVA) pathway, methylerythritol phosphate (MEP) pathway, etc (Croteau et al., 2000; Taiz and Zeiger, 2002; Zhao et al., 2013). Thus understanding the biosynthesis of these metabolites is important for conservation and utilization of plant resources using biotechnological approaches.

P. sokpayensis is a newly reported *Panax* species from Sikkim Himalaya, which contains major ginsenosides in its rhizome (Gurung et al., 2018). Ginsenosides are triterpene saponins with different pharmacological activities such as anti-aging, anti-inflammatory, anticancerous, antihypertensive, antiatherosclerotic, immunomodulatory, antidiabetic and antistress functions (Christensen, 2009). Due to the presence of ginsenosides, several species of *Panax* are in high demand in market making them one of the highly traded medicinal plants (Baeg and So, 2013). Ginsenoside biosynthesis has been studied in various *Panax* species like *P. ginseng*, *P. notoginseng*, *P. quinquefolius* and *P. japonicus* (Table 2.3, 2.4, chapter 2). Recently, ginsenoside biosynthetic pathway genes were reported from *P. vietnamensis* var *fuscidiscus*, which contain higher content of ocotillol type saponin, majonoside R_2 (Zhang et al., 2015a). However, such information is completely absent in case of *P. sokpayensis*, a new species reported from Sikkim Himalaya. Therefore the main

objective of present thesis is to generate information on genes involved in ginsenoside biosynthesis.

5.1. Major ginsenosides quantification in *Panax sokpayensis*

Due to high medicinal value of *P. sokpayensis*, its rhizomes are used in various preparations by local traditional healers for domestic and commercial use (Gurung et al., 2016). However, there is no scientific validation of the chemical constituents in *P. sokpayensis*. Hence, the need was felt to investigate the presence of major ginsenosides (Rb1, Rb2, Rc, Rd, Re, Rf, Rg1 and Rg2) in this newly reported species of *Panax* from Sikkim Himalaya. To the best of our knowledge, this is the first report on *in planta* quantification of major ginsenosides in any *Panax* species from Sikkim Himalaya, India (Gurung et al., 2018). Previous ginsenoside quantification study reported from the Sikkim Himalaya has been of the cell suspension line of *P. sikkimensis* (synonym of *P. bipinnatifidus*) (Biswas et al., 2015). Previously, these major ginsenosides were quantified in the rhizomes of *P. notoginseng* (Wang et al., 2014a), *P. ginseng* (Wang et al., 2016a; Zhang et al., 2014) and *P. quinquefolius* (Kochan et al., 2008).

Among the major ginsenosides, viz., Rb1, Rb2, Rc, Rd, Re, Rf, Rg1 and Rg2, all were detected in the rhizome of 10 years old *P. sokpayensis*, except Rc. Oh et al. (2014) had reported that ginsenoside Rc increased in the roots of hydroponically grown *P. ginseng* after treatment with methyl jasmonate (MJ) and decreased when exposed to low temperatures. Rc was also produced in *P. quinquefolius* callus cultures kept in dark (Kochan and Chmiel, 2013). These outcomes suggested that ginsenoside Rc is differentially accumulated during specific growth period and stress conditions in different *Panax* species. The ginsenoside Rf has been used to differentiate *P. notoginseng*, *P. quinquefolius* and *P. ginseng* (Li et al., 2000). The quantitiy of ginsenoside Rb2 is higher in *P. ginseng* in comparison to *P. quinquefolius* and *P. notoginseng* (Sun et al., 2011). These findings suggest that the presence of both Rf and Rb2 in *P. sokpayensis* is a value addition to this species.

Previously, eight ginsenosides (Re, Rc, Rb1, Rf, Rb2, Rd, Rg1, Ro) have been quantified in the rhizomes of 10 years old *P. ginseng* (Zhang et al., 2014). The total sum of these major ginsenosides was comparable with the total ginsenoside amount of *P. sokpayensis* (Table 4.1) except ginsenoside Ro and Rc. Since the bioactive properties of total ginsenosides vary with the variations in the Rb1 to Rg1 ratio (ginsenoside quality markers), which may also be useful in differentiating *P. ginseng*, *P. notoginseng* and *P.*

quinquefolius, Sun et al. (2011) suggested that the Rb1:Rg1 values for *P. notoginseng* and *P. ginseng* should lie between 1 and 3 and should be around 10 or higher for *P. quinquefolius*. In the current study, Rb1:Rg1 ratio was 2.13 in *P. sokpayensis* (Table 4.1), which is considered as characteristic of Asian ginseng.

All the above findings suggest that *P. sokpayensis* is at par with its Asian congener, *P. ginseng* and could be a ginseng resource for commercialization from Sikkim Himalaya.

5.2. Identification of differentially expressed genes

There are several molecular biological techniques to study the differential expression of genes during specific developmental stage or physiological condition (Carulli et al., 1998; Byers et al., 2000; Casassola et al., 2013). Complementary DNA (cDNA) libraries (Lambert and Williamson, 1993; Ying, 2004), cDNA amplified fragment length polymorphism (cDNA-AFLP) analysis (Bachem et al., 1996), suppression subtractive hybridization (SSH) (Diatchenko et al., 1996), microarray (Schena et al., 1995), differential display (DD) (Liang and Pardee, 1992), representational difference analysis (RDA) (Lisitsyn et al., 1993), serial analysis of gene expression (SAGE) (Velculescu, et al., 1995) and RNA sequencing (RNA-seq) (Finotello and Camillo, 2015; Wang et al., 2009) are some of the commonly used tools for such studies.

SSH has been used to fish out rare transcripts in the same tissue with different treatments or in the two different tissues of interest (Diatchenko et al., 1996) and could be an effective method to get rid of the common housekeeping genes and clone differential transcripts. It has been successfully used to identify differentially expressed genes in many plant species, namely *Corchorus capsularis* (Samanta et al., 2015), *Camellia sinensis* (Paul and Kumar, 2011), *Chlorophytum borivilianum* (Kumar et al., 2012b), *Azadirachta indica* (Narnoliya et al., 2014) and *Citrus sinensis* (Zhou et al., 2015).

5.2.1. Suppression Subtractive Hybridization (SSH) mediated identification and cloning of *FPS, SS* and *DS* involved in ginsenoside biosynthetic pathway

In *Panax* species, ginsenosides distributed in the leaf, root, flower and berry of ginseng plant vary both qualitatively and quantitatively. For instance, the underground parts of *P. notoginseng* are rich in protopanaxodiol and protopanaxatriol-type ginsenosides, while the flowers and leaves possess protopanxodiol-type saponins only. Moreover, Rb2, Rc and Rb3 are reported to be relatively abundant in aerial parts, compared to the parts below the soil in *P. notoginseng* (Wan et al., 2006). Higher ginsenosides levels have been reported in

the berries than in the roots of cultivated *P. ginseng* (Kim et al., 2009a). Total ginsenosides content has been found to be higher in the leaves of *P. quinquefolius* in comparison to their roots (Searels et al., 2013). These findings suggested the movement of ginsenosides during foliation from the leaf to the root or vice versa (Kim et al., 2014a). Hence it was speculated that the variations in the contents of ginsenosides might be due to tissue-specific differential gene expression in the above- and belowground tissues of ginseng plant besides other unknown mechanisms. The information regarding the differentially expressed genes related to ginsenoside biosynthetic pathways and their regulatory components which are tissue specific are scarce in *Panax* spp. Therefore to fish out these differentially expressed transcripts, SSH was performed between rhizome and leaf tissues of *P. sokpayensis*.

In the leaf SSH library, ESTs assembled into 80 unigenes (GenBank accession numbers JZ822892 – JZ822971, Table 4.3). *Galactinol synthase 2* was among the most abundant ESTs. *Galactinol synthase 1* and *galactinol synthase* were also detected. Galactinol synthase (GolS) is an important enzyme in raffinose oligosaccharides (RFOs) biosynthesis and plays a key role in carbon portioning between sucrose and RFOs in plants (Liu et al., 1998). It is a part of glycosyltransferase 8 family, which catalyzes the formation of galactinol from UDP galactose and myoinositol and play a role in plant growth and development. Other ESTs involved in growth and development, namely ribosomal plastocyanin (Schottler et al., 2004), glycolate oxidase protein (Sahu et al., 2010), RNA processing Brix domain protein (Fromont-Racine et al., 2003), ribulose 1,5 bisphosphate carboxylase oxygenase (Rubisco; Quick et al., 1991), cell division cycle 20.1 (Niu et al., 2015), ribonuclease t2 family protein (Luhtala and Parker, 2010) and metallothionein-like protein type 3 (Yu et al., 1998) were also abundant in the leaf SSH library. The abundant presence of these genes in the leaf SSH library suggested that the sampled plant was under high metabolic process.

Analysis of rhizome SSH, yielded 160 unigenes as detailed under section 4.2 (Chapter 4: Results) (GenBank accession numbers JZ822732 – JZ822891, table 4.4). ESTs from rhizome SSH library showed homology to three ginsenoside biosynthetic pathway genes, namely, *FPS*, *SS* and *DS*. Besides, ESTs representing KIAA0664 homologue protein were the most abundant. This is an uncharacterized protein showing sequence homology to eIF-3 (eukaryotic translation initiation factor-3) from other species. It has been suggested that KIAA0664 homologue protein plays a role in pathogenesis triggered by high-fat diet in mice (Luo et al., 2012). The abundance of this gene in the rhizome SSH library suggests its role in pathogenesis and defense response in *P. sokpayensis*. Other ESTs of genes with roles

in plant immunity and defense mechanism that were present in the rhizome library were major latex-like protein (Lytle et al., 2009), ubiquitin-activating enzymes e1 l (Marino et al., 2012), ribonuclease-like storage protein (Kim et al., 2004) and 40s ribosomal protein s26 (García-coronado et al., 2015). The *P. sokpayensis* rhizome remained underground for years. Thus these genes might be involved in plant immunity and defense response.

