Chapter-2
REVIEW OF LITERATURE

Himalayas are rich in biodiversity, so this region has abundant plant species which are valuable for their unique natural products (economical property) and their medicinal with insecticidal property (Tewary et al., 2005). The geological distribution of Cedrus in Himalayan region and the information related to the growing, managing, harvesting, wood formation and utilization has been explained by Tiwari (1994).

Singh et al. (1984), Singh and Rao (1985), Singh and Agrawal (1988) reported that Himalayan Cedar (Cedrus deodara) belongs to family Pinaceae and found throughout the western Himalayas at altitude of 1200-3000m, in loam soil with high content of organic carbon and potash and low content of phosphorus. It contains the essential oils that possess the defence mechanism against the pests and houseflies. The eluates resulted in nine chromatographic fractions on the silica gel. Almost all fractions showed insecticidal activity against both test species. Fractions I and V led to the highest mortality and also produced a quick knockdown effect. Fractions I and V, after rechromatography and purification, yielded himachalol (3%) and β-himachalene (31%), based on essential oil weight, respectively. These sesquiterpenes indicated 97.5% mortality in 0.56 μmol/ insect against pulse beetle. These biologically active natural products of plant origin may serve as suitable prototypes for development of commercial insecticides.

2.1 Wood Formation and Traits

Wood fibres are the natural composite structures in the cellulose fibrils that were together by lignin and hemicelluloses. The major
constituents of wood fibres are lignin, cellulose, hemicelluloses and extractives. Each of these components contributes to fibre properties, which ultimately impact product properties.

Cote et al. (1966) studied the lignin composition and distribution between the normal and compression wood of red spruce. It was reported that 75% lignin located in middle lamella of red spruce. The results obtained with compression wood tracheids showed that there was an isotropic layer with a high lignin concentration located inside S1. This layer contained polysaccharides which were probably oriented in a transverse direction with respect to the fiber axis. The inner part of S2, containing the helical cavities, had a uniform distribution of lignin in the form of a continuous network. There was no evidence for a radial distribution of lignin in compression wood tracheids and presence of lignin from the inner portion of S2. The polysaccharides in the cell wall of the compression tracheids, and especially those located in the outer part of S2, were partially inaccessible to both acids and fungi.

Parham and Cote (1971) studied the lignin distribution in normal and reaction wood of loblolly pine through the lignin skeletonising. The carbohydrates of wood hydrolyzed through the hydrofluoric acid then it left the uniform distributed lignin in cell wall of normal wood tracheids. The total amount of softwood conifer lignin occurred in middle lamella and most of it were found in the secondary cell wall of coniferous tracheids. The reaction wood of loblolly pine contains more lignin than normal wood. They concluded that lignin content is 7% more in compression wood rather than normal wood of pine. Ray cells of normal wood tracheids are also ligninfied to the same extent.

Sarkanen and Ludwig (1971) reported that wood is composed of many primarily extractives, carbohydrates and the lignin which
are of hydrophobic nature. Lignins are treated as the obstacle for water absorption to the cell wall. It confers impermeability to tracheary elements and therefore allows the transport of water and solutes through the vascular system. Lignin plays the intricate role in the transportation of water and nutrients. Lignin has the function in a plant’s natural defence mechanism against degradation by impeding penetration of destructive enzymes through the cell wall.

Westermark et al. (1988) determined the lignin distribution between the middle lamella and the cell wall of spruce fibers through the mecurization technique of the lignin and a concomitant determination of mercury by the SEM-EDXA technique. The ratio of lignin in the middle lamella at the cell corners to the lignin in the secondary wall was 2.5±0.6 for latewood and 2.4±0.6 for earlywood. This gives a lignin content of 55–58% in the true middle lamella in the cell corners. The reactivity to mercuric acetate of different wood elements was determined in separate experiments. Fractions enriched in ray cells, middle lamella and compression wood all reacted at the same rate as the whole wood; about one mole of mercury was incorporated per mole of lignin (C9-unit).

Taylor et al. (1999) studied the irregular xylem3 (irx3) mutant of Arabidopsis which serves the deficiency in secondary cell wall cellulose deposition and leads to collapsed with xylem cells. The irx3 gene has been mapped to the top arm of chromosome V with marker nga106. Expressed sequence tag clone 75G11, which exhibits sequence similarity to cellulose synthase, was found to be tightly linked to irx3, and genomic clones containing the gene corresponding to clone 75G11 complemented the irx3 mutation. Thus, the irx3 gene encodes a cellulose synthase component that is specifically required for the synthesis of cellulose in the secondary cell wall. The irx3 mutant allele contains a stop codon that truncates the gene product by 168 amino acids, suggesting that this allele is null. Furthermore,
in contrast to radial swelling1 (rsw1) plants, irx3 plants show no increase in the accumulation of β-1,4-linked glucose in the noncrystalline cell wall fraction. irx3 and rsw1 fall into a distinct subgroup (Csa) of Arabidopsis genes showing homology to bacterial cellulose synthases.

Muller et al. (2005) reported the enzymatic and topochemical aspects during the lignification in Pinus radiata. Lignification analysis has been done through the UV-spectrophotometery and thioacidolysis. UV spectra absorption has been noticed from the secondary cell wall of tracheids, sclereids and from the compound of middle lamella. Thioacidolysis confirmed the presence of lignin in Pinus radiata.

