The terpenes or isoprenoids are diverse class of secondary metabolites encompassing around 30,000 known products (Connolly and Hill, 1992). Terpenoids includes metabolites that play important role in plant growth and development such as sterols, carotenoids, quinones, and hormones. Some of the important terpenoids play role in plant defense and communication (Lange et al., 2000). Accumulation of terpenoids in plants depends on environmental cues, such as high and low temperature, light regime, drought, alkalinity, salinity, UV stress, and pathogen infection (Gleadow and Woodrow, 2002; Falk et al., 2007; Ballhorn et al., 2011). These factors affect accumulation of terpenoids through modulation of associated metabolic pathways.

*P. kurrooa* (family Scrophulariaceae) is a medicinal herb known for its hepatoprotective and immuno-modulatory properties. In India, the species grows in Himalayan zone at an altitude of 3,000-4,500 m above mean sea level (Haridasan et al., 2002). The biological activity of *P. kurrooa* is attributed to the presence of prominent picrosides, PI and PII. Picrosides are terpenoids with an iridane skeleton of monoterpenoid origin and are derived from 5-carbon precursors IPP and DMAPP which are basic building blocks of all terpenoids. In higher plants, two independent pathways are associated with the biosynthesis of IPP and DMAPP (Wise and Croteau, 1998; Mahmoud and Croteau, 2002; Rodriguez-Concepcion and Boronat, 2002; Dudareva, et al., 2005). In cytosol, IPP is derived from MVA pathway whereas in plastids, IPP is formed via MEP pathway. A deeper investigation of genes of these pathways is required for understanding the regulation of picrosides biosynthesis.

Multiple external signals modulate accumulation of isoprenoids and expression of genes associated with isoprenoids biosynthesis in a coordinated manner in plants. Previous studies showed that temperature modulated accumulation of picrosides in *P. kurrooa* (Kawoosa et al., 2010; Gahlan et al., 2012). Drought is an important environment factor that is known to modulate plant growth, development and secondary metabolism. Under the present scenario of climate change, drought is expected to impact vegetation in Himalayan region. Warming of Himalaya has been reported to occur at an annual rate of 0.05-0.07 °C (Kumar, 2014). Increased temperature in Himalayan region would impact rate of precipitation and hence entire ecosystem. Therefore, drought and temperature are important abiotic stresses to be pursued. Since there is limited molecular data on *P. kurrooa*, the present work was carried out in the species with the following objectives.
1. To study the effect of temperature (15 °C and 25 °C) on the transcriptome of leaf tissue of *P. kurrooa*.

2. To study the effect of drought on picrosides content and gene expression under the above specified temperatures.

3. To characterize selected genes and promoter associated with picrosides biosynthesis pathway.

The following section will discuss the results obtained in the present work vis-à-vis the above objectives.

### 5.1 *De novo* sequencing and characterization of *P. kurrooa* transcriptome at two temperatures showed major transcriptome adjustments

Transcriptome represents the expressed portion of the genome and offers an overall view of the transcribed genes. Various techniques such as microarray, SAGE and massively parallel signature sequencing emerged for high throughput gene expression profiling and to allow for simultaneous interrogation of gene expression on a genome-wide scale (Tyagi, 2000). However, these techniques are time consuming and become expensive, particularly for the analysis at global level. Also, biases are introduced by the inevitable cloning step (Wang, 2007).

With the advent of NGS technology, transcriptome analysis takes lesser time, cost and labor, and at the same time provides major sequence coverage and depth (Rudd, 2003). The technology has been used with success for the analysis of transcriptome of several plant species including *Cajanus cajan* (Dubey *et al*., 2011), *Arabidopsis* (Joanes-Rhoades *et al*., 2007), *Medicago truncatula* (Cheung *et al*., 2006), *Zea mays* (Ohtsu *et al*., 2007), *Hordeum vulgare* (Wicker *et al*., 2008), *Lycopersicum esculentum* (Moxon *et al*., 2008), *Camellia sinensis* (Paul *et al*., 2014), *Cicer arietinum* (Garg *et al*., 2011), *Catharanthus roseus* (Verma *et al*., 2014) to name a few. Large scale data generated through NGS technologies provides knowledge for basic research and have been utilized for understanding of plant sex determination process (Guo *et al*., 2010), identification of tissue specific genes (Libault *et al*., 2010), understanding metabolic process and pathways in plants (Gahlan *et al*., 2012), events of functional conservation, sub functionalization for orthologous genes (Libault *et al*., 2010), identification of genes involved in agronomically important traits, understanding photosynthetic development and transcriptional complexity and evolution of meiosis (Chen *et al*., 2010; Severin *et al*., 2010) etc.
5.1.1 Reads generation and de novo sequence assembly

The de novo assembly of short reads without a reference genome still remains a challenge in spite of the development of many bioinformatics tools for data assembly and analysis (Zerbino and Birney, 2008; Simpson et al., 2009). PE run of 36 cycles, for each of the leaf tissues collected from plants exposed to 15 °C and 25 °C, was performed on Illumina genome analyzer IIx platform (Illumina, USA). Bcl converter was used to produce the reads in qseq format of the PE run of genome analyzer. It contains reads, their coordinates, tile number and quality encoding. Since 3’ ends of reads are prone to sequencing error, for every 36 bp read, only 33 bases (excluding the 3 bases at 3’ end) were considered for further use. In order to select the most appropriate k-mer size for considering de novo assembly, SOAPdenovo was run at different k-mer size ranging between 19 to 29 mers, with read length of 33 bp (Tyagi, 2000). The parameters recorded were the total transcripts obtained after assembling, average coverage, average transcript size, percentage of transcripts having length higher than 1,000 bp and highest transcript length. K-mer size of 23 mer emerged as the best choice for performing assembly, as it had a balance between over-represented and under-represented transcript numbers, coverage, maximum length obtained and average transcript length. Total number of transcripts decreased linearly with the increment of k-mer size suggesting over-representation at lower k-mer and under-representation at higher k-mers. It was observed that the sequences assembled at higher k-mer were enriched for transcripts with higher coverage/higher expression. For assembly process, only those reads were considered that produced high frequency k-mer. PE module was used to perform more sensitive assembling, and utilized pair information and approximate distance between PE reads (200 bp). The data from the two temperature conditions were assembled separately. A total of 31,338 and 63,718 assembled transcript sequences were obtained for 15 and 25 °C. The difference in the number of assembled transcripts generated could be the result of the stress response of the plants at 25 °C (Larkindal and Vierling, 2008). It is also likely that such differences in the number of assembled transcripts could be due to technological noise that might have crept in at some stage (Ewen-Campen et al., 2011). Considering the quality of assemblies produced at two different temperatures, it appeared that increased read data could enhance the total coverage, average contig length and percentage of transcripts longer than 1 kb. Therefore, PE reads of the two lanes (15 °C and 25 °C) were combined, retaining the PE information. Combined de novo assembly was carried out with 64,822,684 filtered reads in PE form (Table 4.1). With this approach, a total of 74,336 transcripts were obtained. Therefore, total 74,336 transcripts, generated from pooled reads, made the final representatives for assembled sequences in this
study. The related read and sequence (contigs) data have been deposited at NCBI in the SRA and TSA database under the accession numbers SRA048843.1 and JR808536-JR842818, respectively. Those sequences which were either less than 200 bp or had a stretch of “N” > 14 nucleotides were not accepted by TSA and is available at URL http://scbb.ihbt.res.in/Picro_information.