About 32% of the ESTs from both libraries were novel with no significant similarities to the sequences available in NCBI database. Our data reveals that some genomic information on *Panax* spp. is still missing despite serveral transcriptome studies. The probable reason for the unmapped unigenes of *P. sokpayensis* could be due to the presence of rare or low-quality singletons and incomplete coverage of genome as suggested for *C. borivilianum* (Kalra et al., 2013) and *A. indica* (Rajakani et al., 2014). Transcriptome and SSH studies have been compared to study the differential gene expression in bio-eroding sponge *Cliona varians* that harbors clade *G Symbiodinium* (Riesgo et al., 2014). These observations suggest that data otained using SSH in this study might be useful for complementing the available *Panax* transcriptomes.

ESTs showed homology to genes from several species (section 4.2.2, Figure 4.4, Chapter 4: results). Surprisingly, the top hit species in both the libraries are not from the genus *Panax*. The high number of ESTs showing homology to genes from species other than genus *Panax* may be due to the absence of reference genome of *Panax* species due to which many gene sequences may not be present in the database.

5.2.2. Gene ontology (GO) based classification of ESTs

GO annotation was done using Blast2GO software for global functional analysis of all the subtractive ESTs (Conesa et al., 2005). In the leaf SSH library, ESTs in the biological process category (Figure 4.5A) revealed that small molecule metabolic process (GO:0044281), oxidation-reduction process (GO:0055114), photosynthetic process (GO:0070271), response to radiation (GO:0009314) and generation of precursor metabolites and energy (GO:0006091) were among the highly represented processes suggesting that the plant was undergoing rapid growth and development by producing precursors for metabolite biosynthesis. In molecular function, ESTs representing genes pertaining to nucleic acid binding (GO:0003676), transferase activity, cation binding (GO:0043169), hydrolase activity acting on ester bonds (GO:0016788), transferring glycosyl groups (GO:0016757) were highly represented categories (Figure 4.5C). Major subgroups of cellular components that were represented included organelle lumen (GO:0043233), intracellular regions

(GO:0005622), envelop (GO:0031975) and viral capsid (GO:0019028) (Figure 4.5E). In the rhizome SSH library, classification of ESTs in biological process categories (Figure 4.5B) revealed that organic substance biosynthetic process (GO:1901576), macromolecule metabolic process (GO:0043170), ribonucleoprotein complex biogenesis (GO:0022613), carbohydrate metabolic process (GO:0005975) and cellular biosynthetic process (GO:0044249) were among the well represented processes hinting that the plant was undergoing cyclization of precursor molecules for metabolites biosynthesis. Nucleic acid binding (GO:0003676), anion binding (GO:0043168), nucleoside binding (GO:0001882) and nucleoside phosphate binding (GO:1901265) were the molecular functions represented by the subtractive rhizome ESTs (Figure 4.5D). ESTs of genes coding for proteins localized to cellular components like organelle lumen (GO:0043233), intracellular regions (GO:0005622), plasmodesma (GO:0009506) and cell periphery (GO:0071944) (Figure 4.5F) were present. These GO annotations provided a snapshot of all differentially expressed sequences encoding diverse proteins in *P. sokpayensis*.

5.2.3. Detection of protein domains using InterPro Scan

Further characterization of the functionally annotated ESTs was done through InterPro Scan analysis (Figures 4.6, 4.7). Nine protein domains were detected in the ESTs of leaf SSH library (Figure 4.7A), while 14 protein domains were found in the ESTs of rhizome SSH library (Figure 4.7B). Nucleotide-diphosphosugar transferases domain was detected in the largest number of ESTs from leaf SSH library. These domains are found in diverse types of glycosyltransferases which catalyze the transfer of sugar moieties from activated donor molecules to specific acceptor molecule during the biosynthesis of disaccharides, oligosaccharides, and polysaccharides. WD40/YVTN repeat-like-containing domain was also present in large number of ESTs. WD40 repeat proteins are members of a large conservative protein family and are key structural signatures in a large number of cellular regulators involved in transcriptional regulation, signal transduction and cell cycle control (Holm et al., 2001; Mishra et al., 2012). Photosynthesis associated protein domains like chlorophyll a/b binding protein domains, photosystem I PsaG/PsaK domain and photosynthetic reaction center were also represented in the leaf SSH library. In the rhizome SSH library, glycosyltransferase family 35 domains and ubiquitin activating enzyme repeat were highly represented. Ubiquitin-activating enzyme plays a key role in ubiquitin-26S proteasome system by coupling ATP hydrolysis and activation of C-terminal glycine of ubiquitin (Marino et al., 2012). This system has a role in plant immune signaling.

Glycosyltransferases, family 35 are enzymes with starch and glycogen phosphorylase activity and play a role in generating phosphorylated glucose molecule. Other protein domains like Myc-type basic helix-loop-helix (bHLH) domain, terpenoid cyclases/protein prenyltransferase alpha-alpha toroid, isoprenoid synthase domain and glyceraldehyde 3-phosphate dehydrogenase catalytic domain were also present in some ESTs. Presence of all these putative domains in the rhizome SSH library ESTs suggested the defensive roles of the corresponding proteins.

5.2.4. KEGG analysis of ESTs

On KEGG analysis, four pathways were represented by some of the ESTs of leaf SSH library. These were galactose metabolism, nitrogen metabolism, carbon fixation in photosynthetic organisms and glyoxylate and dicarboxylate metabolism pathways. It is intriguing that out of all the ESTs encoding galactinol synthases, only one EST (PsF92) was classified in galactose metabolism pathway (Figure 4.8A). These observations suggested that galactinol synthases might be involved in other glycosyltransferase activities as suggested previously (Goswami and Punja, 2008; Sengupta et al., 2012). Similarly, functionally annotated ESTs of the rhizome SSH library represented the following pathways: pentose and glucuronate interconversions, sesquiterpenoid and triterpenoid biosynthesis, terpenoid backbone biosynthesis, porphyrin and chlorophyll metabolism, thiamine metabolism, starch and sucrose metabolism, retinol metabolism, cysteine and methionine metabolism, steroid biosynthesis, riboflavin metabolism, purine metabolism and aminobenzoate degradation (Figure 4.8B). Representation of different pathways in the leaf and rhizome SSH libraries suggest the presence of different set of genes in the two libraries which in turn indicates different metabolism being active in the two tissues.

5.2.5. Other differentially expressed genes with possible roles in ginsenoside biosynthesis

ESTs with homology to GBR5-like gene was obtained in the leaf SSH library (Table 4.3). Previously, five genes, namely *GBR-1*, *GBR-2*, *GBR-3*, *GBR-4* and *GBR-5* have been isolated and characterized in *P. ginseng* (Luo et al., 2003) indicating their important role in the biosynthesis of ginsenosides. Another EST, corresponding to transcription factor, bHLH was detected in rhizome SSH library (Table 4.4). bHLH proteins bind as dimers to specific DNA target sites and are extensively characterized in nonplant eukaryotes for the regulation of diverse biological processes (Toledo-Ortiz et al., 2003). In *A. annua*, a MYC-type bHLH transcription factor has been shown to act as an activator of sesquiterpene synthase and

cytochrome P450 essential for the biosynthesis of artemisinin (Ji et al., 2014; Yu et al., 2012). In tomato, a WRKY and MYC-type bHLH transcription factors have been shown to bind the promoter of sesquiterpene synthase gene (Spyropoulou et al., 2014). In *A. thaliana*, it has been observed that the bHLH directly binds to sesquiterpene synthase promoters, effecting an enhanced release of volatile sesquiterpenoids (Hong et al., 2012). Recently, two bHLH transcription factors, TSAR1 and TSAR2 of *M. truncatula*, have been shown to activate triterpene saponin biosynthesis (Mertens et al., 2016). These findings suggest that *bHLH* may play important role of in the activation of ginsenoside biosynthetic pathway genes in *P. sokpayensis*.

5.3. Degenerate primers approach to clone *AACT, HMGS, MVK, PMVK, DXS, DXR, CMS, CMK, MCS, HDS, HDR, IDI, SE, β-AS, PPDS, PPTS* and *CS*

A degenerate primer is an oligonucleotide with partially degenerate sequence (Telenius et al., 1992). PCR using such degenerate primers allows the user to amplify related nucleic acid sequences as well as targets for which only amino acid sequences are available (Compton, 1990; Kwok et al., 1994; Telenius et al., 1992). In the current study, seventeen partial fragments were obtained using degenerate PCR (Figure 4.23, Table 4.5). These partial fragments were of the following genes: a*cetyl-CoA C-acetyltransferase* (*AACT*), *3-hydroxy-3-methylglutaryl coenzyme A synthase* (*HMGS*), m*evalonate kinase* (*MVK*), *phosphomevalonate kinase* (*PMVK*), *1-deoxy-D-xylulose-5-phosphate synthase* (*DXS*), *1-deoxy-D-xylulose 5-phosphate reductoisomerase* (*DXR*), *2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase* (*CMS*), *4-*(*cytidine 5'-diphospho*)*-2-C-methyl-D-erythritol kinase* (*CMK*), *2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase* (*MCS*), *4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase* (*HDS*), *4-hydroxy-3-methylbut-2-enyl diphosphate reductase* (*HDR*), *isopentenyl diphosphate isomerase 2* (*IDI2*), *squalene epoxidase* (*SE*), *β-amyrin synthase* (*β-AS*), *protopanaxadiol synthase* (*PPDS*), *protopanaxatriol synthase* (*PPTS*) and *cycloartenol synthase* (*CS*).

It is well established that plants possess both cytosolic mevalonate and plastidial non-mevalonate pathways which produce precursors for the isoprenoids (Hemmerlin et al., 2012). In case of genus *Panax*, majority of transcriptomic studies of different *Panax* species have reported MVA genes along with ginsenoside biosynthetic pathway genes (Cao et al., 2015; Chen et al., 2011; Li et al., 2013; Liu et al., 2015; Luo et al., 2011; Rai et al., 2016;

Subramaniyam et al., 2014; Sun et al., 2010a; Wang et al., 2016b; Wu et al., 2013; Zhang et al., 2015a; Zhang et al., 2017). Few transcriptomic studies have also reported genes from MEP pathway (Subramaniyam et al., 2014; Wang et al., 2016b; Zhang et al., 2017). Overall, these studies suggest the presence of both MVA and MEP pathways in *Panax* species. This was further corroborated by our present study which was successful in cloning several partial and full length genes from both MVA and MEP pathways from *P. sokpayensis*.