2.2 Cellulose Biosynthesis

Wu et al. (2000) reported that angiosperms accumulate the large amount of cellulose with the low quantity of lignin in cell walls of tension-stressed tissues Molecular analysis of tree stress response has been done through the cloning procedure of cDNA from developing xylem of aspen and got the full length of Cellulose synthase (PtCesA). PtCesA contains the 90% similarity with the cellulose of cotton (GhCesA). PtCesA gene transcript analysis in various tissues of aspen has been gone through the northern blotting and in situ hybridization. The promoter (GUS) of PtCesA has been fused with the transgenic tomato to demonstrate the expression of PtCesA to developing xylem during normal plant growth. The stem bending induced the mechanical stress. During the stress conditions, GUS expression was remained in xylem and induced in developing phloem fibres. The effect of stress in tissues undergoing the compression was zero in the opposite of bending. The result showed that PtCesA gene has a vital role in cellulose biosynthesis during the tension stressed condition and normal tissues of aspen.
PtCesA gene expression triggers the signalling mechanism during the stress related compensatory deposition of cellulose and lignin.

Endler and Perrson (2011) reported that plant cell walls are complex structures composed of high-molecular-weight polysaccharides, proteins, and lignins. Among the wall polysaccharides, cellulose, a hydrogen-bonded β-1,4-linked glucan microfibril, is the main load-bearing wall component and a key precursor for industrial applications. Cellulose is synthesized by large multi-meric cellulose synthase (CesA) complexes, tracking along cortical microtubules at the plasma membrane. The cellulose synthase proteins are the only known protein of these complexes. In this research migratory pattern of CesAs has been identified through the status of phosphorylation.

Mendu et al. (2011) reported that plant cells have a rigid cell wall that constrains internal turgor pressure which extends in a regulated and organized manner to allow the cell to acquire shape. The cellulose is the primary load-bearing macromolecule of a plant cell wall which forms crystalline microfibrils that are organized with respect to a cell’s function and shape requirements. A primary cell wall is deposited during expansion whereas secondary cell wall is synthesized post expansion during differentiation. A complex form of asymmetrical cellular differentiation occurs in Arabidopsis seed coat epidermal cells, where we have recently shown that two secondary cell wall processes occurs that utilize different cellulose synthase (CesA) proteins. One process is to produce pectinaceous mucilage that expands upon hydration and the other is a radial wall thickening that reinforced the epidermal cell structure. CesA5 facilitates the mucilage attachment to the parent seed and CesA2, CesA5 and CesA9 in radial cell wall thickening and formation of the columella. Here, a model for the complexity of cellulose biosynthesis
in this highly differentiated cell type with further evidence supporting each cellulosic secondary cell wall process.

Lei et al. (2012) reported that cellulose is synthesized at the plasma membrane by protein complexes known as cellulose synthase complexes (CSCs). The cellulose-microtubule alignment hypothesis states that there is a causal link between the orientation of cortical microtubules and orientation of nascent cellulose micro fibrils. CeSA interactive protein 1 (CSI1) interacts with CSCs and potentially links to the cytoskeleton. CSI1 co-localizes with CSCs and also travels bi-directionally in a speed indistinguishable from CSCs. The linear trajectories of CSI1-RFP coincide with the underlying microtubules labelled by YFP-TUA5. In the absence of CSI1, both the distribution and the motility of CSCs are defective and the alignment of CSCs and microtubules is disrupted. These observations led to the hypothesis that CSI1 directly mediates the interaction between CSCs and microtubules. In support of this hypothesis, CSI1 binds to microtubules directly by an in vitro microtubule-binding assay. In addition to a role in serving as a messenger from microtubule to CSCs, CSI1 labels SmaCCs/MASCs, a compartment that has been proposed to be involved in CesA trafficking and/or delivery to the plasma membrane.

Liu et al. (2012) identified the four novel cellulose synthase genes from birch that involved in primary and secondary cell wall biosynthesis. Cellulose synthase (CesA), which is an essential catalyst as the cell wall biomass and encoded by the CesA gene family that contains ten or more members. In this study four full-length cDNAs encoding CesA were isolated from Betula platyphyllo Suk.(timber species) through RT-PCR and followed by RACE method. The isolated genes were named as BplCesA3, BplCesA4, BplCesA7 and BplCesA8. BplCesA are homologous with Arabidopsis and
contains the same typical domains and regions. The genes are different in length as did the locations of the various protein domains inferred from the deduced amino acid sequences, which shared amino acid sequence identities ranging from only 63.8% to 70.5%. Real-time RT-PCR showed that all four BplCesAs were expressed at different levels in diverse tissues. Results indicated that BplCESA8 might be involved in secondary cell wall biosynthesis and floral development. BplCESA3 have the unique expression pattern and involved in primary cell wall biosynthesis and seed development, BplCesA3 might be participate homogalacturonan synthesis. BplCESA7 and BplCESA4 play a role in cellulose synthase complex and participate mainly in secondary cell wall biosynthesis. BplCESAs have extremely low expressions in mature pollen. BplCesA have very little involvement in mature pollen formation of Betula. BplCeSAs may be participated in the various tissue developments due to different expression level.

**Lignin Biosynthesis**

Freudenberg and Neish (1968) demonstrated that lignin is cross-linked racemic macromolecule and it derived through the polymerization of three different hydroxycinnamyl alcohols (monolignols): p-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol. Gymnosperm lignin contains the coniferyl alcohol as the major component. Coniferyl alcohol and sinnapyl alcohol play an important role in composition of lignins in broad-leaved trees. Compression wood and grass lignin are abundant in p-hydroxyphenyl units.