5.1.2 Validation of assembled sequences against the ESTs of *P. kurrooa*

The assembled sequences obtained from pooled reads were validated by sequence based alignments against ESTs of *P. kurrooa* and CYP specific ESTs generated from leaf tissue of *P. kurrooa*. BLASTN analysis against the assembled transcripts was performed. Significant hits were observed for 417 sequences (83.4 %), while no hit could be obtained for 83 ESTs from the assembled transcript set with an E-value threshold of $10^{-05}$. However, no significant hits could be obtained for CYP specific ESTs. The possible reasons for unmappable unigenes in *P. kurrooa* might be due to presence of fusion transcripts, relatively short and low quality singletons, UTR sequences far from the translation start or stop sites (> 1,000 bp), and those having incomplete coverage by the genome. It has been reported that even in *A. thaliana* around 13 % of the ESTs could not be aligned to the predicted genes (Weber *et al.*, 2007) and in human only 64 % of the reads could be mapped to the RefSeq database of well annotated human genes (Mane *et al.*, 2009).

5.1.3 Functional annotation and classification of *P. kurrooa* transcriptome

Annotation against GO database yielded significant annotation for 31,959 out of 72,220 assembled sequences, representing the best possible hits and are classified into two major categories namely, biological process and molecular function. Further classification of these assembled sequences in biological process category (Fig. 4.1A) showed the preponderance of transcripts for metabolic process, transport, regulation of biological processes, response to stimulus and cellular process indicating that the plant is undergoing rapid growth and extensive metabolic activity. Categorization of assembled sequences into molecular function category showed dominance of transcripts involved in DNA binding, catalytic and transferase activity (Fig. 4.1B) indicating dominance of gene regulation, signal transduction and enzymatically active processes.

Classification of *P. kurrooa* transcriptome in enzyme classes indicated the preponderance of transcripts belonging to serine/threonine protein kinase enzyme class alone (14.6 %). Serine/threonine protein kinases belong to a superfamily of protein kinases that phosphorylate serine and/or threonine residues. Phosphorylation of proteins by serine/threonine protein kinases serves as an important regulatory and signaling mechanism
in many processes that transmit external signals regulating cellular growth and differentiation such as ethylene signal transduction (Kiber et al., 1993), leaf and flower development (Roe et al., 1993), resistance to pathogens (Martin et al., 1993), to name a few. Classification of assembled transcriptome sequences into KEGG pathways showed representation of sequences belonged to plant-pathogen interaction pathways (6.13 %) followed by ribosome, spliceosome, protein processing and endoplasmic reticulum, starch and sucrose metabolism, ubiquitin mediated proteolysis, aminoacyl-tRNA biosynthesis, RNA degradation and so on. Highest represented groups included many pathways associated with housekeeping processes as well as plant development and secondary metabolism.

5.1.4 Utilization of transcriptome data for analysis of GC content and identification of SSRs markers

NGS offered an opportunity for the analysis of GC content among unigenes and expanded the scope for molecular markers such as SSRs. GC content gives important indication about the genes and genomic composition including evolution, gene structure (intron size and number), gene regulation and is an indicator of stability of DNA (Carels et al., 1998). Average GC content of P. kurrooa transcripts was 44.6 %, which is in range of GC levels of coding sequences in dicots (44-47 %). SSRs or microsatellites markers have diagnostic and functional significance, and have been usually associated with functional and phenotypic variations (Subramanian et al., 2003). SSRs are multi-allelic in nature, reproducible, highly abundant, cover the genome extensively and exhibit co-dominant inheritance. Transcriptome SSR markers exhibit high interspecific transferability (Wei et al., 2012). Due to the limitation of genomic data available, EST databases have been increasingly screened for development of genic SSRs (Scott et al., 2000; Saha et al., 2001). P. kurrooa is a cross pollinated species and hence the seed raised population will have variability. Variability in vegetative growth as well as for picrosides content of wild populations of P. kurrooa has also been reported (Purohit et al., 2008). Also, considerable variation exists in picrosides content for plants collected from different locations (Katoch et al., 2011). The identification of SSRs in P. kurrooa will help in distinguishing closely related individuals and will also provide useful criteria for enriching and analyzing variation in the gene pool of the plant.

A total of 1,562 SSRs were identified in the assembled transcripts. The trinucleotide SSRs have been observed to be the most prevalent group of markers (45.63 %) with highest occurrence of followed by mononucleotide (35.25 %) and dinucleotide (21.29 %) SSRs. In general, trinucleotides SSRs are the most common ones as compared to dinucleotides or tetranucleotides (Varshney et al., 2002).
5.1.5 Temperature responsive transcriptome of *P. kurrooa*

Our previous work showed the importance of temperature in regulating picrosides accumulation. Picrosides accumulation at 15 °C was higher by 22 % as compared to at 25 °C (Kawoosa *et al*., 2010). Thus temperature responsive transcriptome was studied at these two temperatures. Genome wide analysis of gene expression at two temperatures was assessed using RPKM, where read-counts of a particular transcript represent expression level (Fu *et al*., 2009). This approach was effective in detecting even scarcely expressed transcripts and was reported to be independent of prior knowledge of gene model, making it a natural choice to measure expression in the absence of known gene models and microarray chips. Considering dissimilar sequence clustering to contain over-representation, GO annotation for 19,769 unique genes was obtained along with their RPKM values at the two temperatures. Transcripts associated with monooxygenase activity, 2-iron, 2-sulfur cluster binding, cobalamin binding, beta lactamase activity, aminobutyraldehyde dehydrogenase activity, purine transmembrane transporter activity and metal ions like iron and copper binding activities were found to be over-expressed at 15 °C in molecular function category. Transcripts associated with 2-iron, 2-sulfur cluster proteins and monooxygenase activity are involved in oxidation reduction reactions in plants and play important role in biosynthesis as well as catabolism of various plant specific metabolites. Under the biological process groups, those associated with various biosynthesis and transport processes such as zinc and ammonium transport and protein chromophore linkage were prominent at 15 °C. An enrichment analysis for functional categories pointed out that transcripts associated with response to stress, response to stimulus, phytosteroid metabolic process and brassinosteroid metabolic process were significantly enriched in the group having two fold or higher expression at 25 °C (Fig. 4.6). Brassinosteroids are a group of plant steroidal hormones that regulate various aspects of plant growth and development, including cell elongation, photomorphogenesis, xylem differentiation, seed germination (Sasse, 2003) and adaptation to abiotic and biotic environmental stresses (Krishna, 2003). Brassinosteroids promote tolerance in plants against a wide range of stresses, including heat, cold, drought and salinity, possibly through up-regulating the expression of stress related genes (Divi and Krishna, 2009). Compared to the expression at 15 °C, up-regulation of stress responsive transcripts is suggestive of *P. kurrooa* to be under stress at 25 °C. The above described global analysis of gene expression provided comprehensive dataset with each gene represented by its absolute expression level at the two temperatures. Several processes such as metabolic processes, cellular processes, transport processes (Fig. 4.4), and catalytic processes (Fig. 4.5) were...
enriched at 15 °C indicating the importance of these processes in plant growth and survival; although different set of genes in these processes might determine the response of plant to temperature change. Transcripts associated with the processes involving lipid metabolism were highly enriched at 25 °C (Fig. 4.6), suggesting a change in lipid profile. Indeed, temperature has been shown to modulate lipid profile in plant (Smertenko et al., 1997).

Concomitant modulation of several plant processes suggested involvement of TFs for coordinated regulation of gene expression. TFs are sequence specific DNA-binding proteins that interact with the promoter regions of target genes and modulate gene expression. These proteins regulate gene transcription depending upon tissue type and in response to internal signals, for example plant hormones, and to external signals such as temperature, UV light, pathogen attack, and drought. In P. kurrooa, 6,305 transcript sequences exhibited homology with TF families, which were reduced to 2,500 after dissimilar sequence clustering. The most abundant TFs overrepresented at 15 °C were MYB-related, FAR1, G2-like, NAC, SNF2I, WRKY, FWA, HB, bHLH, Orphans, C3H, C2H2 and MADS. Of the total TFs, 16.9 % exhibited abundance increment two fold or above at 25 °C and belonged to C3H, MADS, bHLH, PHD and FAR1 family. The above mentioned TFs have been associated with varied processes. For example, members of C3H family are involved in embryogenesis (Li and Thomas, 1998), whereas PHD proteins are found in nucleus and regulate chromatin-mediated transcription (Aasland et al., 1995). Most members of MADS family TFs are involved in the regulation of flower-related physiological and developmental processes (Parenicova et al., 2003), whereas members of bHLH are involved in controlling cell proliferation and in the development of specific cell lineages (Heim et al., 2003). FAR1, yet another family of TFs, is involved in phytochrome signaling (Lin et al., 2007).