The partial sequences of MVA, MEP and ginsenoside biosynthetic pathway that were amplified using degenerate primer approach are discussed as follows:

5.3.1. *Acetyl-CoA C-acetyltransferase* (*AACT*)

AACT is the first enzyme of the MVA pathway which catalyzes the condensation of two acetyl-CoA molecules into acetoacetyl-CoA (Chen et al., 2017; Miziorko, 2011). Thus it is a key enzyme that channelizes metabolic flux from the primary metabolism towards secondary metabolism, i.e., towards isoprenoid biosynthesis. Though complete characterization of AACT belonging to *Panax* species has not been reported so far, several transcriptome studies of *P. ginseng*, *P. notoginseng*, *P. vietnamensis*, *P. quinquefolius* and *P. japonicus* have reported the presence of its transcripts in their datasets (Cao et al., 2015; Chen et al., 2011; Li et al., 2013; Liu et al., 2015; Luo et al., 2011; Rai et al., 2016; Sun et al., 2010a; Wu et al., 2013; Zhang et al., 2015a; Zhang et al., 2017). *AACT* ORFs from medicinal plants *B. monniera* and *G. biloba* have been successfully cloned (Chen et al., 2017; Vishwakarma et al., 2013). *B. monniera* functional AACT was produced in *E. coli* (Vishwakarma et al., 2013) while *G. biloba AACT* introduced in *AACT* deficient *Saccharomyces cerevisiae* could functionally complement the lack of host gene/enzyme (Chen et al., 2017).

5.3.2. *3-Hydroxy-3-methylglutaryl coenzyme A synthase* (*HMGS*)

HMGS catalyzes the second reaction of MVA pathway wherein acetyl-CoA is condensed with acetoacetyl-CoA to give HMGCoA (Liu et al., 2016; Miziorko, 2011). Transcripts of *HMGS* from *Panax* species have been reported by several transcriptomic studies (Cao et al., 2015; Chen et al., 2011; Li et al., 2013; Liu et al., 2015; Rai et al., 2016; Subramaniyam et al., 2014; Sun et al., 2010a; Wang et al., 2016b; Wu et al., 2013; Zhang et al., 2015a; Zhang et al., 2017). However, full length cDNA of *HMGS* was cloned only recently from *P. notoginseng* (Liu et al., 2016). ORFs of *HMGSs* from medicinal plants, *M. chamomilla* and *C. nobil*e have been successfully used to functionally complement *HMGS*-deficient *S. cerevisiae* (Cheng et al., 2016; Tao et al., 2016).

5.3.3. *Mevalonate kinase* (*MVK*)

MVK catalyzes the fourth step of MVA pathway in which mevalonate is phosphorylated to mevalonate-5-phosphate (Lluch et al., 2000; Miziorko, 2011). Several RNA sequencing studies related to different *Panax* species have reported the MVK transcripts (Cao et al., 2015; Li et al., 2013; Liu et al., 2015; Luo et al., 2011; Rai et al., 2016; Sun et al., 2010a; Wang et al., 2016b; Wu et al., 2013; Zhang et al., 2015a; Zhang et al., 2017). *In silico* analysis of *MVK* cloned from *P. notoginseng* has been found to possess 387 amino acids without a transmembrane region (Guo et al., 2012). Using the respective ORFs, functional MVKs from medicinal plants, *B. monniera* and *G. biloba* have been produced in *E. coli* and MVK-deficient *S. cerevisiae*, respectively (Chen et al., 2017; Kumari et al., 2015).

5.3.4. *1-Deoxy-D-xylulose-5-phosphate synthase* (*DXS*)

DXS catalyzes the condensation of pyruvate with D-glyceraldhyde-3-phosphate to form 1-deoxy-D-xylulose-5-phosphate (DXP) which is the first step of MEP pathway (Lichtenthaler, 1999). *P. quinquefolius DXS* transcripts were reported from the transcriptome of adventitious roots treated with methyl jasmonate (Wang et al., 2016b) whereas *P. ginseng* transcriptome revealed that *DXS* was highly expressed in the periderm region of the root (Zhang et al., 2017). All other transcriptomes of *Panax* species have failed to detect reads belonging to *DXS*. We could successfully amplify partial *DXS* from *P. sokpayensis* using degenerate primer pair. The information of this partial sequence can be used to clone the full length cDNA sequence of this elusive gene. Full length cDNAs of *DXS* from medicinal plants *W. somnifera*, *Tripterygium wilfordii* and *Conyza blinii* have been cloned and characterized (Gupta et al., 2013; Sun et al., 2014; Tong et al., 2015). Functional *T. wilfordii DXS* have also been expressed in *E. coli* (Tong et al., 2015).

5.3.5. *1-Deoxy-D-xylulose 5-phosphate reductoisomerase* (*DXR*)

DXR is a second enzyme of MEP pathway which catalyzes the formation of 2-*C*-methyl-D-erythritol-4-phosphate (MEP) from DXP (Kuzuyama, 2002; Schwender et al., 1999). Like DXS, *P. quinquefolius DXR* transcripts were reported from the transcriptome of adventitious roots treated with methyl jasmonate (Wang et al., 2016b). In *P. ginseng* transcriptome, *DXR* was highly expressed in the cortex region of the root (Zhang et al., 2017). In the current study, partial *DXR* from *P. sokpayensis* was amplified using degenerate primers that would serve as a basis for further characterization of the gene. Full length cDNAs of *DXR* from medicinal plants *T. wilfordii*, *W. somnifera* and *Camptotheca*

acuminata have been cloned and characterized (Gupta et al., 2013; Tong et al., 2015; Yao et al., 2008). *A. thaliana DXR* ORF have been used to successfully produce recombinant DXR using *E. coli* as host (Schwender et al., 1999). The same study also confirmed the functional activity of heterologous *A. thaliana* DXR using *in vitro* enzyme assay. Recombinant *T. wilfordii* DXS was also produced in *E. coli* whose functional activity was confirmed by color complementation in lycopene accumulating strain of *E. coli* (Tong et al., 2015).

5.3.6. *2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase* (*CMS*)

CMS catalyzes the formation of 4–diphosphocytidyl-2C-methyl-d-erythritol (CDP-ME) from MEP and is a third step of plastidial MEP pathway (Kuzuyama, 2002; Rohdich et al., 2000a). We were successful in amplifying a partial fragment of CMS of *P. sokpayensis* using degenerate primers. The only other *Panax* CMS reported was from the recent transcriptome study of *P. ginseng* root (Zhang et al., 2017). Among other medicinal plants, full length CMSs have been cloned from *G. biloba* and *R. verticillata* and their functions have also been validated (Kim et al., 2006b; Lan, 2013).

5.3.7. *2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase* (*MCS*)

MCS is the fifth enzyme of the MEP pathway that catalyzes the first cyclization step of the pathway by converting 4-(diphosphocytidyl)-2-C-methyl-D- erythritol-2- phosphate (CDP-MEP) into 2C- methyl-D-erythritol 2,4-cyclodiphosphate (MEcPP) (Hsieh and Goodman, 2006; Kuzuyama, 2002). Partial fragment of *P. sokpayensis* MCS was successfully cloned using degenerate primers. The only other *Panax* MCS reported was from the recent transcriptome study of *P. ginseng* root (Zhang et al., 2017). Full length cDNAs of MCS have been cloned from medicinal plants like *T. media*, *S. rebaudiana* and *G. biloba* (Jin et al., 2006; Kim et al., 2006a; Kumar et al., 2012a). The MCSs from the latter two plants have also been functionally characterized.

5.3.8. *4-Hydroxy-3-methylbut-2-en-1-yl diphosphate synthase* (*HDS*)

HDS catalyzes the sixth step and penultimate step of MEP pathway by converting 2C- methyl-D-erythritol 2,4-cyclodiphosphate (MECDP) into 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate (HMBPP) (Querol et al., 2002; Rohdich et al., 2003). Partial fragment of *P. sokpayensis HDS* was successfully cloned using degenerate primers. *HDS* transcripts were also reported from the 3 years old cDNA SSH root library of *P. notoginseng* (He et al., 2008b) and transcriptome study of *P. ginseng* root (Zhang et al., 2017). However, cloning and characterization of full length *HDS* from *Panax* species has

not been reported yet. Full length cDNA of *HDS* have been cloned and characterized from other plants like *G. biloba* and *R. verticillata* (Kim and Kim, 2010; Zheng et al., 2011).

5.3.9. *4-Hydroxy-3-methylbut-2-enyl diphosphate reductase* (*HDR*)

HDR catalyzes the seventh and final step of MEP pathway in which HMBPP is converted to IPP and DMAPP (Adam et al., 2002; Rohdich et al., 2002). Using degenerate primers, partial fragment of *P. sokpayensis HDR* was successfully cloned. *HDR* transcripts were also reported from the transcriptome studies of *P. ginseng* roots (Subramaniyam et al., 2014; Zhang et al., 2017). However, cloning and characterization of full length *HDR* from *Panax* species is yet to be achieved. *HDRs* from other medicinal plants like *A. annua* and *G. biloba* have been cloned and functionally characterized (Hao et al., 2013; Ma et al., 2017).

5.3.10. *Isopentenyl diphosphate isomerase* (*IDI*)

IDI catalyzes the reversible isomerization of IPP to DMAPP (Berthelot et al., 2012; Ramos-Valdivia et al., 1997). IPP and DMAPP are synthesized by cytosolic MVA and plastidial MEP pathways and are the precursors for the biosynthesis of all terpenoids. To the best of our knowledge, *IDI* has not been cloned and characterized from the genus *Panax*. However, transcripts have been reported from the transcriptome studies of *P. ginseng*, *P. notoginseng*, *P. japonicus*, *P. vietnamensis* and *P. quinquefolius* (Cao et al., 2015; Chen et al., 2011; Li et al., 2013; Liu et al., 2015; Luo et al., 2011; Rai et al., 2016; Sun et al., 2010a; Wang et al., 2016b; Wu et al., 2013; Zhang et al., 2015a; Zhang et al., 2017). *IDI* from medicinal plants *Salvia miltiorrhiza* and *T. wilfordii* have been cloned and functionally characterized (Tong et al., 2016; Zhang et al., 2015b). In the present study, *P. sokpayensis IDI* amplicon was amplified using degenerate primers. The sequence information obtained could be used for the full length cloning and further characterization of *P. sokpayensis IDI* in future.