Eriksson *et al.* (1980) treated the lignin-carbohydrate material from black spruce (*Picea mariana*) with hemicellulase mixture. The carbohydrates in the remaining material were studied through the treatment of sodium borohydride followed by hydrolysis of furanosidic bonds and through the Smith degradation method. The
results obtained indicate that lignin is bound to all types of sugar units in the hemicelluloses. For arabino-4-O-methylglucuronoxytan, linkages to the side-chain units seem to dominate. The links to the 4-O-methylglucuronic acid units are ester bonds, probably benzyl ester bonds. Those to the other sugar units in the hemicelluloses are ether bonds, probably benzyl ether bonds. It showed that fractions of cellulose are linked to lignin.

Goldberg et al. (1983) reported that peroxidases are responsible for the final oxidation of cinnamyl alcohols to form lignin. In poplar, two anionic peroxidases have been preferentially expressed in developing xylem affecting oxidation on syringaldazine. Cell wall peroxidases from poplar stem were investigated through parallel biochemical and histochemical techniques. Oxidation of syringaldazine was obtained only in lignifying cells. It has been demonstrated through the peroxidase activity with the exogeneous H₂O₂ for in vitro assays while syringaldazine oxidation could be obtained in situ in the absence of exogeneous hydrogen peroxide probably because of the production of H₂O₂ by the lignifying cell walls itself. Syringaldazine oxidase activity was strongly bound to the cell walls and fairly resistant to heat inactivation.

Lewis and Yamamoto (1990) reported that lignin is a complex polymer of high carbon content distinct from carbohydrates and layered the plant cell wall. This phenolic biopolymer has no extended sequences of repeating units and it is characterized by a set of variable cross-linkages. Lignins are composed of three main units named p-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units. These components originate from the polymerization of the three monolignols, the p-coumaryl, coniferyl, and sinapyl alcohols, respectively. The monolignols are synthesized through various successive steps such as deamination, reduction, hydroxylation, and methylation. Gymnosperms lignin is composed of G unit where as
the angiosperms lignin composed of G and S unit while monocotyledons have more H unit. He reported the amount of lignin in Norway spruce. Norway spruce contains 98:2 proportions of guaiacyl unit and p-hydroxyphenyl unit.

Li et al. (2001) showed the involvement of cinnamyl alcohol dehydrogenase (CAD) in the reduction of cinnamaldehydes into cinnamylalcohols, which explained as the last step of the monolignol biosynthesis pathway. During lignin biosynthesis in angiosperms, coniferyl and sinapyl aldehydes are believed to be converted into their corresponding alcohols by CAD and by sinapyl alcohol dehydrogenase (SAD), respectively. An aspen gene was discovered that encodes a SAD enzyme which is specifically involved in the reduction of syringyl monolignol through in vitro technique. SAD gene is phylogenetically distinct from aspen CAD (PtCAD). Liquid chromatography–mass spectrometry-based enzyme functional analysis and substrate level–controlled enzyme kinetics consistently demonstrated that PtSAD is sinapaldehyde specific and that PtCAD is coniferaldehyde specific. The enzymatic efficiency of PtSAD for sinapaldehyde was approx 60 times greater than that of PtCAD. SAD function is essential to the biosynthesis of syringyl monolignol in angiosperms. In aspen stem primary tissues, PtCAD was immunolocalized exclusively to xylem elements in which only guaiacyl lignin was deposited, whereas PtSAD was abundant in syringyl lignin–enriched phloem fiber cells. In the developing secondary stem xylem, PtCAD was most conspicuous in guaiacyl lignin–enriched vessels, but PtSAD was nearly absent from these elements and was conspicuous in fiber cells.

Dixon et al. (2002) reported that phenylpropanoid compounds play a role in plant defence mechanism from preformed or inducible physical and chemical barriers against the infection to signal
molecules which involves in systematic signaling for defence gene induction. Defensive mechanism has been noticed in hydroxycinnamic acids and other monolignols such as flavonoids, isoflavonoids and stilbenes. Phenylpropanoid biosynthesis and its enzymatic steps are now well known with many corresponding genes. Many of the biosynthetic pathway enzymes are encoded by gene families. The complete genome sequences of *Arabidopsis thaliana*, and expressed sequence tag (EST) resources in other species such as rice, soybean and tomato allow for the comparative genetic complexity of the phenylpropanoid pathway across species. Gene expression array analysis and metabolic profiling approaches are useful to find the comparative parallel analyses of global variation at the genome and metabolome levels.

Boerjan *et al.* (2003) found that over the past decade lignin biosynthetic pathways have been studied. The genetic, combined bioinformatics and biochemistry approaches have been made the significant progress in cloning new genes. In vitro enzymatic assays and detailed analyses of mutants and transgenic plants altered in the expression of lignin biosynthesis genes have provided a solid basis for redrawing the monolignol biosynthetic pathway, and structural analyses have shown that plant cell walls can tolerate large variations in lignin content and structure. In some agriculture transgenic with modified lignin structure has been demonstrated and contains the potential.