It was interesting to note that global gene expression analysis exhibited modulation of processes which were responsive to heat, responsive to biotic stimulus, lipid catabolic process and glycogen biosynthetic process (Fig. 4.6) at 25 °C as compared to 15 °C. Members of TFs families such as bHLH, WRKY, MYB, AP2/EREBP that are known to regulate the above processes, also exhibited modulation in accordance with the said transcripts, suggesting their role in regulating the mentioned processes. P. kurrooa is a plant of high altitude temperate region which does not tolerate high temperature for a longer period of time. A systematic analysis of these transcription factors would open door for imparting tolerance to P. kurrooa at high temperature (25 °C). It would also be worthwhile to study, how such a temperature mediated transcriptomic adjustment affects picrosides biosynthesis.
5.1.6 Pathways associated with picrosides biosynthesis exhibited temperature dependent modulation

Since the two temperatures targeted in the present work modulated picrosides content, it was of interest to analyze various genes associated with picrosides biosynthesis. Picrosides are classified as PI and PII depending upon the functional-group moieties. PI has cinnamate moiety, whereas PII has vanillate moiety. The cinnamate and vanillate moieties are derived from PP pathway (Fig. 1.4). Synthesis of cinnamate requires the action of PAL on phenylalanine whereas COMT is the key enzyme for vanillate biosynthesis (Dixon and Paiva, 1995).

Iridoid moiety is derived from GPP (Fig. 1.4). GPP is synthesized by sequential head to tail addition of IPP and its allelic isomer DMAPP (Wise and Croteau, 1998). Cytosolic MVA pathway and the plastid localized MEP pathway synthesize IPP and DMAPP (Mahmoud and Croteau, 2002; Rodriguez-Concepcion and Boronat, 2002) with cross talks between these two pathways (Newman and Chappell, 1999; Schuhr et al., 2005). Thus MVA, MEP and PP are regarded as central pathways for the synthesis of picrosides. MVA pathway starts from the condensation of acetyl-CoA (Qureshi and Porter, 1981; Newman and Chappell, 1999), whereas MEP pathway needs pyruvate and GAP (Eisenreich et al., 2001; Rohmer, 1993). Biosynthesis of picrosides involves synthesis of iridoid moiety from GPP through series of oxidation and cyclization steps followed by the condensation of glucose and cinnamate/vanillate with iridoid unit (Fig. 1.5).

Analysis of MVA, MEP and PP pathways have general implications as well. These are involved in the biosynthesis of large number of secondary metabolites including those having commercial implications. These compounds include taxol (Palazon et al., 2003), artemisinin (Souret et al., 2002), β-carotene (Hirschberg, 2001), α-tocopherol (Dellapenna and Pogson, 2006), vincristine, vinblastine and coumarins (Cragg and Newman, 2003). These pathways play important roles in growth and development including secondary metabolism, and hence identification of major regulatory steps would be key to modulate plant performance and secondary metabolism, if need be.

Using BLAST analysis against the UniProt and KEGG databases, various genes associated with MVA, MEP and PP pathways were identified. RPKM-based expression showed 2 fold increase for several genes of MEP pathways at 15 °C as compared to those at 25 °C (Fig. 4.8). Data was in agreement with the data on picrosides content that showed its increased accumulation at 15 °C (Kawoosa et al., 2010). While a previous work on P. kurrooa also showed a positive correlation between PkDXS (a gene of MEP pathway) and
picrosides accumulation (Kawoosa et al., 2010), the present work detailed on all the genes of MEP pathway highlighting their importance in picrosides accumulation (Fig. 4.8).

RPKM data showed up-regulation of various genes of MVA and PP pathway at 25 °C as compared to those at 15 °C (Fig. 4.8). Accumulation of picrosides decreased at 25 °C as compared to at 15 °C, whereas various genes of PP pathway exhibited up-regulation at 25 °C. Since PP pathway is important in supplying cinnamate and vanillate for picrosides biosynthesis, an up-regulation of various genes of the pathways were envisioned at 15 °C as compared to 25 °C. However the results were opposite, suggesting rerouting of the metabolites towards the synthesis of other metabolites at 25 °C. And at 15 °C, the observed expression of genes of the MVA and PP pathway might be sufficient enough to meet the requirement of cinnamate and vanillate. In fact, increased activity of PAL (a gene of PP pathway) in response to thermal stress was considered as an acclamatory response of cells to heat stress in Citrus vulgaris (Rivero et al., 2001). Expression of various genes of PP pathway including PAL is regulated by TF family LIM. LIM proteins have conserved cysteine-histidine rich, zinc-coordinating domain consisting of two zinc fingers repeated in tandem. In transgenic tobacco, down-regulating the expression of LIM proteins through antisense approach lowered the expression of various genes of PP pathway (Kawaoka et al., 2000). In P. kurrooa, RPKM based expression analysis showed that the expression of LIM was up-regulated at 25 °C suggesting its role in regulating PP pathway. Thus at higher temperature, up-regulation of MVA and PP pathways could have a role in temperature stress acclimation as well.

5.1.7 Transcriptome data identified CYPs and GTs as a source of hitherto unknown genes involved in the biosynthesis of picrosides

The intermediates and enzymes involved in cyclization of GPP for the synthesis of iridoid moiety and later its condensation with glucose and cinnamate/vanillate moieties are yet to be deciphered in P. kurrooa. In vivo tracer studies in Catharanthus roseus and Lonicera morrowii showed that iridoid is synthesized by cyclization of 10-oxogeranial to yield iridoial (Uesato et al., 1984). This is subsequently converted into iridoid compounds via iridotrial intermediate, and involves multiple oxidation/hydroxylation and glycosylation reactions. Most of the oxidative reactions, including hydroxylations, epoxidation, dealkylation, dehydration and carbon-carbon bond cleavage are catalyzed by CYP group of enzymes (Schuler, 2003), whereas glycosylation reactions are catalyzed by GTs. Therefore, it would be relevant to discuss CYPs and GTs in the transcriptome of P. kurrooa.
CYPs are membrane bound hemoproteins involved in array of pathways in primary and secondary metabolism. Some of the example of CYPs include lauric acid hydroxylase, limonene-3-hydroxylases (CYP71D13 and CYP71D15), (+)-menthofuran synthase, geraniol hydroxylase, camphor-6-exo-hydroxylase, cinnamate 4-hydroxylase (4-CH), flavonoid 3'-hydroxylase, flavones synthase 2- berbamunine synthase and tyrosine N-hydroxylase. Based on phylogenetic studies, plant CYPs can be divided into 10 separate clans that cover the current 61 families (Nelson et al., 2004).

Monoterpenes, sesquiterpenes and diterpenes are intimately associated with CYPs. For example, 10-hydroxylation of geraniol and nerol has long been known to be a CYP function (Meehan and Coscia 1973). Allylic hydroxylation of cyclic monoterpenes by CYPs is well documented. Such reactions are known to occur in the hydroxylation of limonene (Karp et al., 1990), pinene (Karp et al., 1992), sabinene (Karp et al., 1987), camphor (Funk and Croteau, 1993), abietin (Funk and Croteau, 1994) and terpeniol (Bolwell et al., 1994). All the CYPs associated with monoterpenes are divided into CYP71 clan. In the present study, 33 unigenes annotated as putative CYPs were identified (Fig. 4.14). Ten CYPs showed more than two fold increase in the expression and 8 CYPs exhibited down-regulation, respectively at 15 °C as compared to at 25 °C. Increased picrosides content (Fig. 4.13) and up-regulation of CYPs at 15 °C suggested these to be the possible candidates associated with picrosides biosynthesis through their possible role in cyclization of GPP and iridoid moiety as indicated in Fig. 1.5.