5.3.11. *β-amyrin synthase* (*β-AS*)

β-amyrin synthase (β-AS) catalyzes the conversion of 2,3-oxidosqualene to *β-amyrin,* which is the proposed aglycone for oleanane type ginsenosides. 2,3-oxidosqualene is an important intermediate in the plant secondary metabolism as it is used for both phytosterol and saponin biosynthesis. In the genus *Panax*, oxidosqualene cyclases, viz., CS and lanosterol synthase (LS) (Suzuki et al., 2006) uses 2,3-oxidosqualene for the synthesis of phytosterol precursors whereas DS and β-AS use it for the synthesis of dammarane and

oleanane type aglycone of ginsenosides, respectively (Wang et al., 2014b; Zhao et al., 2015a). These oxidosqualene cyclases play important roles in channelizing the intermediate 2,3-oxidosqualene towards respective end products and thus their cloning and characterization is important. In *P. sokpayensis* partial *β-AS* was cloned using degenerate primer pair in the present study. This partial sequence would serve as base for further characterization of this important gene of the ginsenoside biosynthetic pathway responsible for the synthesis of oleanane type of aglycone. In the genus *Panax*, *β-AS* was first cloned and functionally characterized from *P. ginseng* (Kushiro et al., 1998). Several transcriptomic studies have also reported the transcripts of *β-AS* from *P. ginseng*, *P. notoginseng*, *P. quinquefolius*, *P. japonicus* and *P. vietnamensis* (Cao et al., 2015; Jayakodi et al., 2014; Li et al., 2013; Liu et al., 2015; Rai et al., 2016; Sun et al., 2010a; Wang et al., 2016b; Wu et al., 2013; Zhang et al., 2015a; Zhang et al., 2017). Use of RNA interference (RNAi) against *P. ginseng β-AS* led to a reduced level of β-amyrin and oleanane type ginsenoside, Ro while increasing the level of dammarane type of ginsenosides indicating its important regulatory role (Zhao et al., 2015a). *β-AS* have also been cloned and characterized from other medicinal plants like *W. somnifera* and *Glycyrrhiza uralensis* (Chen et al., 2013; Dhar et al., 2014).

5.3.12. *Protopanaxadiol synthase* (*PPDS*)

Previously, experimental proof was provided that cyt P450 enzyme, CYP716A47 from *P. ginseng* catalyzes the formation of protopanaxadiol from dammarenediol-II and named it protopanaxadiol synthase (PPDS) (Han et al., 2011). PPDS is a cytochrome P450 hydroxylase that catalyzes the hydroxylation of dammarenediol-II at C-12 position to give protopanaxadiol (Han et al., 2011). Several transcriptomic studies have also reported the transcripts of *PPDS* from *P. ginseng*, *P. notoginseng* and *P. vietnamensis* (Jayakodi et al., 2014; Li et al., 2013; Liu et al., 2015; Subramaniyam et al., 2014; Zhang et al., 2015a; Zhang et al., 2017). Functional role *P. quinquefolius PPDS* was also characterized (Sun et al., 2013). The same study also found that *PPDS* played a regulatory role affecting the production of ginsenosides. Thus considering the key role of PPDS in the synthesis of dammarane type of ginsenosides, partial *P. sokpayensis PPDS* (*PsPPDS*) was amplified using degenerate primer pair. The sequence information of this partial fragment of *PsPPDS* would be used to clone and characterize this important gene in future.

5.3.13. *Protopanaxatriol synthase* (*PPTS*)

PPTS is a cytochrome P450 (CYP716A53v2) enzyme that catalyzes the hydroxylation of protopanaxadiol at C-6 position to give protopanaxatriol (Han et al., 2012). Thus it is also called protopanaxadiol 6-hydroxylase. The functional characterization of PPTS was first reported by Han et al using the ORF of PPTS from *P. ginseng* (Han et al., 2012). *PPTS* expression was found to increase in response to continuous chilling stress increasing the production of protopanaxatriol type of ginsenosides (Jiang et al., 2016). It is suggested that the increased production of protopanaxatriol type of ginsenosides under chilling stress may be a strategy used by ginseng cells to detoxify itself from reactive oxygen species (ROS) that accumulates under such stress (Jiang et al., 2016). Overexpression and RNA interference of *P. ginseng PPTS* led to the increased and decreased production of protopanaxatriol type of ginsenosides, respectively, indicating the role of *PPTS* in the regulation of ginsenosides (Park et al., 2016). Thus to study this important gene in *P. sokpayensis*, its partial fragment was amplified using degenerate primers. The sequence information of this partial *PgPPTS* would be used for full length cloning of this gene for further characterization in future. Cloning and functional characterization of *P. quinquefolius PPTS* have also been reported (Wang et al., 2014c). Transcripts of *P. ginseng*, *P. notoginseng*, *P. japonicus* and *P. vietnamensis PPTS* have been reported in their respective transcriptome studies (Jayakodi et al., 2014; Liu et al., 2015; Rai et al., 2016; Zhang et al., 2015a; Zhang et al., 2017).

Partial fragments of *PMVK*, *CMK*, *SE* and *CS* were also amplified using degenerate primer approach. Full length cDNA sequences of these four genes were obtained using RACE. *PMVK* and *CMK* are discussed under section 5.4 while *SE* and *CS* are discussed under section 5.5.

5.4. Characterization of full length genes from MVA and MEP pathways

5.4.1. *Phosphomevalonate kinase* (*PMVK*)

PMVK catalyzes the fifth step of mevalonate pathway, converting mevalonate-5-phosphate to mevalonate-5-diphosphate (Olivier et al., 1999). We used homology based PCR to amplify the partial fragment of *PsPMVK* (Figure 4.23D) and then used RACE to obtain the full length (Figure 4.32). Then *in silico* characterization of the gene was done using various bioinformatics tools.

5.4.1.1. Full-length cloning and *in silico* analysis of *PsPMVK*

A single polyadenylation signal, AATAAA was detected in the 3′ UTR of *PsPMVK* (Figure 4.37). *PMVKs* from other *Panax* species that are reported in the NCBI database lacked such PAS. The predicted molecular weight of 54.72 kDa of PsPMVK was similar to that of PMVK reported from *C. nobile* L. (Yan et al., 2016). Fairly high similarity observed in global alignments of PsPMVK with the PMVK of *P. ginseng* and *P. notoginseng* shows that the major part of this protein is conserved. Also the multiple sequence alignments revealed that the C – terminal end of PMVK was less conserved (Figure 4.35). Though Kyte and Doolittle hydropathy plot shows two hydrophobic regions (Figure 4.34C), TMHMM analysis failed to detect any transmembrane helix in PsPMVK (Figure 4.34D). Though MVA pathway machinery occurs in cytoplasm (Wang et al., 2012), studies on human and rat PMVK has suggested of it being present in peroxisomes (Chambliss et al., 1996; Olivier et al., 1999). On phylogenetic analysis, PsPMVK showed a close evolutionary relationship with PMVKs from *Panax* species by aligning itself close to them (Figure 4.39).

The transcripts of PMVK have also been reported from the transcriptome studies of different *Panax* species (Li et al., 2013; Luo et al., 2011; Sun et al., 2010a; Wang et al., 2016b; Zhang et al., 2015a). However, its further characterizations, either *in silico* or experimental, have not been done. Hence, *PsPMVK* could not be compared with the *PMVK* of other *Panax* species.

5.4.2. *4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase* (*CMK*)

CMK catalyses the fourth step of plastidial MEP pathway, converting 4-(diphosphocytidyl) -2-C-methyl-D-erythritol into 4-(diphosphocytidyl)-2-C- methyl- D-erythritol-2- phosphate (Concepcion et al, 2015; Kuzuyama and Seto, 2012; Zhao et al., 2014). Although, several transcriptome studies on many *Panax* spp. have reported various genes belonging to the ginsenoside biosynthesis, none of these studies have reported CMK transcripts (Cao et al., 2015; Chen et al., 2011; Jayakodi et al., 2014; Li et al., 2013; Liu et al., 2015; Luo et al., 2011; Rai et al., 2016; Subramaniyam et al., 2014; Sun et al., 2010a; Wang et al., 2016b; Wu et al., 2013; Zhang et al., 2015a). Thus, this is the first report of CMK from genus *Panax*, due to which comparative studies against the same enzyme from other *Panax* species couldn′t be conducted.

5.4.2.1. Full-length cloning and *in silico* analysis of *PsCMK*

Canonical PASs (AATAAA and ATTAAA) were not detected in the 3′ UTR of *PsCMK*. The predicted molecular weight of 44.33 kDa of deduced 403 amino acid long

polypeptide of PsCMK was similar to that of CMK reported from *Nicotiana benthamiana* (Ahn and Pai, 2008). Global alignments revealed that PsCMK was highly similar to the CMK from *H. helix*, which is also from Araliaceae family and fairly similar to those from *L. japonica* var. *chinensis*, *S. tuberosum* and *W. somnifera* showing that it is more conserved within the genus. Multiple sequence alignments of the reported CMKs from across genera and families found that the GHMP kinases N terminal domain was highly conserved while the IspE superfamily domain showed considerable variations at both N and C terminals (Figure 4.39). Kyte and Doolittle hydropathy plot and TMHMM analysis did not detect any potential transmembrane helix indicating that PsCMK is not bound to any membrane (Figures 4.38C, D). CMKs from other *Panax* species could not be used for phylogenetic analysis as they have not been reported. On phylogenetic analysis with CMKs from other genus, PsCMK aligned close to a CMK from *H. helix* which also belongs to Araliaceae family indicating their common ancestry (Figure 4.44).