Goujan *et al.* (2003) reported Hydroxylation, methylation step have been carried out to form complex polymer of the monolignols precursor from phenylalanine This complex polymer lignin composed by the hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units of the polymer. The set of genes potentially involved in the monolignol pathway in the model plant, *Arabidopsis thaliana*. Genes encoding
enzymes implicated in constitutive lignin synthesis were identified on
the basis of their homology to monolignol biosynthesis genes of other
plants and their high expression in lignified tissues (floral stems,
roots). This research shows that most of the genes belong to
multigene families while few genes such as PAL, 4CL, CAD are
duplicated in it. The genes encoding the cytochrome P450 mono-
oxigenases (C4H, C3H, F5H) are unique except for F5H because one
homologue gene present in the complete genome. The deregulation of
monolignols pathway in mutants and transgenic Arabidopsis lines
has been reported with the impact of targeted gene, structure.

Ralph et al. (2004) Lignins are complex natural polymers
resulting from oxidative coupling of, primarily, 4-
hydroxyphenylpropanoids. An understanding of their nature is
evolving as a result of detailed structural studies, recently aided by
the availability of lignin-biosynthetic-pathway mutants and
transgenics. The currently accepted theory is that the lignin polymer
is formed by combinatorial-like phenolic coupling reactions, via
radicals generated by peroxidase-H2O2, under simple chemical
control where monolignols react endwise with the growing polymer.
As a result, the actual structure of the lignin macromolecule is not
absolutely defined or determined. The randomness of linkage
generation (which is not truly statistically random but governed, as
is any chemical reaction, by the supply of reactants, the matrix, etc.)
and the astronomical number of possible isomers of even a simple
polymer structure, suggest a low probability of two lignin
macromolecules being identical. A recent challenge to the currently
accepted theory of chemically controlled lignification, attempting to
bring lignin into line with more organized biopolymers such as
proteins, is logically inconsistent with the most basic details of lignin
structure. Lignins may derive in part from monomers and conjugates
other than the three primary monolignols (p-coumaryl, coniferyl, and
sinapyl alcohols). The plasticity of the combinatorial polymerization
reactions allows monomer substitution and significant variations in final structure which, in many cases, the plant appears to tolerate. As such, lignification is seen as a marvelously evolved process allowing plants considerable flexibility in dealing with various environmental stresses, and conferring on them a striking ability to remain viable even when humans or nature alter required lignin-biosynthetic-pathway genes/enzymes. The malleability offers significant opportunities to engineer the structures of lignins beyond the limits explored to date.

Sibout et al. (2005) reported that coniferyl and sinapyl aldehydes are to be converted into its corresponding alcohols by cinnamyl alcohol dehydrogenase (CAD) and by sinapyl alcohol dehydrogenase (SAD), respectively in angiosperms lignin biosynthesis. CAD-C and CAD-D are the initial genes involved in lignin biosynthesis in the floral stem of Arabidopsis thaliana. Floral stem of Arabidopsis thaliana has been made through supply of both coniferyl and sinapyl alcohols. The double mutants of CAD (cad-c cad-d) showed the limp floral stem at maturity due to the phenotypic change and cause the changes in the pattern of lignin staining. He reported that lignin content of the mutant stem was reduced by 40%, related to the wild type which contains 94%. The lignin mutation affects the xylem vessels and fibres which have been demonstrated by Fourier transform infrared spectroscopy. RT-PCR analysis reveals the transcription of CAD homologs. In tissue specific expression of CAD suggests the different abilities of these genes/proteins to produce syringyl-lignin moieties but there is no requirement of SAD gene.

Bomal et al. (2008) evaluated that the involvement of two R2R3-MYB genes from Pinus taeda L., PtMYB1 and PtMYB8, in phenylpropanoid metabolism and secondary cell wall biogenesis.
These pine MYBs were constitutively over-expressed in *Picea glauca* (Moench) Voss, used to study the heterologous conifer expression system. Morphological, histological, chemical (lignin and soluble phenols), and transcriptional analyses such as microarray and reverse transcription quantitative PCR (RT-qPCR) were used for extensive phenotyping of MYB-over-expressing plantlets of spruce. The lignin deposition was also a common feature but *PtMYB8-OE* gene strongly associated with in secondary cell wall deposition. Microarray and RT-qPCR results expressed the up-regulation of many genes encoding phenylpropanoid enzymes involved in lignin monomer synthesis.

Wagner *et al.* (2009) explained the suppression of 4-coumarate-coenzyme A ligase (*4CL*) in the coniferous gymnosperm *Pinus radiata* that affected plant phenotype and resulted in dwarfed plants with a “bonsai tree-like” appearance. Microscopic analyses of stem sections revealed wood and bark tissues changes in 2 year old plant. In 2yr old plant weak lignified tracheids have the signs of collapse and the development of circumferential bands of axial parenchyma. Acetyl bromide-soluble lignin assays and proton nuclear magnetic resonance studies revealed the reduction of lignin upto 50% in the most severely affected transgenic plants. Lignin reduction has been studied through the 2-D nuclear magnetic resonance and pyrolysis-gas chromatography-mass spectrometry which indicates that reductions were mainly due to depletion of guaiacyl but not *p*-hydroxyphenyl. *4CL* silencing also caused modifications in the lignin inter unit linkage distribution, including elevated β-aryl ether (β-0-4 unit) and spirodienone (β-1) levels, accompanied by lower phenylcoumaran (β-5), resinol (β-β), and dibenzodioxocin (5-5/β-O-4) levels. A sharp depletion in the level of saturated (dihydroconiferyl alcohol) end groups was also observed. Severe suppression of *4CL* also affected carbohydrate metabolism. It
causes the 2-fold increment in galactose content in wood from transgenic plants due to increased compression wood formation. The molecular, anatomical, and analytical data verified that the isolated 4CL clone is associated with lignin biosynthesis and illustrated that 4CL silencing leads to complex, physiological and morphological changes in *P. radiata*.