GTs constitutes a large family of enzymes that catalyze transfer of glycosyl group from activated sugars UDP sugars to aglycone acceptor molecules (Coutinho et al., 2003). GTs are grouped into 69 families based on the substrate recognition and sequence relatedness, of which family 1 is the largest and is over-represented UGTs. UGTs use UDP-glucose as the donor in the GT catalyzed reactions (Jones and Vogt, 2001). Therefore, the reactions involving transfer of glucose utilize UGTs. UGTs have 44-amino acid C-terminal signature motif designated as PSPG box and are encoded by large multigene families, sometime comprising several hundred of genes. For example, family 1, UGTs are encoded by 120 UGT genes in Arabidopsis and by 165 UGTs in Medicago truncatula (Gachon et al., 2005). The UGT superfamily in higher plants is thought to encode enzymes that glycosylate a broad array of aglycones, including plant hormones, all major classes of plant secondary metabolites, and xenobiotics such as herbicides (Jones and Vogt, 2001). Picrosides are present as 1-O-glucosides and hence an analysis of UGTs would be central to identify the gene associated with the glycosylation of iridoid moiety.
BLAST search identified 154 unigenes encoding GTs, out which 17 encoded for UGTs in *P. kurrooa* (Fig. 4.15). Expression of these 17 UGTs through RPKM analysis showed that 2 UGTs are up-regulated and 4 were down-regulated at 15 °C as compared to those at 25 °C. Up-regulation of picrosides at 15 °C vis-à-vis up-regulation of 2 UGTs suggested these to be the possible candidates associated with picrosides biosynthesis.

5.2 Full-length cloning and characterization of selected genes
Transcriptome sequencing produced huge amount of biological information, yet limited information is available on the derived protein sequences needed for functional analysis of cDNAs. Full-length cDNAs are valuable for gene annotation, correct annotation of transcriptional units and gene products of genomic sequence data, functional analysis of the genes and also for cloning of 5′ upstream sequences (Seki et al., 2002). Therefore, full length cDNAs involved in picrosides biosynthesis were cloned following RACE (Frohman et al., 1988). Since the development of RACE protocol (Frohman et al., 1988), RACE has been used to clone large number of genes from varied plant and animal systems (Singh et al., 2009; Singh et al., 2010; Kumar et al., 2011; Singh et al., 2013; Zhang et al., 2014). RACE is a powerful technique for obtaining 5′ and 3′ ends of cDNAs that allows the amplification of mRNA from a known internal sequence region to either 5′ or 3′ end of the mRNA target. Briefly, an adaptor with a defined sequence is attached to one end of the cDNA; then, the region between the adaptor and the known sequences is amplified by PCR.

Primers for RACE were designed from partial gene sequences obtained from transcriptome data (Gahlan et al., 2012). The 5′- and 3′-RACE sequences of nine full length cDNAs namely *PkHMGS, PkMVK, PkPMD, PkMCT, PkMDS, PkHDS, PkMTS1, PkMTS2* and *PkCOMT* were aligned and their sequences were submitted to NCBI gene databank (Appendix IV, Table 4). Detailed bioinformatics analyses of the above full-length cDNAs revealed the presence of conserved functional domain characteristic of individual protein families. Previous work indicated the importance of the identified conserved domain in functionality of the genes, for example HMGS cloned from *Brassica juncea* had HMG-CoA domain and the gene was found to be functional (Alex et al., 2000). Similarly, *MDS* from *Stevia rebaudiana* with MeCDP_synthase domain was found to be functional in complementation experiments (Kumar et al., 2012). Likewise, dihydroflavonol 4-reductase from *Camellia sinensis* was found to be functional in heterologous system. Dihydroflavonol 4-reductase catalyzes the reduction of dihydroflavonols to leucoanthocyanins, a key step in the biosynthesis of catechins (Singh et al., 2008).
Before start of the present work only 8 cDNAs of picrosides biosynthesis pathway were reported (Singh et al., 2013). Transcriptome analysis coupled with RACE further yielded 7 more full-length cDNA and 5 partial cDNAs. Thus, at present we have a total of 15 full-length and 5 partial cDNAs of MVA, MEP and PP pathways which are associated with picrosides biosynthesis (Fig. 5.1).

Biosynthesis of picrosides involves cyclization of GPP to iridoid wherein MTS would play a central role (McGarvey and Croteau, 1995). Two putative MTS were cloned (accession numbers KJ187308 and JX119191) in the present thesis, which are yet to be analyzed for functionality.

Generally various genes of a pathway are regulated by common TFs, for example TFs bHLH and MYB regulate the genes associated with anthocyanin biosynthesis in maize in a coordinated manner (Von Endt et al, 2002). Similarly, in Catharanthus roseus, AP2/ERF domain containing TFs regulates several genes associated with terpenoid indole alkaloid biosynthesis in response to jasmonic acid (Van der Fits and Memelink, 2000). TFs regulate various genes through binding of the protein on cis-acting elements which are usually present at 5' upstream region of the gene. An analysis of upstream sequence of PkHMGR, PkDXS and PkCMK showed the presence of WRKY domain suggesting WRKY could be a key TF regulating gene expression (Fig. 5.1; Kawoosa et al., 2010; Kawoosa, 2012).
5.3 Functional characterization of selected genes of *P. kurrooa*

Although, cloned genes had requisite catalytic domains, evaluation of the encoded protein is important to establish them as functional genes. Of the total 15 full-length genes, the expression of *PkDXS* was in accordance with the picrosides content (Kawoosa *et al.*, 2010) and hence functionality of *PkDXS* was evaluated in heterologous system. DXS has been cloned and functionally characterized from several plant species such as *Zea mays* (Cordoba *et al.*, 2011), *Ammomum villosum* (Yang *et al.*, 2012), *Picea abies* (Phillips *et al.*, 2007), *Stevia rebaudiana* (Totte *et al.*, 2003), *Medicago truncatula* (Walter *et al.*, 2002), *Arabidopsis* (Paul *et al.*, 2012), *Populus trichocarpa* (Banerjee *et al.*, 2013), grapevine (Battilana *et al.*, 2011). The enzyme catalyzed irreversible condensation of (hydroxyethyl) thiamine derived from pyruvate with the C1 aldehyde group of GAP to produce DXP (Sprenger *et al.*, 1997; Lange *et al.*, 1998; Lois *et al.*, 1998).
DXS (EC 2.2.1.7) has been reported to be a regulatory enzyme in isoprenoids biosynthesis pathway. The enzyme showed a positive correlation with the accumulation of carotenoids in *Lycopersicon esculentum*. Over-expression and silencing of DXS was correlated with changes in the levels of isoprenoids end products namely α-tocopherol, chlorophylls, carotenoids, and ABA (Este´vez *et al*., 2001) in *Arabidopsis*. Over-expression of a bacterial DXS gene in potato tubers perturbed and enhanced the phytoene content to 7-fold (Morris *et al*., 2006).

DXP served as a precursor in the biosynthesis of isoprenoids derived from the MEP pathway. These include hemiterpenes, monoterpenes such as geraniol, menthone, pulegone, thymol etc, monoterpenes iridoid glycosides such as secologanin, artemisiaketone, chrysanemic acid etc, diterpenoids such as taxol, labdane, steviol, lipiferolide, phytol, salvinorin A, chlorophylls, carotenoids and plastoquinone, ABA (Lichtenthaler, 1999). In addition to it, DXS also serves as precursor for synthesis of few sesquiterpenes such as germacrene, anthecotuloid, epoxide hodgsonox from *Lepidolacna hodgsoniac*. DXP also served as a precursor in the biosynthesis of vitamin B1 (thiamine) in plastids (Julliard and Douce, 1991).