5.5. Characterization of full length genes (*FPS*, *SS*, *SE*, *DS* and *CS*) from ginsenoside biosynthetic pathway

5.5.1. *Farnesyl pyrophosphate synthase* (*FPS*)

FPS is a key enzyme that channelizes the precursors from MVA and MEP pathway towards terpenoid biosynthesis by producing FPP (Zhao et al., 2014). FPP is located at a branch point in terpenoid biosynthetic pathway and serves as a precursor for the synthesis of large number of sesquiterpenes and other higher terpenes (Bohlmann et al., 1998; Ferriols et al., 2015; Sawai and Saito, 2011). Generally, FPS has the functions of both geranyltransferase which forms FPP from GPP and IPP, and dimethylallyl transferase which forms GPP from DMAPP and IPP (Su et al., 2015). It is the first crucial enzyme of terpene biosynthesis and produces FPP that serves as a precursor for the synthesis of other intermediates which ultimately leads to the production of ginsenosides and phytosterols in *Panax* species (Kim et al., 2010a; Kim et al. 2014). *FPSs* from other organisms have also been suggested to play a pivotal role in the regulation and synthesis of terpenes and have been cloned, characterized and produced in heterologous systems (Srivastava et al. 2015; Su et al. 2015; Zhao et al. 2015). Moreover, differential expression of *FPS* was observed in rhizome specific suppression subtractive hybridization (SSH) library in *P. sokpayensis*, (Gurung et al., 2016). All the above findings suggest the regulatory role of *FPS* in terpenoid/ginsenoside biosynthesis.

5.5.1.1. Full-length cloning and *in silico* analysis of *PsFPS*

A full length *FPS* with 1029 bp long open reading frame (ORF) was successfully reconstructed using RACE. Three identical polyadenylation signals (PASs) were detected in 3′ UTR of full length *PsFPS* (Figure 4.11). PASs are defining features of eukaryotic protein – coding genes. They are essential for 3′ end processing, cleavage and polyadenylation of pre-messenger RNA and transcriptional termination (Proudfoot, 2011).

In silico analysis of deduced 342 amino acids of PsFPS predicted a molecular weight of 39.6 kDa with theoretical pI of 5.63 which is corroborated by the earlier report of FPS from *P. ginseng* (Kim et al., 2010a). Global alignments of deduced amino acids sequence of PsFPS with that of different *Panax* species revealed 99.1 % identity with *P. ginseng* FPS (AAY87903) and 98.8 % identity with FPS of *P. notoginseng* (AGS79228), *P. quinquefolius* (ADJ68004) and *P. japonicus* (AKN52395). This shows that the enzyme is highly conserved despite wide geographical separations among these species. The conserved domain search annotated PsFPS with trans – isoprenyl diphosphate synthase, head to tail domain containing protein under isoprenoid_biosyn_C1 superfamily (Figure 4.12A). The predicted secondary structure consisted of 57.89 % alpha helix, 10.82 % extended strand, 7.60 % beta turn and 23.68 % random coil (Figure 4.12B). We couldn't find any report on the *in silico* predicted secondary structures of FPS from *Panax* species. Hence, when we did *in silico* prediction of deduced amino acid sequence of *P. ginseng FPS* (DQ087959), we found that the results corroborated to the secondary structures of PsFPS (data not shown). Kyte and Doolittle hydropathicity plot and prediction of transmembrane region using TMHMM program predict PsFPS to be a cytosolic protein (Figures 4.12C, D). Previous study in *A. thaliana* have concluded that FPS is a cytosolic protein (Keim et al., 2012). Seven conserved features, viz., substrate binding pocket (22/22 residues), substrate – Mg^{2+} binding site (13/13 residues), active site lid residues (27/28 residues), chain length determination region (10/10 residues), catalytic residues (8/8 residues), first aspartate rich motif (FARM) (7/7 residues) and second aspartate rich motif (SARM) (6/6 residues) were mapped to PsFPS (Figure 4.13). DDXXD, a highly conserved motif of FARM and SARM was also found on PsFPS. Multiple sequence alignment of FPSs from across the genera and families from the plant kingdom revealed that all the domains present in this protein were highly conserved (Figure 4.13). The presence of aspartate rich FARM and SARM with the highly conserved DDXXD which is the main catalytic site of prenyltransferases (Marreros et al., 1992; Song and Poulter, 1994) suggests that PsFPS carries the same functionality as the FPSs of other organisms. Moreover, it has been proved through *in vitro* enzyme assay

that the purified recombinant PgFPS catalyzed condensation of IPP with GPP to produce FPP (Kim et al., 2010a). Thus a very high similarity in sequence, secondary structure and functional domains between PsFPS and PgFPS strongly suggests their functional similarity. Phylogenetic analysis revealed that FPS of *Panax* species were highly related to each other and thus clustered close to one another in the cladogram (Figure 4.14). Another species from Araliaceae family, *A. elata* was also placed very close to the *Panax* species showing their close evolutionary relationship.

5.5.2. *Squalene synthase* (*SS*)

SS is a bifunctional enzyme bound to the membrane of endoplasmic reticulum (Abe et al., 1993; Busquets et al., 2008). It initially catalyzes the condensation of two FPP molecules in a head to head fashion to presqualene diphosphate (PSPP) and then converts PSPP to squalene in the presence of NADPH and Mg^{2+} (Takatsuji et al., 1982; Rong et al., 2016). This is the first committed step towards the synthesis of higher terpenoids, sterols and brassinosteroids (Busquets et al., 2008). In case of *Panax* species, this step is the first committed step towards synthesis of ginsenosides and phytosterols (Sathiyamoorthy et al., 2010). The SS is shown to have a regulatory role in the ginsenoside biosynthesis (Lee et al., 2004; Seo et al., 2005). SS gene is present in variable copy number in different plants. For example, a single SS gene is reported from plants like *Taxus cuspidata* (Huang et al., 2007), *O. sativa* (Hata et al., 1997), *Lotus japonicus* (Akamine et al., 2003) and *Euphorbia Pekinensis* (Zheng et al., 2013) whereas two genes have been reported from *Glycyrrhiza glabra* (Hayashi et al., 1999), *N. tabacum* (Devarenne et al., 1998), *W. somnifera* (Gupta et al., 2012) and *G. max* (Nguyen et al., 2013). In *P. ginseng* three genes encoding SS1, SS2 and SS3 have been reported (Kim et al., 2011).

5.5.2.1. Full-length cloning and *in silico* analysis of *PsSS*

The *in silico* deduced molecular weight of 47.13 kDa corroborated with the molecular weights of SSs reported from *P. ginseng* (Lee et al., 2004; Kim et al., 2011). The multiple sequence alignments of deduced amino acid of PsSS with the SS from other *Panax* species, viz., *P. ginseng*, *P. quinquefolius*, *P. notoginseng* and *P. japonicus* (Figure 4.17) and their global alignments showed that the SS is well conserved in the genus despite wide geographical separations. The multiple sequence alignments also show that catalytic domains and the entire enzyme seem to be highly conserved in the plant kingdom (Figure 4.17). In fact, the functional domains of SS has been found to be well conserved in eukaryotes (Linscott et al., 2016; Robinson et al., 1993). The presence of two main catalytic

motifs of FARM and SARM with canonical conserved sequence, DXXXD (DTVED and DYLED in PsSS) suggests that PsSS may have similar functionality as other SSs (Gupta et al., 2012; Zheng et al., 2013). High proportion of predicted alpha helix in the PsSS secondary structure was in line with the SS reported from *S. miltiorrhiza* (Rong et al., 2016). Hydropathicity plot and TMHMM program predict transmembrane helix at the C-terminal of PqSS (Figures 4.16C, D) like in SS reported from other plants (Gupta et al., 2012; Kim et al., 2011). It has been suggested that this hydrophobic transmembrane helix helps SS to anchor to the membrane of endoplasmic reticulum (Kim et al., 2011). Deduced PsSS forming a separate clade along with SSs of other *Panax* species in the phylogenetic tree indicates close evolutionary relationships among these proteins and suggests that these species are closely related (Figure 4.19).

5.5.3. *Squalene epoxidase* (*SE*)

SE functions as a rate-limiting enzyme in the phytosterol and triterpenoids biosynthesis and catalyzes the oxidation of squalene to 2,3-oxidosqualene. In addition to 2,3-oxidosqualene, SE activity can result in the formation of other oxidosqualenes such as 6,7-oxidosqualene, 10,11-oxidosqualene and dioxidosqualene (Bai and Prestwich, 1992). Being a rate limiting enzyme, SE has a cascading influence on the up-regulation of downstream genes (Han et al., 2010) and has significant implications in biosynthetic pathway intensification (Razdan et al., 2013). Previously, *SE* genes have been isolated from different plant species such as *C. asiatica* (Kim et al., 2005), *W. somnifera* (Razdan et al., 2013), *P. notoginseng* (Niu et al., 2014), *C. borivilianum* (Kalra et al., 2015), *Gynostemma pentaphyllum* (Guo et al., 2016a), etc. However, unlike yeast and mammals, SE of several plants belong to multigene families (Rasbery et al., 2007). For example, genes encoding for SEs of *P. notogineng*, *P. ginseng*, *P. vietnamensis* var *fuscidiscus*, *A. thialiana* and *G. pentaphyllum* are multigenic in nature (Guo et al., 2016a; Han et al., 2010; Ma et al., 2016; Niu et al., 2014; Rasbery et al., 2007). Also, Laranjeira et al (2015), have suggested the evolutionary role of SEs due to their phylogenetic diversity in plants.