Barakat *et al.* (2010) analyzed the expression of the phylogeny of CAD and CAD-like genes in lycophyte and bryophyte sequences. Many CAD/CAD-like genes do not seem to be associated with wood development under normal growth conditions. The expression of CAD/CAD-like genes in Populus plant tissues has been analyzed through the response against the feeding damage by gypsy moth larvae (*Lymantria dispar L.*) to give the emphasis on the functional evolution. Expression of CAD/CAD-like genes in Populus tissues (xylem, leaves, and barks) of treated (herbivore) and non-treated plants was analyzed by using the qRT-PCR.

Hu *et al.* (2010) reported that phenylpropanoid pathway has the key enzyme 4-Coumaric acid:CoA ligase (4CL). It catalyzes the synthesis of hydroxycinnamate-CoA thioester i.e. precursors of lignin and other important phenylpropanoids, the catalysis covers the two-step reactions involving the formation of hydroxycinnamate-AMP anhydride and followed by nucleophilic substitution of AMP by CoA. In this study, the crystal structures of *Populus tomentosa* 4CL1 in the unmodified (apo) form and in forms complexed with AMP and adenosine 5beta-(3-(4-hydroxyphenyl)propyl)phosphate (APP), an intermediate analog, at 2.4, 2.5, and 1.9 Å resolution, respectively. 4CL1 consists of two globular domains connected by a flexible linker region. The larger N-domain contains a substrate binding pocket, while the C-domain contains catalytic residues. Upon binding of APP, the C-domain rotates 81° relative to the N-domain. The crystal
structure of 4CL1-APP reveals its substrate binding pocket. We identified residues essential for catalytic activities (Lys-438, Gln-443, and Lys-523) and substrate binding (Tyr-236, Gly-306, Gly-331, Pro-337, and Val-338) based on their crystal structures and by means of mutagenesis and enzymatic activity studies. The size of the binding pocket is the most important factor in determining the substrate specificities of 4CL1 protein to demonstrate the biochemical activity.

Li et al. (2010) demonstrated that conifers have highly conserved xylem transcriptome region but it diverse with angiosperms. The functional domains of genes in the xylem transcriptome are moderately to highly conserve in vascular plants. It showed the existence of a common ancestral xylem transcriptome. The whole transcriptome analysis from the range of tissues derived that the xylem transcriptome is relatively conserved in vascular plants the conserve region such as xylem transcriptome, cell wall genes, ancestral xylem genes, known proteins and transcription factors are relatively more conserved in vascular plants. The putative orthologs (527) are unevenly distributed across the Arabidopsis chromosomes with eight hot spots observed. Phylogenetic analysis revealed that evolution of the xylem transcriptome has paralleled plant evolution. He identified that 274 conifer-specific xylem Unigenes functions are unknown. These xylem orthologs and conifer-specific unigenes are likely to have played a crucial role in xylem evolution. The xylem transcriptomes of vascular plants are more conserved than the total transcriptomes. Evolution of the xylem transcriptome has largely followed the trend of plant evolution.

Morita et al. (2011) crystallized the 4-Coumarate-ligase –CO A ligase (4CL2) in Arabidopsis thaliana. In this research he has done the preliminary X-ray analysis of the coumarate ligase gene. 4CL2 in Arabidopsis thaliana catalyzes the ATPdependent formation of the 4-coumaroyl-CoA thioester through the formation of 4-coumarate-AMP.
Recombinant 4CL2 protein was expressed in Escherichia coli and crystallized through the sitting-drop vapour-diffusion method. The crystals belonged to space group P21, with unit-cell parameters $a = 91.6$, $b = 55.5$, $c = 124.4 \, \text{Å}$, $\alpha = \gamma = 90.0$, $\beta = 111.1^\circ$.

Naioki et al. (2011) focused on four type of genes such as 4-coumarate-CoA ligase (4CL), LIM domain transcription factor (LIM), coniferaldehyde 5-hydroxylase (CAld5H) and the three catalytic units of cellulose synthase (CesA1, CesA2 and CesA3) influencing wood quality. There is correlation between the relative expression levels of genes related to wood quality. The genes encoding LIM, 4CL, CAld5H and the cellulose synthase (CesA1, CesA2 and CesA3) from *E. globulus* has been cloned by the method of cDNA library which was developed through purified mRNA from the stems of 4 month old eucalyptus. The expression levels of LIM in basal stems of ten independent *E. globulus* lines showed similar patterns to those of 4CL, indicating that the LIM may control 4CL expression. The correlation has been occurred between gene expression levels and wood qualities such as Klason lignin (KL) content, syringyl/guaiacyl (S/G) ratio and holocellulose (HC) content. Expression of the LIM and 4CL were positively correlated with KL content. A highly significant positive correlation was observed between CAld5H expression and S/G ratio. Furthermore, a ratio of the sum of the expression levels of three CesA1, CesA2 and CesA3 to 4CL showed positive correlation with a ratio of HC/KL content that positively correlated to the chemically extracted fiber content in this woody plant. As the resultant candidate genes such as LIM, 4CL, CAld5H and CesA are responsible for the desirable wood qualities and biomass.