Realizing the importance of DXS in isoprenoids biosynthesis, the function of the cloned gene from *P. kurrooa* was evaluated in the present thesis. Also, protein binding ability of the two TFs *PksWRKY* and *PkdWRKY* was evaluated in the present because of their possible involvement in regulating picrosides biosynthetic pathway genes (Fig. 5.1).

### 5.3.1 Functional characterization of PkDXS in *E. coli*

Expression of PkDXS in *E. coli* resulted in appearance of the expressed protein as inclusion bodies (IBs). Therefore, there was need to solubilize and refold the protein to assess the functionality of the enzyme.

Several reagents are used to solubilize IBs, these include urea, guanidine hydrochloride, SDS, N-cetyltrimethylammonium chloride and N-lauroylsarcosine. Choice of the solubilizing agent depends on the protein in question and downstream purification process. Detergents such as SDS, N-cetyltrimethylammonium chloride and N-lauroylsarcosine interfere with the downstream purification process by binding with the columns and filters. Also, these solubilize membrane proteases along with IBs leading to poor yield of the target proteins. Urea and guanidine hydrochloride have been reported to offer advantages over the other reagents in terms of offering flexibility (Burgess, 1996; Stockel *et al*., 1997). Therefore, urea was selected to solubilize IB of PkDXS.
Since PkDXS has theoretical pI of 8.60 as deduced from its amino acid sequence, pH of various buffers used in the solubilization, refolding and enzyme assay was maintained at 8.0. PkDXS was solubilized in 8 M urea (pH 8.0) in the presence of reducing agent. DTT reduce the formation of undesirable inter- and/or intra-molecular disulfide bonds formed due to air oxidation during cell disruption (De Bernardez and Clark, 2001).

Solubilization of IBs in urea results in proteins devoid of their native conformation. For a protein to be functionally active, refolding of the solubilized protein was carried out by transferring the protein into conditions that allow formation of its native structure (e.g. low denaturant concentration). At high denaturant concentrations, proteins are unfolded, disordered, well solvated, and flexible. Therefore procedure involving removal of urea from denatured proteins is an important step in the recovery of the proteins. Many protocols have been developed for refolding of a functionally inactive protein to an active protein, these include dialysis, dilution, size-exclusion chromatography (Li et al., 2004), reversed micelle systems (Sakono et al., 2004), zeolite absorbing systems (Nara et al., 2009), and natural GroEL–GroES chaperone system (Weissman et al., 1994).

Solubilized PkDXS was slowly diluted with refolding buffer. Our data showed that the best results were obtained using refolding buffer 1 (0.5 M arginine). L-arginine is one of the most commonly used refolding agent. It has been used for refolding of proteins such as casein kinase II (Lin and Traugh, 1993), gamma interferon (Arora and Khanna, 1996), p53 tumor suppressor protein (Bell et al., 2002), and interleukin-21 (Asano et al., 2002). L-arginine functions by increasing the solubility of protein by shielding the hydrophobic regions and thus impeding aggregate formation. Arginine is generally used in a concentration range of 0.1-2.0 M in the in vitro refolding buffer at pH 8.0-8.5 (Umetsu et al., 2003).

The other refolding aiding agents such as glycerol stabilize proteins by enhancing the rate of in vitro protein refolding. The glycerol is known to increase the solubility and hence refolding of proteins even at high protein concentration (Gekko and Timasheff, 1981; Sawano et al., 1992). Glycerol functions as a protein stabilizer by increasing the solvent ordering and hydrophobic interactions of the protein chains. Similarly, PEG has been reported to stabilize protein structures (Nian et al., 2009). For example, PEG was used as refolding aiding agent in refolding of several proteins such as insulin-like growth factor (Wang et al., 2006) and interferon (Hart et al., 1994). PEG aids the refolding process by binding to the intermediates of refolding pathway and decrease the rate of protein aggregation by interacting with hydrophobic side chains of denatured proteins.
A number of methods have been developed for measurement of DXS activity. The most common assay method employed for determining the activity of DXS involved measurement of radioactivity incorporated into the product DXP from radiolabeled pyruvate (Sprenger et al., 1997; Lois et al., 1998). Coupled spectrophotometric assay and fluorescence detection methods were developed to avoid use of radioactivity (Han et al., 2003). Recently, a method was developed by Banerjee et al. (2013) that involves measurement of DXP produced by LC-MS/MS. The method involves direct measurement of enzyme activity that uses selectivity and sensitivity of UPLC for separation of product from precursors. The enzymatic production of DXP resulted in a peak at retention time of 1.59 min that was confirmed by running a standard under the same conditions as that of test reaction. Thus, data clearly showed that PkDXS is a functional gene and has the ability to catalyze the conversion of GAP and pyruvate to DXP. Our data was in line with the work done by Banerjee et al. (2013).

5.3.2 Functional evaluation of TFs PksWRKY and PkdWRKY in Y1H

The signature feature of WRKY TFs is the presence of highly conserved DNA-binding WRKY domain (WRKYGQK) at the N-terminus (Rushton et al., 1996). WRKY TFs binds to the cis-element (TTGACC/T) known as the W-box of the target gene (Yamasaki et al., 2005) and modulate their expression. A number of W-box motifs were identified in the putative promoter regions of PkHMGR, PkDXS and PkCMK. The frequency of W-box in promoter of PkHMGR, PkDXS, and PkCMK is 11, 7 and 12, respectively (Kawoosa, 2012). It is to be noted that WRKY box appeared with highest frequency in the upstream region of these genes.

Functionality of PksWRKY and PkdWRKY was assessed by Y1H. Y1H is a powerful technique to study the interaction of TF with a specific regulatory DNA sequence of interest. Data clearly showed that both PksWRKY and PkdWRKY encoded functional protein since the interaction of functional WRKY protein with W box resulted in yeast auxotroph’s to grow on a minimal medium lacking histidine. Thus, Y1H confirmed the prediction of functional protein encoded by putative WRKY genes. Although both the TFs exhibited DNA binding capability, it will be worthwhile to assay transcriptional activation/repressor activity of these TFs.

5.4 Molecular and physiological analysis of drought stress in P. kurrooa

Drought stress is one of the most significant abiotic stresses limiting growth and productivity of plants by altering the metabolism at physiological, biochemical and molecular levels (Yue
Physiological changes under drought stress are often reflected at the transcription level, where the transcript levels of genes associated with growth and development of plants were altered. Metabolic adjustments in response to drought stress helps restore chemical and energetic imbalances and is crucial to acclimation and survival.

Drought has been shown to alter biosynthesis of primary as well as secondary metabolites in plants. Changes have been observed in plant hormones such as ABA, GA₃, cytokinins and brassinosteroids; photosynthetic pigments such as carotenoids, tocopherols and flavonoids; emission of terpenes such as isoprene and diterpenes (Demming-Adams and Adams, 1996; Niyogi et al., 1997; Munne-Bosch and Alegre, 2000; Beckett et al., 2012).

The response of *P. kurrooa* plants to drought stress is not yet known and present thesis is the first effort to analyze the response of plant to drought stress at physiological, biochemical and molecular level. Impact of drought was studied at 15 °C and 25 °C. As discussed previously, temperature modulated picrosides biosynthesis at molecular and biochemical level. High temperature results in decreased picrosides biosynthesis with concomitant down-regulation of genes of the pathway and global change in the transcriptome (Kawoosa et al., 2010; section 4.1). Molecular and metabolic responses of plants to a combination of drought and heat stress are unique (Mittler, 2006) and cannot be extrapolated from the response of plant to individual cues. Since temperature increase under drought is inevitable and that present work was carried out in plant growth chamber, the effect of drought was studied at 15 °C as well as 25 °C.