5.5.3.1. Full-length cloning and *in silico* analysis of *PsSE*

Two polyadenylation signals (PASs) with canonical sequences AATAAA and ATTAAA were detected in the 3′ UTR of *PsSE*. Such multiple PASs were also found in the *SE*s of *P. notoginseng* (KJ804171) and *P. ginseng* (BAD15330, BAA24448) submitted in NCBI database. Studies on alternative polyadenylation signals in SE gene have not been reported so far. Hence, functional and structural implications of presence of such multiple

PASs are not clear. The predicted molecular weight and pI of deduced polypeptide of PsSE were consistent with those of SEs from other *Panax* spp. (He et al., 2008a; Han et al., 2010; Ma et al., 2016). The percentages of secondary structures of PsSE are in range with those of SEs reported earlier (Kalra et al., 2015; Ma et al., 2016; Razdan et al., 2013). The global alignments of SE from *P. sokpayensis* with the SEs from other *Panax* species showed that they were highly similar. The amino acid sequence alignment and conserved domain analysis revealed the presence of NAD domain further supporting the proposal for PsSE being a true SE. Previous reports have revealed that plant SEs can be divided into two clades: true SEs and SE-like proteins. These true SEs have specific NAD domains whereas SE-like proteins have specific FAD domains (Laranjeira et al., 2015). Two SE genes (*PgSE1* and *PgSE2*) were investigated in *P. ginseng* which suggested their critical role as regulators (Han et al., 2010). This study further revealed the differential regulation of *PgSE1* in ginsenoside biosynthesis and *PgSE2* in phytosterol biosynthesis in *P. ginseng*. Also, three SE isoforms were cloned and characterized in *P. vietnamensis* and suggested differential role in sterol and ginsenoside biosynthesis (Ma et al., 2016). Four transmembrane helices predicted using Kyte and Doolittle hydropathicity plot and TMHMM program corroborated with the similar earlier reports (Han et al., 2010; Laranjeira et al., 2015; Ma et al., 2016). These transmembrane helices are used to anchor SEs to the membrane of endoplasmic reticulum (Han et al., 2010; Laranjeira et al., 2015; Ma et al., 2016; Niu et al., 2014). Like other enzymes of the ginsenoside biosynthetic pathway, SEs of *Panax* species including PsSE formed a separate clade in the phylogenetic tree inferring their evolutionary close relatedness (Figure 4.29).

5.5.4. *Dammarenediol synthase* (*DS*)

DS is one of the two enzymes that cyclizes 2,3 oxidosqualene to form intermediate for the formation of dammarane type ginsenosides, the other one being β-AS, which forms intermediate for the oleanane type ginsenosides (Tansakul et al., 2006; Wang et al., 2012). Thus DS is a first committed enzyme for the biosynthesis of dammarane type ginsenosides (Tansakul et al., 2006). Using cDNA of ORF coding for DS of *Panax* spp., functional DS has been produced in heterologous systems like yeast (Liang et al., 2012; Tansakul et al., 2006; Wang et al., 2014b) and bacteria (Hu et al., 2013). Transgenic tobacco plants producing dammarenediol II with tolerance to Tobacco Mosaic Virus (TMV) (Lee et al., 2012) and transgenic tobacco cell suspension culture producing 5.2 mg dammarenediol-II per liter has been established using cDNA for *P. ginseng* DS (Han et al., 2014). Thus full

length cDNA of this important enzyme from *P. sokpayensis* was cloned and its *in silico* characterization was done.

5.5.4.1. Full-length cloning and *in silico* analysis of *PsDS*

As in case of *PsFPS*, three polyadenylation signals (PASs) with canonical sequence AATAAA were detected in the 3′ UTR of *PsDS*. Such multiple PASs and their roles have not been reported in DS so far. However, when *DS* nucleotide sequence belonging to *Panax* species were retrieved from NCBI database and checked for the polyadenylation signals, two *DSs* (ACZ71036 and AB265170) reported from *P. ginseng* and one *DS* (KJ804174) from *P. notoginseng* were found to contain three PASs as in *PsDS* (data not shown). The predicted molecular weight of 88.37 kDa corroborated with the actual molecular weight of *P. ginseng* DS determined experimentally (Liang et al., 2012; Hu et al., 2013). Secondary structures could not be compared with those of DS from other species as there are no previous such reports. Hence, few deduced polypeptide sequences of DS from other *Panax* species were retrieved from the NCBI database and their secondary structures were predicted using *in silico* tool as mentioned in Table 3.3 (section 3.10, Chapter 3: Materials and methods). Their secondary structures were highly similar to that of PsDS (35.50 % alpha helix, 17.43 % extended strand, 12.35 % beta turn and 34.72 % random coil) (data not shown). The global alignments of *P. sokpayensis* DS with the DS of other *Panax* species showed very high similarities among them. Similarly, the catalytic acid and active site cavity residues that were mapped to PsDS were highly conserved (Figure 4.21) indicating the importance of these features for the function of the enzyme. Though hydropathy plot predicts four possible hydrophobic regions (Figure 4.20C), the TMHMM analysis has predicted a single transmembrane helix comprising amino acids 612 – 634 (Figure 4.20D). The heterologously produced *P. ginseng* DS was localized in lipid particles (Liang et al., 2012). This is most likely due to the presence of hydrophobic transmembrane helix predicted above, whose main function may be to help the protein to anchor to membrane. A distinct phylogenetic clade formed by the DSs of *Panax* species revealed that PsDS is evolutionarily related to the DSs from other *Panax* species (Figure 4.24).

5.5.5. *Cycloartenol synthase* (*CS*)

CS converts 2,3 oxidosqualene to cycloartenol, which is an intermediate for the synthesis of phytosterols. Thus, CS channelizes metabolic flux towards phytosterol biosynthesis and in doing so reduces the metabolic flux towards the ginsenoside biosynthesis (Liang et al., 2009). Study of CS is thus important in order to understand the

metabolic flux and regulatory aspects of ginsenoside biosynthesis. By suppressing *CS* in *P. ginseng* and *P. notoginseng* using antisense suppression and RNA interference, respectively, transgenic lines producing high levels of ginsenosides have been produced (Liang et al., 2009; Yang et al., 2017b).

5.5.5.1. Full-length cloning and *in silico* analysis of *PsCS*

The 2277 bp long ORF of *PsCS* coding for 758 amino acids had a predicted molecular weight of 86.01 kDa which was similar to the molecular weights of CSs characterized in other species (Corey et al., 1993; Kim et al., 2005; Zhao et al., 2017). Global alignments with the CSs reported from other *Panax* spp. revealed that PsCS showed high similarities with them. On multiple sequence alignments, it was found that the two conserved features, viz., catalytic acid and catalytic site cavity mapped to PsCS were conserved not only in CSs of other *Panax* spp. but also in plants belonging to other genera indicating the importance of these features (Figure 4.31). Though hydropathy plot predicts a hydrophobic region near the N terminal end suggesting the presence of transmembrane region (Figure 4.30C), TMHMM analysis showed that the probability of presence of such region is very low (Figure 4.30D). *In silico* analysis of CS from another medicinal plant, *Siraitia grosvenorii* also shows the absence of transmembrane region in the protein (Zhao et al., 2017). PsCS aligning close to CSs from other *Panax* species in the phylogenetic tree reveals their common ancestry and close evolutionary relationship (Figure 4.34).

5.6. Key genes involved in ginsenoside biosynthesis exhibit differential expression in leaf, stem and rhizome

5.6.1. *PsFPS*

The qRT expression studies found that the relative expression of *PsFPS* was ~1.5- and ~3.7-fold higher in rhizome than that of leaf and stem, respectively (Figure 4.22A). Similar differential expression of *FPS* in different organs have also been reported in *P. ginseng* where highest expression was seen in root followed by leaf and stem (Kim et al., 2010a). Previous studies have shown that *FPS* expression is upregulated by methyl jasmonate (MJ) treatment in *P. ginseng* and *P. notoginseng* (Kim et al., 2010a; Niu et al., 2014). Moreover, MJ treatments have also increased the ginsenoside contents in the above two *Panax* species (Kim et al., 2009b; Lu et al., 2001; Oh et al., 2014). Thus the differences in spatial expression of *FPS* in *P. sokpayensis* and other *Panax* species as well as

upregulation of both *FPS* transcripts and ginsenoside contents by methyl jasmonate strongly suggest that *FPS* plays an important regulatory role in ginsenoside biosynthesis.

5.6.2. *PsSS*

The expression of *PsSS* was ~3-fold higher in the rhizome as compared to that of leaf and stem (Figure 4.22B). Higher expression of *SS* in rhizome compared to leaf and stem was also reported in *P. ginseng* (Lee et al., 2004). Similarly, spatial expressions of *SS* in other plants have also shown similar patterns with roots containing higher transcript levels compared to leaves and stems (Rong et al., 2016; Vishwakarma et al., 2015; Zheng et al., 2013). *SS* from *Panax* species are upregulated on treatment with an elicitor, methyl jasmonate (Kim et al., 2011; Lee et al., 2004). Moreover, there are several studies that have reported that ginsenoside content is enhanced by MJ treatment (Kim et al., 2009b; Lu et al., 2001; Oh et al., 2014). Also, the transgenic adventitious roots of *P. ginseng* created using *P. ginseng SS* (Lee et al., 2004) and *P. notoginseng* cells co-transformed with *P. notoginseng SS* and *3-hydroxy-3-methylglutaryl CoA reductase* (*HMGR*) (Deng et al., 2017) showed enhanced ginsenosides contents in comparison to the non transgenic ones. The differential expression of *PsSS* in rhizome of *P. sokpayensis* suggests its regulatory roles in the ginsenoside biosynthesis.

5.6.3. *PsSE*

qRT-PCR analysis revealed differential expression pattern of *PsSE* with leaf showing higher expression than stem and rhizome tissues (Figure 4.40A). Previously, same trends of over expression of *SE* gene have been observed in *P. ginseng* and *P. notoginseng* (Han et al., 2010; Niu et al., 2014). These results also support the expression of *SE* from *Euphorbia tirucalli* (Uchida et al., 2007), *W. somnifera* (Razdan et al., 2013) and *C. borivilianum* (Kalra et al., 2015) where the gene expression was higher in leaf as compared to stem and root tissues. These findings suggest that the formation of ginsenoside precursors occurs in leaf and are then transported to root/ rhizome in *P. sokpayensis*.