Bukh et al. (2012) studied the phylogeny and structure of CAD in *Brachypodium distachyon*. Cinnamyl alcohol dehydrogenase (CAD) catalyses the final step of the monolignol biosynthesis, the conversion of cinnamyl aldehydes to alcohols, using NADPH as a
cofactor. Seven members of the CAD gene family were identified in the genome of Brachypodium distachyon and five of these were isolated and cloned from genomic DNA. BdCAD5 has been expressed in all tissues and highest in root and stem while BdCAD3 was only expressed in stem and spikes. A phylogenetic analysis of CAD-like proteins placed BdCAD5 on the same branch as bona fide CAD proteins from maize (ZmCAD2), rice (OsCAD2), sorghum (SbCAD2) and Arabidopsis (AtCAD4, 5). The predicted three-dimensional structures of both BdCAD3 and BdCAD5 have the resemblance with AtCAD5. The domain active site of BdCAD3 and BdCAD5 are distributed symmetrically and BdCAD3 is similar to the poplar sinapyl alcohol dehydrogenase (PotSAD). BdCAD3 and BdCAD5 expressed and purified from Escherichia coli both showed a temperature optimum of about 50 °C and molar weight of 49 kDa. The optimal pH for the reduction analysis of coniferyl aldehyde was pH 5.2 and 6.2 and the pH for the oxidation analysis of coniferyl alcohol was pH 8 and 9.5, for BdCAD3 and BdCAD5 respectively. Kinetic parameters for conversion of coniferyl aldehyde and coniferyl alcohol showed that BdCAD5 was clearly the most efficient enzyme in its gene family. The results showed that BdCAD5 is the main CAD enzyme for lignin biosynthesis while BdCAD3 has a different role in Brachypodium. All CAD enzymes are cytosolic except for BdCAD4, which has a putative chloroplast signal peptide adding to the diversity of CAD functions.

Huang et al. (2012) sequenced the transcriptomes of nine tissues from Chinese fir and analyzed by using the Illumina HiSeq™ 2000 sequencing platform. Ilumina HiSeq generated the ~40 million pair ends to generate the 3.62 Gbps of sequencing data. These reads were assembled into 83,248 contigs with an average length of 449 bp, (37.40 Mb). A total of 73,779 Unigenes were supported by more than 5 reads, 42,663 (57.83%) had homologs in the NCBI non-redundant and Swiss-Prot protein databases, corresponding to
27,224 unique protein entries. 16,750 were assigned to Gene Ontology classes, and 14,877 were clustered into orthologous groups. A total of 21,689 (29.40%) were mapped to 119 pathways by BLAST comparison against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. The majority of the genes encoding the enzymes in the biosynthetic pathways of cellulose and lignin were identified in the Unigene dataset by targeted searches of their annotations. Eighteen genes related to cellulose and lignin biosynthesis were cloned for experimental validating of transcriptome data. Overall 49 Unigenes, covering different regions of these selected genes, were found by alignment. The expression analysis of genes has been done through qRT-PCR. The assembled datasets has been used to identify the candidate genes of cellulose and lignin synthesis.

2.4 Sequencing Perspective in Genome Analysis

Wakamiya et al. (1993) reported the genome size and environmental factors affect on Pinus genera. In post genomic era, there was no initiative for comprehensively sequence a gymnosperm genome. The largest genus in the coniferous family Pinaceae is Pinus, whose 110-120 species have extremely large genomes (c. 20-40 Gb, 2N = 24). Conifer genomes are considered with highly repetitive region, but there is little information available on the nature and identity of repetitive units in gymnosperms. The pines have extensive genetic resources, with approximately 329000 ESTs from eleven species and genetic maps in eight species, including a dense genetic map of the twelve linkage groups in Pinus taeda.

The Arabidopsis genome Initiative (2000) concluded that the flowering plant Arabidopsis thaliana is an important model system for identifying genes and determining their functions. The sequenced regions cover 115.4 Mbps of the 125-Mbp genome and extend into
centromeric regions. The evolution of *Arabidopsis* involved whole-genome duplication, followed by subsequent gene loss and extensive local gene duplications, giving rise to a dynamic genome enriched by lateral gene transfer from a cyanobacterial-like ancestor of the plastid. The genome contains 25,498 genes encoding proteins from 11,000 families, similar to the functional diversity of *Drosophila* and *Caenorhabditis elegans*. *Arabidopsis* has many families of new proteins but also lacks several common protein families, indicating that the sets of common proteins have undergone differential expansion and contraction in the three multicellular eukaryotes. This was the first complete genome sequence of a plant and provides the foundations for more comprehensive comparison of conserved processes in all eukaryotes.

Tuskan et al. (2006) reported the draft genome of the black cottonwood tree, *Populus trichocarpa*. The genome sequence analysis has been integrated through the shotgun sequence assembly method with genetic mapping that enabled chromosome-scale reconstruction of the genome. > 45,000 putative protein-coding genes were identified. Analysis of the assembled genome revealed a whole-genome duplication event; about 8000 pairs of duplicated genes from that event survived in the *Populus* genome. Poplar was the second evidence for the genome duplication and divergence of poplar and *Arabidopsis* lineages. Nucleotide substitution, tandem gene duplication, and gross chromosomal rearrangement appear to proceed substantially more slowly in *Populus* than in *Arabidopsis*. *Populus* has more protein-coding genes than *Arabidopsis*, ranging on average from 1.4 to 1.6 putative *Populus* homologs for each *Arabidopsis* gene. Poplar genome has the genes which were associated with lignocellulosic wall biosynthesis, meristem development, disease resistance, and metabolite transport.
Yang et al. (2009) reported the genome sequencing of *Populus trichocarpa* which is commonly known as “black cotton wood” by using the female genotype Nisqually 1. *P. trichocarpa* was selected as the model forest species for genome sequencing not only because of its modest genome size but also because of its rapid growth. *P. trichocarpa* has substantially more protein-coding genes than *Arabidopsis* reported by the whole-genome shotgun strategy adopted for sequencing and assembling the Populus genome. *P. trichocarpa* chloroplast genome was assembled from 139,442 sequence reads and available on website www.genome.ornl.gov. The resulting genome consists of 157,033 bp with the presence of 280,792 sequence reads of mitochondrial genome.