### 5.4.1 Physiological and biochemical response of *P. kurrooa* to abiotic stresses

With progressive drought, SMC and RWC exhibited significant decrease, whereas REL exhibited significant increase both at 15 and 25 °C as compared to respective control of the study in response to drought. A decrease in RWC under drought stress was observed in other plant species as well. RWC decreased by 30 and 16 % in maize seedlings and apple plants in response to drought at 24 h and at 50 % SMC, respectively (Valentovic et al., 2006; Bolat et al., 2014). RWC exhibited decrease by 66 and 80 % in drought stressed *Solanum lycopersicum* (Loyola et al., 2011) and *Agrostis stolonifera* (Merewitz et al., 2011), respectively at day 18 of drought stress and when SMC was recorded to be 5 %. Thus, drought mediated decrease in RWC is dependent on the plant species under study and extent of drought stress. An increase in REL also varied in different plant species. REL in drought stressed maize and apple increased by 390 % and 33.56 % at 24 h and at 50 % SMC, respectively (Valentovic et al., 2006; Bolat et al., 2014), whereas a 250 % increase in REL was observed in drought stressed *Agrostis stolonifera* at 5 % SMC. An increase in REL by
226.09 %, and a decrease in RWC by 36.96 % were observed in tea plants subjected to osmotic stress as a result of PEG treatment (Paul et al., 2014).

Accumulation of proline was significantly higher in response to drought stress at both the temperature studied in *P. kurrooa*. Proline is a well-known compatible osmolyte, antioxidant, ROS scavenger, molecular chaperone and an osmoprotectant (Rhodes et al., 1986; Hare and Cress, 1997; Verbruggen and Hermans, 2008; Szabados and Savoure, 2010). Proline accumulates to very high concentrations in plants experiencing osmotic stress (Szabados and Savoure, 2010). Proline accumulated from 0.69 to 26.1 μmoles/g FW in tobacco plants exposed to drought stress after 10 days (Kishor et al., 2005). Proline concentration increased from 0.5 to 2.3 μmol/g FW in drought stressed rice (Hien, et al., 2003). A role of proline metabolism in turgor maintenance (Blum, 2005; DaCosta and Huang, 2006) and redox buffering (Szabados and Savoure, 2010) is also proposed that possibly helps in acclimation and survival of plants under stress.

### 5.4.2 Picrosides content under water deficit stress

Plants respond to drought by progressive adjustment of their metabolism that possibly helps in acclimation and survival of plants under stress. Metabolic adjustment in response to unfavorable conditions is a dynamic and multifaceted process. It depends on the type and strength of stress and also on the plant species under study.

In the present study, drought stress leads to accumulation of picrosides in leaf tissue of *P. kurrooa* (Fig. 4.63, 4.64) at both the temperatures under study. Accumulation of several secondary metabolites has been reported in plant species experiencing drought stress. Increase in endogenous α tocopherol and carotenoids levels has been reported in plant species experiencing drought stress (Havaux, 1998; Noctor and Foyer, 1998; Asada, 1999; Smirnoff and Wheeler, 1999). Tocopherols and carotenoids exhibit antioxidant activity in the chloroplast by scavenging singlet oxygen and lipid peroxy radicals. Thereby, maintaining integrity of the photosynthetic membranes under oxidative stress such as those caused by drought.

Abietane diterpenes such as carnosic acid, rosmanol, isorosmanol were also accumulated in the drought stressed leaves of *Salvia officinalis* and *Rosmarinus officinalis* (Bosch and Alegre, 2003). *In vitro* studies showed antioxidant properties of these compounds which protected photosynthetic membranes against oxidative damage (Aruoma et al., 1992).

Drought stressed *Salvia officinalis* plants have higher monoterpene content as compared to control plants (Nowak et al., 2010). Induction of glucosinolates accumulation by drought conditions has been reported as part of the plant response to stress through the
process of OA (Schreiner, et al., 2009). Iridoid glycosides such as catalpol, aucubin, harpagide, and harpagoside accumulated in drought stressed *Scrophularia ningpoensis* plants (Wang et al., 2010). Catalpol and aucubin are also present in *P. kurrooa*, but in small concentration (Kumar et al., 2005).

Plant hormones are also modulated under drought stress. GA$_3$ and ABA shows increased accumulation, whereas cytokininins show a decline (Behringer et al., 1990; Davies and Zhang, 1991; Rivero et al., 2010). Increased ABA and decreased cytokininins induced stomatal closure that helped reduction of water loss but also inhibit photosynthesis.

Antioxidant activity of different fractions of leaf extract of *P. kurrooa* has been established (Kant et al., 2013). High antioxidant activity has been reported from butanol and ethyl acetate fraction as compared to ethanol fraction (Kant et al., 2013). Methanolic extracts of leaf tissue of *P. kurrooa*, was reported to consist of P1, whereas ethyl acetate fraction, contained PII (Bhandari et al., 2009). Increase in picrosides content in leaf tissue could be an adaptive mechanism of plant to oxidative stress induced by drought stress.

In contrast to leaves, picrosides content in rhizome tissue decreased with increase in stress (Fig. 4.63, 4.64), which could be due to re-routing of the pathway toward synthesis of various metabolites other than picrosides, since the pathway is involved in synthesis of several other metabolites such as hemiterpenes, monoterpenes, diterpenes, sesquiterpenes, triterpenes etc. Decrease in metabolite content in rhizome tissue in drought stress could also be due to catabolism of iridoid glycosides.

**5.4.3 Effect of water stress on expression of genes of picrosides biosynthetic pathway**

The expression of genes involved in picrosides biosynthesis was studied to understand the molecular basis of changes in picrosides content in leaf and rhizome tissues.

In the leaf tissue, expression of most of the genes of MVA and MEP pathway, with the exception of *PkCMK*, *PkHMGS* and *PkAACT* was up-regulated at day 10 of drought stress (Fig. 4.65A). An increase in gene expression could explain higher picroside content in the leaf tissue at day 10 of withholding irrigation. *DXS*, *DXR* and *HDR* of MEP pathway and *HMGR* of MVA pathway are the rate limiting genes for the synthesis of isoprenoids precursors IPP and DMAPP (Harker and Bramley, 1999; Kuzuyama et al., 2000; Matthews and Wurtzel, 2000). An early induction followed by decrease in expression of *DXS*, *HMGR* and *GPS* with stress were also reported in *Cistus creticus* (Pateraki and Kanellis, 2010). Whereas a decrease in expression of *DXS* and *IPI* during drought was reported in tomato plants (Loyola et al., 2011). It is likely that during early period of drought stress increased expression of MVA and MEP pathway genes promoted synthesis and accumulation of IPP.
and DMAPP substrates, respectively. These substrates could be used by downstream enzymes for the synthesis of various metabolites during drought stress as proposed previously (Loyola et al., 2011).

DXS is reported to be sensitive to feedback inhibition from metabolites of the pathway (Wolfertz et al., 2004). Increased accumulation of IPP and DMAPP could result in feedback inhibition of DXS leading to overall decrease in the metabolite flux through MEP pathway. This is caused by likely response of decrease in expression of PkDXS and other genes of MEP pathway at day 20 and 30 of withholding irrigation (Fig. 4.65A).

Apart from affecting the key genes of MEP pathway, regulatory genes of MVA pathway were also affected by drought stress. Initial increase followed by decrease in HMGR expression could be due to increased levels of ABA in drought stressed plants (Davies and Zhang, 1991). ABA has been reported as a negative regulator of HMGR activity (Brooker and Russell, 1979; Moore and Oishi, 1994; Cowan et al., 1997). Yet another key enzyme of MVA pathway, PkHMGS was found to be down-regulated under drought stress in P. kurrooa (Fig. 4.65). HMGS has been reported to be down-regulated under osmotic stress in Brassica juncea (Alex et al., 2000).