5.6.4. *PsDS*

The expression of *PsDS* was ~3.7- and ~19- fold higher in the rhizome than that of the leaf and stem, respectively (Figure 4.22C). Higher expression of *DS* in rhizome in comparison to leaf and stem were also reported from other *Panax* species (Han et al., 2006a, 2006b). According to previous reports, methyl jasmonate upregulates *DS* as well as enhances ginsenosides production in *Panax* species (Han et al., 2006a; Kim et al., 2009; Lu

et al., 2001; Oh et al., 2014; Um et al., 2017; Wang et al., 2014b) suggesting a correlation between the expression of *DS* and the ginsenoside production. This correlation has been further substantiated by creating transgenic line of *P. quinquefolius* hairy roots wherein overexpression of *DS* from the same species led to an increased accumulation of ginsenosides (Wang et al., 2014b). Interestingly, higher ginsenosides content in rhizome of *P. sokpayensis* in comparison to leaf (Gurung et al., 2018) and also higher expression levsel of *PsDS* in rhizome (Gurung et al., 2016) tempts us to suggest that *PsDS* is one of the genes that regulates ginsenoside biosynthesis in *P. sokpayensis*.

5.6.5. *PsCS*

Realtime expression analysis shows the differential nature of expression of *PsCS* with the highest expression being in rhizome (Figure 4.40B). *P. ginseng* roots have been found to contain high amount of phytosterols (Beveridge et al., 2002; Lee et al., 2016). The upregulation of CS gene has also been positively correlated with the phytosterol content in the adventitious roots of *P. ginseng* (Lee et al., 2004) indicating a regulatory nature of CS. Thus higher expression of *PsCS* in the rhizome in comparison to leaf and stem tissues may indicate that *P. sokpayensis* rhizome may have higher phytosterol content than the two other tissues.

5.7. Differential expression in leaf, stem and rhizome of other genes obtained through SSH suggests their roles in plant growth and development, defense and ginsenoside biosynthesis

From the leaf SSH library, qRT-PCR study detected *galactinol synthase 2* expression only in the leaf but not in the rhizome and stem (Figure 4.47A). *Galactinol synthases* (*GolS*) are reported as stress-responsive genes in several plant species. For instance, two *GolS* genes that were cold-inducible were reported from the leaves of *Ajunga reptans* (Sprenger and Keller, 2000). Osmotic stress and low temperature upregulated rice *GolS* (Takahashi et al., 1994). In *A. thaliana*, out of seven *GolS* identified, the expression of *GolS1* and *GolS2* were induced by heat, salt and drought stress in leaf tissues (Taji et al., 2002). Overexpression of *GolS2* in transgenic *A. thaliana* increased the levels of endogenous raffinose and galactinol under normal growth conditions (Taji et al., 2002). *GolS1* and *GolS2* expressions were also induced by high light and heat stress or H_2O_2 treatment suggesting their roles in protecting plants from oxidative stress (Nishizawa et al., 2008). These observations suggest that *galactinol synthase 2* may have a role in the

protection of cellular metabolism in *P. sokpayensis*. The expression of *cell division cycle 20.1* (*CDC 201.1*) gene was highest in leaf when compared to rhizome and stem (Figure 4.47B). *CDC 20.1*, a cofactor of APC complex protein, is part of anaphase promoting complex/cyclosome that is involved in G1 phase of cell cycle. It plays multiple roles in chromosome segregation during meiosis in mice, bovine oocytes and *Drosophila melanogaster* (Chu et al., 2001; Jin et al., 2010; Sun and Kim, 2012; Yang et al., 2014). Five *CDC 20*-like genes, namely *CDC 20.1*, *CDC 20.2*, *CDC 20.3*, *CDC 20.4* and *CDC 20.5* have been reported from *A. thaliana*. The *A. thaliana CDC 20.1* has been found to play a role in plant fertility and meiosis (Niu et al., 2015). Differential expression of *CDC 20.1* gene in *P. sokpayensis* suggests its role in growth and development. Another gene, *metallothionein-like protein type 3* showed more expression in stem in comparison to the leaf and rhizome tissues (Figure 4.47C). Previous study has shown that overexpression of *Tamarix androssowii metallothionein-like protein type 3* gene (*TaMT3*) in transgenic tobacco played a protective role in photosynthesis apparatus and antioxidant system (Zhou et al., 2014). Expression of *GBR-5* gene was high in the leaf when compared to rhizome and stem (Figure 4.47D). Previously, isolation and characterization of *GBR-1*, *GBR-2*, *GBR-3*, *GBR-4*, and *GBR-5* form *P. ginseng* suggested that they play important roles in the ginsenoside biosynthesis (Luo et al., 2003). *PsbA* expression was higher in leaf as compared to the stem and rhizome (Figure 4.47E). It is an important gene encoding D1 protein which forms the reaction core of PSII along with the D2 protein. The efficiency of photosynthesis is drastically reduced by abiotic stress (Saibo et al., 2009). Previously, differential regulation of *Hordeum vulgare PsbA* and *PsbD* genes has been reported (Christopher and Mullet, 1994). Since *P. sokpayensis* grows in a shady habitat, the differential expressions of all these transcripts suggest their role in photosynthetic adjustment and carbon influx for the biosynthesis of ginsenosides. From the rhizome SSH library, expression of gene encoding protein KIAA0664 homologue was high in the rhizome when compared to stem and leaf (Figure 4.47F). Study of differential expression of this protein in porcine cells infected with rA/FMDVor rA/FMDV ΔmSAP virus suggested its role in pathogenesis, immune response and cellular metabolism (Zhu et al., 2015). These results suggest that this novel and uncharacterized gene might have roles in immunity and pathogenesis in *P. sokpayensis*. *Major latex-like protein* (*MLP*) gene showed higher expression in rhizome than in leaf and stem (Figure 4.47G). Major latex proteins were first reported from the latex of *P. somniferum* (Lytle et al., 2009). Their function in plants is not known but their patterns of expression showed similarity with some of the intracellular pathogenesis-related proteins

(Sun et al., 2010b). They have been suggested to be important in defense response to *Fusarium equiseti* infection in the roots of *P. quinquefolius* (Goswami and Punja, 2008). *Gh-MLP* was upregulated in the roots of *Gossypium hirsutum* under salt stress (Chen and Dai, 2010). *MLPs* have also been suggested to be involved in many biological processes like transcription, translation, plant response to environmental stimuli and protein degradation (Choi et al., 2005). Higher *MLP* expression in *P. sokpayensis* rhizome suggests their role in plant immunity and defense response to some pathogens. *RNase-like major storage protein* gene expression was detected only in rhizome but not in leaf and stem (Figure 4.47H). Previously, expression studies of RNase-like major storage protein in *P. ginseng* suggested its role in defense mechanism (Kim et al., 2004). *GAPDH* gene expression was upregulated in the rhizome as compared to leaf and stem (Figure 4.47I). GAPDH is an important enzyme in glycolysis and gluconeogenesis which are the main pathways for supplying energy and substrates for metabolic pathways. It may mediate stress-induced metabolic responses and provide additional energy for the cellular adjustment during growth and development under stress conditions by diverting carbon away from glycerol into the pathway leading to glycolysis and ATP formation (Jeong et al., 2000). *GAPDH* expression was also upregulated during drought stress in potato indicating its role in adaption during to drought (Kappachery et al., 2015). Constitutive overexpression of *GAPDH* in transgenic potato plants have helped them to survive under drought stress (Kappachery et al., 2015). These results suggest a role of *GADPH* in growth and development of *P. sokpayensis* in a niche environment. *Ankyrin repeat protein gene* was also significantly upregulated in rhizome as compared to two other tissues, stem and leaf (Figure 4.47J). Ankyrin (ANK) repeat proteins are have roles in several functions that includes signal transduction, cell cycle regulation, mitochondrial enzymes and cytoskeleton interactions (Sedgwick and Smerdon, 1999). Previously, ANK repeat-containing protein has been reported to be involved in the regulation of antioxidation metabolism during disease resistance and stress responses in *Arabidopsis* (Yan et al., 2002). In pepper, *ankyrin-repeat containing protein* (*CaKR1*) has been shown to play roles in both biotic and abiotic stress responses (Seong et al., 2007). Differential expression of ankyrins in different tissues have also been reported indicating their role in plant growth and development (Yuan et al., 2013). Higher expression of *β-amylase* was observed in rhizome in comparison to the leaf and stem (Figure 4.47K). β-Amylase produces maltose by hydrolyzing alpha-1,4 glycosidic linkages of polyglucan chains at the nonreducing end (Kaplan and Guy, 2004). *β-Amylase-like protein* has been found to be noncatalytic and play a regulatory role in starch degradation in chloroplast (Fulton et al., 2008). A study has

revealed that the two genes encoding BAM7 and BAM8 have roles in plant growth and development (Reinhold et al., 2011). Expression of transcription factor, *bHLH*, was higher in rhizome in comparison to stem and leaf tissues (Figure 4.47L). bHLH proteins bind as dimers to specific DNA target sites and are extensively characterized in nonplant eukaryotes for the regulation of diverse biological processes (Toledo-Ortiz et al., 2003). In *A. annua*, a MYC-type bHLH transcription factor has been shown to act as an activator of sesquiterpene synthase and cytochrome P450 essential for the biosynthesis of artemisinin (Ji et al., 2014; Yu et al., 2012). In tomato, a WRKY and MYC-type bHLH transcription factors have been shown to bind the promoter of sesquiterpene synthase gene (Spyropoulou et al., 2014). In *A. thaliana*, it has been observed that the bHLH directly binds to sesquiterpene synthase promoters, effecting an enhanced release of volatile sesquiterpenoids (Hong et al., 2012). Recently, two bHLH transcription factors, TSAR1 and TSAR2 of *M. truncatula*, have been shown to activate triterpene saponin biosynthesis (Mertens et al., 2016). These findings suggest that *bHLH* may play important role of in the activation of ginsenoside biosynthetic pathway genes in *P. sokpayensis*. *Ubiquitin-conjugating protein* and *polyubiquitin-like protein* were also highly upregulated in the rhizome in comparison to leaf and stem (Figures 4.47M, N). Ubiquitination is a pivotal regulatory step in protein degradation mechanism and is known to regulate pertinent functions in different plant growth and developmental processes (Zhiguo et al., 2015). Ubiquitin-conjugating protein is involved in modifications of targeted proteins by catalyzing the transfer of ubiquitin to substrate or E3 ligase. It mediates the formation of polyubiquitin chains on target proteins (Ye and Rape, 2009) and, together with the E3 enzyme, determines substrate specificity in the ubiquitination system (Bae and Kim, 2014). Recently, five *OsUBC* genes showing upregulation in *O. sativa* roots suggests their role in plant immunity, growth, and development (Zhiguo et al., 2015). Similarly, another gene *heat shock protein 70* (*HSP70*) showed higher expression in the rhizome when compared to the leaf and stem tissues (Figure 4.47O). HSP70 plays a role in preventing aggregation and support refolding of nonnative proteins during normal and as well as stress conditions. It is involved in import, translocation and proteolytic degradation of unstable proteins by targeting them to proteosomes or lysosomes (Wang et al., 2004). It could be suggested that differential expression of all these genes in the leaf, stem and rhizome may be responsible for the plant growth and development as well as for the biosynthesis of different ginsenosides in *P. sokpayensis* growing in the niche environment.