Kovach et al. (2010) sequenced and annotated the ten *P. taeda* BAC clones through the Sanger sequence method. Genome Analyzer II shows that whole genome shotgun (WGS) sequences representing 7.5% of the genome. Computational annotation of ten BACs predicts three putative protein-coding genes and at least fifteen likely pseudogenes in nearly one megabase of sequence. LTR retroelements have been found in three conifer BACs, and identified at least 15 others based on evidence from the distantly related angiosperms. Alignment of WGS sequences to the BACs indicates that 80% of BAC sequences have similar copies with ≥ 75% nucleotide identity elsewhere in the genome, but only 23% have identical copies which contains 99% identity. The three most common repetitive elements in the genome were identified and, when combined, represent less than 5% of the genome. This study indicates that the majority of repeats in the *P. taeda* genome are 'novel'. The pine genome contains a very large number of diverged and probably defunct repetitive elements.

Lin et al. (2010) studies the comparative analysis of chloroplast genomics and analysis of concatenated 49 chloroplast protein-coding
genes common to 19 gymnosperms, including 15 species from 8 Pinaceous genera. The complete cpDNAs of *Cathaya argyrophylla* and *Cedrus deodara* (Abitoideae) and draft cpDNAs of *Larix decidua*, *Picea morrisonicola*, and *Pseudotsuga wilsoniana* are reported. There are 21- and 42-kb inversions in congeneric species and different populations of Pinaceous species, which indicates evolution of structural polymorphism in Pinaceae. The phylogenetic analyses reveal that *Cedrus* is clustered with *Abies–Keteleeria* rather than the basal-most genus of Pinaceae and that *Cathaya* is closer to *Pinus* than to *Picea* and *Larix–Pseudotsuga*. Topology, structural change tests and position of indels-distribution also provide the evidence for phylogenetic analysis. The phylogeny has been analyzed through the maximum-likelihood divergences. As the result concluded that 2 (Abietoideae and *Larix–Pseudotsuga–Piceae–Cathaya–Pinus*), 4 (*Cedrus*, non-*Cedrus* Abietoideae, *Larix–Pseudotsuga*, and *Piceae–Cathaya–Pinus*), or 5 (*Cedrus*, non-*Cedrus* Abietoideae, *Larix–Pseudotsuga*, *Picea*, and *Cathaya–Pinus*) groups/subfamilies are more reasonable delimitations for Pinaceae.

Palle et al. (2011) analyzed the gene expression through the contribution of native populations to understand the plant development and adaptation in multiple ways. These include the identification of candidate genes and genetic polymorphisms affecting expression and phenotypic traits and characterization of transcriptional networks. The expression analysis of 111 genes has been done which have probable roles in xylem/wood development in a population of loblolly pine (*Pinus taeda* L.). Loblolly pine is one of the most commercially important forest tree species in the United States, and the discovery of genes and alleles contributing to desirable wood properties would be valuable. The xylem/wood development genes has been analyzed through the quantitative reverse transcription–polymerase chain reaction, there were
significant differences in gene expression between clones for 106 genes. Genes encoding lignin biosynthetic enzymes and arabinogalactan proteins were more variable than those encoding cellulose synthases or those involved in signal transduction. Several groups of genes with related functions form clusters. The transcription factors are the key regulators for the xylem development in pine. Secondary wall-associated NAC domain protein 1 (SND1) in particular appears to be involved in the regulation of many other genes. The cluster analysis using clones identify the expression differences in various regions.

Lin et al. (2012) determined the complete chloroplast genome (cpDNA) of Ginkgo biloba (common name: ginkgo), the only relict of ginkgophytes from the triassic Period. The cpDNA molecule of ginkgo is quadripartite and circular, with a length of 156,945 bp, which is 6,458 bp shorter than that of Cycas taitungensis. In ginkgo cpDNA, rpl23 becomes pseudo, only one copy of ycf2 is retained, and there are at least five editing sites. It explored the retained ycf2 is a duplicate of the ancestral ycf2, while one copy of ycf2 has been lost from the inverted repeat A (IRA). These inverted repeats conversion led to the contraction of IRs after ginkgos diverged from other gymnosperms. A novel cluster of three transfer RNA (tRNA) genes, trnY-AUA, trnC-ACA, and trnSeC-UCA, was predicted to be located between trnC-GCA and rpoB of the large single-copy region. The phylogenetic analysis strongly suggests that the three predicted tRNA genes are duplicates of trnC-GCA.