PP pathway is involved in the synthesis of diverse phenolic metabolites such as flavonoids, tannins, hydroxycinnamate esters and the structural polymers lignin. Phenolic metabolites play an important role in protecting plants under stress (Dixon and Paiva, 1995; Grace and Logan, 2000). No significant change in expression of PkPAL and PkCOMT was observed, however the expression of C4H increase at day 30 of stress (Fig. 4.6). The cinnamate and vanillate moieties in picrosides are derived from PP pathway (Fig. 1.4). Synthesis of cinnamate requires the action of PAL on phenylalanine whereas COMT is the key enzyme for vanillate biosynthesis. The initial three steps of PP pathway are catalyzed by PAL, C4H and 4-coumaroyl CoA ligase and offer the mother molecule for synthesis of several other metabolites. The expressions of these three genes are species specific and vary with stress intensity/duration of stress (Vincent et al., 2005).

It was worth noting that in spite of decreased expression of various gens of the pathway picrosides content increased in leaf tissue. Previous studies showed the existence of positive co-relation between gene expression and picrosides content in response to different temperatures (Gahlan et al., 2012). A negative co-relation between gene expression and picrosides content in response to drought stress could be possibly due to re-routing of pathway flux for the synthesis of metabolites other than picrosides. Possibility also arises for catabolism of picrosides. Though this is not yet reported for picrosides, transport and
catabolism of monoterpenes glycosides to mixture of ether, methanol and water soluble metabolites has been shown in *Mentha* and *Salvia* (Croteau, *et al*., 1985).

In rhizome tissue the expression of most of the genes of MEP, MVA and PP pathway genes was up-regulated at day 20 of stress imposition and remained high thereafter (Fig. 4.65 and 4.66). Expression of *PkHMGR* decreased during drought stress in rhizome. Expression of *PkPAL* and *PkHMGS* was high during day 10 of stress and thereafter expression was comparable to that of control plants (Fig. 4.65 and 4.66). Increased expression of *COMT* was also reported in needles of *Pinus pinastres* (Costa *et al*., 1998) and in roots of *Arabidopsis* experiencing osmotic stress (Bianchi *et al*., 2002).

In spite of increased expression of majority of MVA, MEP and PP pathways genes during drought stress, picrosides content decreased in rhizome tissue. MVA, MEP and PP are universal pathways for the biosynthesis of all plant isoprenoids such as xanthophylls, carotenoids, tocopherols, isoprene, abscisic acid and these metabolites impart adaptive advantage to plants under stress conditions (Thameur *et al*., 2011). Re-routing of MEP pathway towards synthesis of ABA and other isoprenoids in improving stress tolerance under drought stress has been reported (Thameur *et al*., 2011).

Our data suggested that drought stress at higher temperature enhanced the rate of change in various physiological processes under drought condition. There was no apparent qualitative change of the parameters under the present work.

**5.5 Functional evaluation of promoter of *PkDXS***

Accumulation of secondary metabolites in plants is affected by several environmental factors such as light intensity and drought. Of the several genes, *DXS* has been reported to be a regulatory gene determining terpenoids content (Cordoba *et al*., 2009). Expression studies showed that DXS played important regulatory role in biosynthesis of IPP and DMAPP. Over- and down-expression of *DXS* in transgenic *Arabidopsis* resulted in altered content of chlorophyll, carotenoids, tocopherols, and abscisic acid (Estévez *et al*., 2001). Similar results were obtained in other plants including tomato (Lois *et al*., 2000; Enfissi *et al*., 2005), potato (Morris *et al*., 2006) and *Ginkgo biloba* (Gong *et al*., 2006). Our work showed a positive correlation between picrosides content and expression of *PkDXS* under light and low temperature (15 °C) (Kawoosa *et al*., 2010; Gahlan *et al*., 2012). The expression of *PkDXS* was also modulated in response to drought stress in *P. kurrooa*.

*In silico* analysis of upstream region of *PkDXS* (*PropkDXS*) indicated the preponderance of light responsive motifs (Kawoosa *et al*., 2010) and dehydration responsive
motifs. Therefore PropkDXS was analyzed to understand the molecular mechanism of PkDXS regulation. Targeted light regulated elements included GATA, and SORLIP; whereas motif involved in abiotic stress was WRKY and a few putative novel motifs were studied for their protein-binding capabilities.

5.5.1 DNA–protein interaction analysis of selected motifs in P. kurrooa by EMSA

GATA motif, in general, is responsible for increased promoter activity under light as being the binding site of light regulatory GATA factors (Kehoe et al., 1994; Terzaghi and Cashmore 1995; Teakle et al., 2002). The role of GATA factors in mediating the response of plants under light has been studied by In vitro EMSA and DNase I footprinting experiments. The preponderance of GATA motifs in promoters of genes modulated by light suggested the role for plant GATA factors in mediating light regulated gene expression (Lam and Chua, 1989; Schindler and Cashmore, 1990; Borello et al., 1993). Other GATA-related motifs such as I boxes (GATAA) were present in many light-regulated genes such as those encoding RBCS, chlorophyll a/b binding protein and GAP (Castrresana et al., 1987; Giuliano et al., 1988; Gilmartin et al., 1990; Jeong and Shih, 2003). Mutation and deletion of some of the GATA elements strongly reduces promoter activity. Data on P. kurrooa showed stronger protein binding of nuclear extract of light-exposed plants with GATA (Fig. 4.72A lane 2). Binding was not observed with mutated GATA box (mGATA; TCGT). A positive correlation between gene expression and appearance of fast-migrating complex under light suggested the importance of GATA in regulating light-mediated gene expression of PkDXS in P. kurrooa.

Another motif characterized in the present work was SORLIP that was originally identified in phytochrome A induced promoters of Arabidopsis (Hudson and Quail, 2003). Also, the motif was reported to be over-represented in the promoters of flavonoid pathway genes (Pan et al., 2009). SORLIP has also been identified in promoters of (a) circadian phenomenon-related genes like cytochrome P450 monooxygenases (Pan et al., 2009), (b) LHY gene (Spensley et al., 2009) in Arabidopsis and (c) photosynthesis-related genes (Bilgin et al., 2010). However, SORLIP motif from P. kurrooa has been shown to be regulated by dark and showed DNA-protein interaction with nuclear extract of dark exposed plants (Fig. 4.72B).

Yet another motif evaluated in the present thesis was for WRKY TFs. WRKY family TFs mediates plant responses to different environmental cues. Plant response to drought and salinity is mediated by several WRKY proteins (Golldack et al., 2014). Increased expression of OsWRKY11 downstream of HSP101 promoter led to increased drought tolerance in transgenic rice plants as indicated by the slower leaf-wilting and increased survival rate of
green plant parts (Wu et al., 2009). Osmotic stress tolerance of transgenic *Arabidopsis* plants was improved by up-regulation of *OsWRKY08*, in response to PEG, NaCl or ABA treatment. *OsWRKY08* positive regulated expression of two ABA-independent abiotic stress responsive genes, *AtCOR47* and *AtRD21* (Song et al., 2009). ABA plays an essential role in plant responses to drought stresses. Previous research had showed that WRKY proteins functioned as activators or repressors in ABA signaling and play important role in ABA signaling. Activator WRKY protein included *OsWRKY72* and *OsWRKY77*, whereas *OsWRKY24* and *OsWRKY45* acted as repressors to same ABA inducible promoter.

A strong DNA-protein interaction was observed when nuclear extract of leaf of drought stressed *P. kurrooa* was incubated with WRKY motif (Fig. 4.73A). Though, weak binding was also observed with nuclear extract of leaf of well-irrigated plants (Fig. 4.73A). These results were in accordance to expression of *PksWRKY* and *PkdWRKY* in leaf tissue of drought stressed *P. kurrooa* plants. Data on preponderance of WRKY motif in promoter of various genes of picrosides biosynthesis pathway, capability of *WRKY* genes to produce functional protein that binds to WRKY domain and capability of WRKY boxes to exhibit DNA-protein binding suggests *WRKY* to be a target gene for modulating picrosides biosynthetic pathway.