5.8. Heterologous expression of PsFPS in *E. coli*

Colony PCR and sequencing confirmed the in-frame cloning of *PsFPS* in pQE30. The recombinant PsFPS fusion protein of approximate size of 40 kDa was observed on 12 % SDS–PAGE gel (Figure 4.41). The time course analysis of expression of recombinant protein revealed high level of expression within 1 h after induction by 1 mM IPTG (Figure 4.42). High levels of protein expression were observed upto 5 h after induction. These results indicate the successful expression and production of the recombinant PsFPS in host *E. coli* cells. FPSs from many important plants including *P. ginseng* have been successfully cloned and expressed in bacteria using recombinant DNA technology. For example, functional *P. ginseng* FPS (cDNA accession number, DQ087959) was successfully produced in *E. coli* (Kim et al., 2010a). Recently, Srivastava et al., (2015) have co-expressed FPS with two sesquisabinene synthases, SaSS1 and SaSS2 in bacterial culture to produce sesquisabinene, an important sesquiterpene of *Santalum album*. Similarly *FPS* from other medicinal plants like *Matricaria recutita*, *T. wilfordii* and *A. annua* have been expressed in bacteria (Su et al., 2015; Zhao et al., 2015b; Matsushita et al., 1996). Our results revealed high expression of PsFPS in *E. coli* cells which can be harnessed for the biosynthesis of isoprenoid precursors.

Heterologous production of recombinant proteins in microorganisms could be an alternative approach to synthesize active metabolites or their intermediates found in medicinal plants through synthetic biology approach. As discussed earlier, FPS plays a pivotal role in diverting the precursors IPP and DMAPP towards the isoprenoid biosynthesis triggering the synthesis of large number of sesquiterpenes and other higher terpenes (Bohlmann et al., 1998; Ferriols et al., 2015; Sawai and Saito, 2011). In genus *Panax*, FPS converges the precursors IPP and DMAPP towards ginsenoside biosynthesis and is an important enzyme for ginsenoside biosynthesis. Thus heterologous production of PsFPS is an important step in the synthetic biology/ metabolic engineering of ginsenosides.

5.9. Future works

Understanding of molecular mechanisms of genes and their regulations is a prerequisite for metabolic engineering in plants. These genes could be used for genetic engineering especially in cell and tissue cultures for increasing ginsenoside contents. SSH cDNA libraries helped in identifying differentially expressed genes in leaf and rhizome tissues of *P. sokpayensis*. Also using degenerate primer approach, we could obtain partial

fragments of different ginsenoside biosynthetic pathway genes, namely, a*cetyl-CoA C-acetyltransferase* (*AACT*), *3-hydroxy-3-methylglutaryl coenzyme A synthase* (*HMGS*), m*evalonate kinase* (*MVK*), *phosphomevalonate kinase* (*PMVK*), *1-deoxy-D-xylulose-5-phosphate synthase* (*DXS*), *1-deoxy-D-xylulose 5-phosphate reductoisomerase* (*DXR*), *2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase* (*CMS*), *4-*(*cytidine 5'-diphospho*)*-2-C-methyl-D-erythritol kinase* (*CMK*), *2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase* (*MCS*), *4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase* (*HDS*), *4-hydroxy-3-methylbut-2-enyl diphosphate reductase* (*HDR*), *isopentenyl diphosphate isomerase 2* (*IDI2*), *squalene epoxidase* (*SE*), *β-amyrin synthase* (*β-AS*), *protopanaxadiol synthase* (*PPDS*), *protopanaxatriol synthase* (*PPTS*) and *cycloartenol synthase* (*CS*). The rich repository of ESTs generated in this thesis work would be a valuable resource for gene discovery in *P. sokpayensis* which would complement the available transcriptomes from other *Panax* species. In future, the current work could be followed up with the following works to enhance our knowledge on ginsenoside biosynthesis:

1. Full length cloning of other genes of the pathway using partial sequences generated in this study.
2. Promoter analysis of differentially expressed genes.
3. Correlation of gene expression with ginsenosides production by studying the regulation of genes involved in ginsenoside biosynthesis.
4. Mining of isoforms of regulatory genes involved in ginsenoside biosynthesis.
5. Strategies could be devised to transplant the ginsenoside biosynthetic pathway in heterologous system.

Summary

The present work describes the cloning and characterization of ginsenoside biosynthetic pathway genes from *P. sokpayensis*, an important medicinal plant from Sikkim Himalaya. In addition, major ginsenosides were also quantified in the *P. sokpayensis* rhizome. The summary of works covered by the current thesis is as follows:

1. We generated and analyzed subtractive libraries of high quality ESTs representing leaf and rhizome tissues of *P. sokpayensis*. This led to the identification of tissue specific ESTs involved in plant growth and development, immunity and defence. Functional annotation identified some ginsenoside biosynthetic pathway genes, namely, *farnesyl pyrophosphate synthase* (*PsFPS*), *squalene synthase* (*PsSS*) and *dammarenediol synthase* (*PsDS*), which might have a regulatory role in the biosynthesis of different ginsenosides. Some ESTs neither showed significant similarity to proteins in the NCBI database nor were found in the transcriptomes of other *Panax* species. 13.75 % of unigenes from the leaf SSH library were not represented in the leaf transcriptome of *P. ginseng* and around 18.12, 23.75, 25, and 6.25 % of unigenes from the rhizome SSH library were not represented in the root/rhizome transcriptomes of *P. ginseng*, *P. notoginseng*, *P. quinquefolius* and *P. vietnamensis*, respectively, implying that these were novel and rarely expressed transcripts that can complement the available *Panax* transcriptomes. Moreover, they are valuable resources for gene discovery in *P. sokpayensis*.
2. Seventeen partial genes belonging to MVA, MEP, ginsenoside biosynthetic and phytosterol biosynthetic pathways were amplified using degenerate primers. These amplicons belong to a*cetyl-CoA C-acetyltransferase* (*PsAACT*), *3-hydroxy-3-methylglutaryl coenzyme A synthase* (*PsHMGS*), m*evalonate kinase* (*PsMVK*), *phosphomevalonate kinase* (*PsPMVK*), *1-deoxy-D-xylulose-5-phosphate synthase* (*PsDXS*), *1-deoxy-D-xylulose 5-phosphate reductoisomerase* (*PsDXR*), *2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase* (*PsCMS*), *4-*(*cytidine 5'-diphospho*)*-2-C-methyl-D-erythritol kinase* (*PsCMK*), *2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase* (*PsMCS*), *4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase* (*PsHDS*), *4-hydroxy-3-methylbut-2-enyl diphosphate reductase* (*PsHDR*), *isopentenyl diphosphate isomerase 2* (*PsIDI2*), *squalene epoxidase* (*PsSE*), *β-amyrin synthase* (*Psβ-AS*), *protopanaxadiol synthase* (*PsPPDS*), *protopanaxatriol synthase* (*PsPPTS*) and *cycloartenol synthase* (*PsCS*). These fragments can be used for full

length cloning and further characterization of the respective genes.

3. Seven full length genes were cloned through RACE based on partial fragments obtained through SSH library and degenerate primer appraoch. These included four genes, viz., *PsFPS*, *PsSS*, *PsSE* and *PsDS* from the ginsenoside biosynthetic pathway, one gene, viz., *PsCS* from phytosterol biosynthetic pathway, one gene, viz., *PsPMVK* from MVA pathway and one gene, viz., *PsCMK* from MEP pathway.
4. *In silico* characterizations of the above genes identified polyadenylation signals in some of them, which are involved in the 3′ end processing of pre-messenger RNA and transcription termination. Protein molecular weights, protein secondary structures, conserved domains and transmembrane regions predicted using deduced polypeptide sequences of the above full length genes using various bioinformatics tools revealed that these proteins and their domains that are important for their respective functions were highly conserved. Phylogenetic analyses inferred close evolutionary relationships among the same proteins belonging to different *Panax* species.
5. Expression studies of five genes, viz., *PsFPS*, *PsSS*, *PsSE*, *PsDS* and *PsCS* using qRT PCR showed that they were differentially expressed among the leaf, stem and rhizome tissues of *P. sokpayensis* prompting us to suggest that these genes might play regulatory roles in ginsenoside biosynthesis.
6. Expression construct of *PsFPS* in expression vector, pQE30 was successfully prepared using recombinant DNA technology. Subsequently, heterologous production of PsFPS in *E. coli* was successfully achieved. This can pave the way for genetic engineering of ginsenoside biosynthesis in future.
7. On HPLC study of major ginsenosides in the rhizomes of *P. sokpayensis*, ginsenosides Rb1, Rb2, Rd, Re, Rf, Rg1 and Rg2 could be quantified and their total content was at par with its Asian congener, *P. ginseng*. Our study provided the first scientific validation of the presence of ginsenosides in *P. sokpayensis* which could be a potential ginseng resource for commercialization from Sikkim Himalaya.