Nystedt et al. (2013) studied that Conifers have dominated forests for more than 200 million years and have huge ecological and economic importance. The draft assembly of the 20-gigabase genome of Norway spruce (Picea abies), has been available. The number of well-supported genes (28,354) is similar to the >100 times smaller
genome of *Arabidopsis thaliana*, and there is no evidence of a recent whole-genome duplication in the gymnosperm lineage. Instead, the large genome size seems to result from the slow and steady accumulation of a diverse set of long-terminal repeat transposable elements. There is the lack of an efficient elimination mechanism. Comparative sequence analysis of *Pinus sylvestris, Abies sibirica, Juniperus communis, Taxus baccata* and *Gnetum gnemon* reveals that the transposable element diversity is shared among extant conifers. Expression of 24-nucleotide small RNAs, previously implicated in transposable element silencing, is tissue-specific and much lower than in other plants. Numerous introns have been identified with the length of >10,000 base pairs, these are the gene-like fragments, uncharacterized long non-coding RNAs and short RNAs occurred in the genome. This opens up new genomic avenues for conifer forestry and breeding.

### 2.5 Genomic Databases

Suarez and Galperin (2012) published the 20th annual database issue of nucleic acid research which includes the 176 articles few databases describes the online molecular biology databases and others provides the updates. It includes two databases namely DNA repeat element database and databases with the various aspects of protein structure and protein-protein interaction. The increased emphasis on using the genome data to improve human health is reflected in the development of the databases of genomic structural variation (NCBI’s dbVar and EBI’s DGVa), the NIH Genetic Testing Registry and several other databases centered on the genetic basis of human disease, potential drugs, their targets and the mechanisms of protein–ligand binding. Two new databases present genomic and RNAseq data for monkeys, providing wealth of data on our closest relatives for comparative genomics.
Ahara et al. (2012) published the molecular ecology resource database. This database contains the 299 microsatellite marker loci and nine pairs of single-nucleotide polymorphism (SNP) EPIC primers. Loci were developed for the following species: Alosa pseudoharengus, Alosa aestivalis, Aphis spiraecola, Argopecten purpuratus, Coreoleuciscus splendidus, Garra gotyla, Hippodamia convergens, Linnaea borealis, Menippe mercenaria, Menippe adina, Parus major, Pinus densiflora, Portunus trituberculatus, Procontarinia mangiferae, Rhynchophorus ferrugineus, Schizothorax richardsonii, Scophthalmus rhombus, Tetraponera aethiops, Thaumetopoea pityocampa, Tuta absoluta and Ugni molinae. These loci were cross-tested on the following species: Barilius bendelisis, Chiromantes haematocheir, Eriocheir sinensis, Eucalyptus camaldulensis, Eucalyptus cladocalix, Eucalyptus globulus, Garra litaninsis vishwanath, Garra para lissorhynchus, Guindilla trinervis, Hemigrapsus sanguineus, Luma chequen. Guayaba, Myrceugenia colchaguensis, Myrceugenia correifolia, Myrceugenia exsucca, Parasesarma plicatum, Parus major, Portunus pelagicus, Psidium guayaba, Schizothorax richardsonii, Scophthalmus maximus, Tetraponera latifrons, Thaumetopoea bonjeani, Thaumetopoea ispartensis, Thaumetopoea libanotica, Thaumetopoea pinivora, Thaumetopoea pityocampaena clade, Thaumetopoea solitaria, Thaumetopoea wilkinsoni and Tor putitora and nine EPIC primer pairs for Euphaea decorata, Euphaea formosa, Euphaea ornata and Euphaea yayeyamana.

2.6 Proteomics Analysis

Cygler et al. (1993) studied the protein sequence alignment among the 32 amino acid sequences related to the esterases, lipases and other proteins. The alignment showed that 24 residues were invariant among the 29 hydrolytic sequences and 49 were well conserved. The conserve residue contains the presence of active
sites, disulphide bridges, salt bridges and residue of core proteins. The structural analysis showed that invariant residues positioned on the edges of secondary structural elements which have regular repetitive structures.

Bystroff et al. (1996) reviewed that the study of relationship between the local sequences and its structures has a pivotal role to predict the secondary structure of proteins which is the first step to identify the tertiary structure through the modeling approach. The local sequences also explain the folding of polypeptide chain and show the number of limit for configuration availability to each side of polypeptide chain. In this article the structure prediction analysis from the local sequences and the relationship between them has been analyzed through the protein structure predicting tools. The correlation between the sequence and its structures showed that the similarity among the short sequences or motif reason of proteins.

Kopp and Schwede (2004) review and concluded that comparative protein structure modeling is the accurate method to generate the suitable models for different applications such as drug designing or virtual screening. The accuracy of predicted model through the homology modeling is based on the sequence similarity. The high quality models contain the more than 50% similarity between the target and template sequences with the low resolution value of X-ray crystal structure. If the target-template identity is less 30% means there are significant errors to predict the models. The 3D structure models are very much valuable to understand the molecular function of proteins.

Sitbon and Pietrokovski (2007) studied that the protein structure can be used to identify the traits which have the similarity in the related proteins. Conserve pattern in protein sequences have an important role to identify and study the function as well as the
structure of particular conserve region. Homologous sequences contain the conserve region among the sequences but very few sequences are homogenous along their length. If the sequence has the divergence then it contains the particular site as conserve site.

Shenoy and Jayram (2010) reported, the era of structural biology the central challenge is to rationalize the mass of sequence information into the biochemical, biophysical knowledge to decipher the structural and evolutionary clues. The computational biology investigate the sequences through the two main approaches first is the pattern recognition approach which finds the similarity among the sequences and infer the related structures and functions. The second *ab initio* approach applied to detect the 3D structure and ultimate to infer the function through the linear sequences.