Also, three novel motifs were selected to evaluate their capability to bind to nuclear extract of leaves of drought stressed *P. kurrooa* plants. Nuclear protein from IC and IW plants were used for functional characterization of various motifs. The capacity of binding of protein to the boxes suggested that target box could be functional one. However, absence of binding could be due to low concentration or absence of protein in the nuclear extract. TF are specific to the motif and are induced in response to particular developmental changes or environmental cues. Out of the three motifs, only one novel motif exhibited DNA-protein interaction. This motif needs detailed analysis and could provide insight into regulation of isoprenoid biosynthesis.

### 5.6 Functional evaluation of *PropkDXS* in *Arabidopsis*

#### 5.6.1 *PropkDXS* drives developmental and tissue specific expression in *Arabidopsis*

The present thesis established *PkDXS* to be a functional gene (Section 4.3). Also, analysis of six motifs in the promoter region of *PkDXS* showed protein binding ability with four motifs. Capability of these motifs for protein binding varied with the exposure of plant to light and drought. Therefore, 1,100 bases upstream of *PkDXS* were evaluated for its functionality. Since the transgenic system of *P. kurrooa* did not exist, *PropkDXS* was evaluated in
Arabidopsis using GUS as a reporter gene. During early developmental stages (up to 5 days of germination), PropkDXS driven GUS activity was undetectable in any tissue. GUS expression was visible in seven day old seedlings but was restricted to tips of leaves and in mature leaves, GUS expression was observed in leaf stalks, emerging apical leaves, and apical part of the leaves. Previous work also showed that DXS promoter from Arabidopsis drives high GUS expression in young developing leaves as compared to mature leaves (Lorenzo et al., 2002).

5.6.2 PropkDXS up-regulated gene expression in response to light in Arabidopsis

Previous research in our lab showed higher accumulation of monoterpenoid picrosides in P. kurrooa under continuous light with concomitant up-regulation of PkDXS as compared to those under dark (Kawoosa et al., 2010). In silico analysis of PropkDXS indicated the preponderance of several light responsive motifs throughout the promoter (Kawoosa et al., 2010) and these include: GATA box (WGATAR; Gilmartin et al., 1990), GT-1 (GRWAAW; Villain et al., 1996), and I box (GATAA; Terzaghi and Cashmore, 1995). Other promoter specific light responsive motifs are CGGATA (REbeta) box and GGTTAA motif. CGGATA box is required for phytochrome regulation in Lemna gibba Lhcb21 gene promoter (Degenhardt) and GGTTAA motif, are present in upstream sequences of the light-responsive pea rbcS-3A gene (Green et al., 1988).

Light is an environmental cue that has the most profound effect on biosynthesis of a range of terpenoids through regulation of genes of the associated biosynthetic pathways. Light positively regulates the MEP pathway genes. DXS is the regulatory enzyme of MEP pathway and its expression is modulated by light. The accumulation of transcript of DXS has been observed in photosynthetic tissues of maize, medicago, tomato and tobacco. Knockout mutations in DXS gene resulted in lethal albino plants. These albino plants have been unable to synthesize photosynthetic pigments and had impaired chloroplast activity (Mandel et al., 1996). Expression analyses in Arabidopsis showed accumulation of transcript of DXS upon exposure to light in seedlings during the development of the first true leaves (Carretero-Paulet et al., 2002; Guevara-Garcı’a et al., 2005; Hsieh et al., 2008). Positive regulation of light and accumulation of transcript of DXS by light provides an advantage during early seedling development. There is an elevated demand of photosynthetic pigments during early seedling development.

The activity of PropkDXS was evaluated in response to light and dark in Arabidopsis. The activity of PropkDXS was recorded to be higher in light as compared to that in dark at 12 and 24 h of light exposure. Our previous work showed up-regulation of PkDXS in light
(Kawoosa et al., 2010). Also PropkDXS had the required motif responsible for light regulation of gene expression which showed protein binding capability (section 4.7.1).

The decrease in GUS expression at 48 and 72 h of exposure might be due to feedback inhibition of DXS activity by the metabolites of pathway as a protective mechanism of plants against continuous exposure of light; since MEP pathway is responsible for synthesis of metabolites involved in pigment-protein complexes such as chlorophylls, carotenoids, and their oxygenated xanthophylls in chloroplasts (Pulido et al., 2012).

5.6.3 PropkDXS down-regulated gene expression under drought in Arabidopsis

In silico analysis of PropkDXS indicated the presence of several drought responsive motifs (Table 4.5) present in dehydration responsive genes. These motifs were ACGTATERD1 (ACGT), MYB1AT (WAACCA), MYBCORE (CNGTTR), MYBST1 (GGATA), and MYCCONSENSUSAT (CANNTG). Other motifs present in PropkDXS and involved in abiotic stresses response were WBBOXPCWRKY1 (TTTGACY), WBOXATNPR1 (TTGAC), and WRKY71OS (TGAC). The present thesis evidently showed binding of nuclear extracts of drought stressed plants to WRKY motifs (section 4.7).

Drought adversely affects growth, metabolism and yield in higher plants. Plants produce a number of secondary metabolites for ameliorating abiotic stress, one group of such compounds are isoprenoids. DXS is the regulatory gene of the MEP pathway. The expression of PkDXS in drought stressed P. kurrooa was studied and PkDXS was found to be down-regulated in water stressed leaves of P. kurrooa.

The activity of PropkDXS was evaluated in response to PEG induced drought stress in Arabidopsis. PEG lowers the water potential of nutrient medium by getting diffused in the nutrient medium (Van der Weele et al., 2000). The expression of PropkDXS was higher at 12 h of osmotic stress as compared to control plants. Down-regulation in GUS expression was observed at 48 and 72 h of treatment as compared to that of control. Decrease in GUS activity at later stages of osmotic stress was in accordance with expression of PkDXS in P. kurrooa plants subjected to drought stress. Thus, experiments of PropkDXS in Arabidopsis suggested PropkDXS to be a functional promoter.

To conclude, the present work analyzed transcriptome of P. kurrooa and provided insight into molecular as well as metabolic responses of the species to temperature and drought stresses. Molecular regulation of picrosides metabolism through promoter analysis was also evident. A temperature of 25 °C appeared to impose stress in the species. Drought had differential response in term of picrosides accumulation for leaf and rhizome tissues. Leads obtained in the present thesis can be perused further as follows.
5.7 Future line of work

i. Transcriptome sequencing generated gene resource for plant which can be used to understand molecular response of plant to various cues.

ii. Developed gene resources including TFs can be used to generate transgenic plants for modulating picrosides content and stress tolerance.

iii. Functionality of WRKY TFs could further be studied for transcriptional activation/repressor activity.

iv. Putative MTS cloned in the present work could be evaluated for their functionality for characterization of unknown steps in cyclization of GPP to iridoid moiety.

v. CYPs and GTs identified in the present study could serve as a source of hitherto unknown genes associated with picrosides biosynthesis.

vi. It would be interesting to understand the role of picrosides in plant biology, if any.

vii. Identification of putative cis-acting elements motifs could help in understanding regulation of picrosides biosynthesis and the plant response to environmental cues.

viii. Gene expression is affected by both-transcription and translation modification. miRNA are known to down-regulate target genes at the post-transcriptional level thus regulating the expression of key genes and TFs. Stress memories can be stable and carried forward as epigenetic changes thereby assisting plants to cope effectively with subsequent stresses. Both miRNAs and epigenetic systems provide opportunities to study the response of plants in response to various environmental cues.