Studies on Regulation of Lignin Biosynthesis gene(s) in *Leucaena leucocephala*

A THESIS
SUBMITTED TO THE
UNIVERSITY OF PUNE

FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
BIOTECHNOLOGY

BY
SUMITA OMER

UNDER THE GUIDANCE OF
Dr. B. M. KHAN

DIVISION OF PLANT TISSUE CULTURE
NATIONAL CHEMICAL LABORATORY
PUNE – 411 008
INDIA

November, 2011

Dedicated to my loving family

TABLE OF CONTENTS

Content **Page**

ACKNOWLEDGEMENTS

The work presented in this thesis was possible due to the direct or indirect contributions of one and all who have been around me since the time I joined this prestigious lab of Plant Tissue Culture Division, NCL.

My foremost heartfelt gratitude to my eminent research guide Dr. B. M. Khan, who gave me the opportunity to be a part of his research group cum family and was always there to guide and support me like a fatherly mentor. The continuous guidance, support and critical feedback given by him was the prime driving force to keep me going inspite of hard times in research tenure.

I would like to specially thank Dr. S. K. Rawal (former HOD, PTC Division, NCL, Pune) for always being there with his invaluable suggestions, which were really useful for me in getting through my work.

I also wish to acknowledge Dr. Sourav Pal, Director, NCL for providing necessary infrastructure to carry out my research.

My sincere thanks to all the scientists of PTC Division, especially Dr. S.V. Kendurkar for sharing her spiritual thoughts and helping me calm my soul. I would like to thank Dr. D.C. Agrawal, Dr.(Mrs.) Thengane and Dr. Urmil Mehta for allowing me to use their lab facilities in crucial times.

I would like to thank Mr. Suryaprasad for using his technical expertise for timely repair and maintenance of the electric and electronic equipments and Mr. Jagtap for maintenance of the transgenic plants.

I am sincerely grateful to my lab seniors Dr. Manish, Dr. Sushim, Dr. Sameer, Dr. Noor, Dr. Arun, Dr. Abhilash and Dr. (Mrs) Pallavi for their suggestions and moral support whenever required. Dr. Mrs. Sucheta deserves special thanks for her cordial nature and support beyond lab limits. Special thanks are due to my current colleagues in the lab, R. J. Santosh, Ruby, Rishi, Somesh, Kannan, Prashant, Parth, Krunal, Neha, Trupti, Uma and Shakeel for all the interactive discussions and making a cordial environment in the lab. I also thank my ex-labmates, Malini, Azfar, Poonam, Rajshree, and Sameena for their friendly presence in the lab.

Special thanks to my trainees, Sneha, Richa, Rajeev and Shikha for their helping hands in crucial period of my research work.

I would also like to thank all my friends in the PTC division, Dr.Swapna, Dr.Varsha, Dr. Bhuban, Dr. Meena, Dr. Prasad, Poonam, Raju and Rita, for their friendly presence and cooperation in the lab vicinity.

I take this opportunity to thank my hostel friends, Urvashi and Tanpreet who have shared several lighter moments with me and been an extended family in my hostel days and till date. I also thank Trupti, Madhurima, Pallavi, Ravindra, Reetika and Sonali for being around me for their cooperation.

I would like to acknowledge the junior and senior research fellowship awarded by the Department of Biotechnology (DBT), India for my research work.

My deepest gratitude goes towards my labmate cum life partner, Mr. Santosh Kumar, who has always stood beside me in my good and bad times with patience and encouragement. I really look forward to a prospective future sharing my life with his love and affection. I equally acknowledge my mother and father-in-law for their care and blessings.

I am falling short of words to mention the unconditional love and support I have received from my parents and dearest sister, Namita in every pace of life, including the difficult times in my PhD tenure.

I might have forgotten to mention many names, but I have to thank my Almighty God for blessing me with good health and balanced state of mind to successfully complete my work as well as in every step of life.

Sumita

DECLARATION

I hereby declare that the thesis entitled **"Studies on Regulation of Lignin Biosynthesis gene(s) in *Leucaena leucocephala*"**, submitted for the Degree of **Doctor of Philosophy** to the University of Pune, has been carried out by me at Plant Tissue Culture Division, National Chemical Laboratory, Pune - 411 008, Maharashtra, India, under the supervision of **Dr. B. M. Khan** (Research Guide). The work is original and has not been submitted in part or full by me for any other degree or diploma to any other University.

November, 2011

Sumita Omer

(Research scholar)
PTC Division, NCL
Pune-411008

CERTIFICATE

This is to certify that the work incorporated in the thesis entitled **"Studies on Regulation of Lignin Biosynthesis gene(s) in *Leucaena leucocephala*"** submitted by **Sumita Omer** was carried out under my supervision at the Plant Tissue Culture Division, National Chemical Laboratory, Pune. Materials obtained from other sources have been duly acknowledged in the thesis.

Dr. B. M. KHAN

(Research Guide)

Abstract

Leucaena leucocephala, a leguminous tree species is mainly used as pulpwood and also as a forage crop in India. However, the presence of lignin which constitutes 20-30% of its wood is considered an undesirable factor in paper making and also limits its digestibility as a fodder crop. Thus there is a lot of economic interest in engineering the lignin content and composition in this plant. Currently, the biosynthetic pathway of lignin has been well defined in only a few tree species. However our understanding on the transcriptional regulation of the pathway is still limited. Transcriptional control of lignin biosynthesis is mediated by a coordinated interplay of specific transcription factors (mostly MYB) which interact directly or indirectly with the cis-regulatory elements (AC elements) found on the promoters of lignin biosynthetic pathway genes. To further our understanding on the regulation of lignin biosynthetic genes in *L. leucocephala*, the present study was conducted.

The first part of the study involved isolation and characterization of partial promoter sequences for a few key enzymes of the lignin biosynthetic pathway: C4H, CCoAOMT, CCR and CAD from *L. leucocephala*. *In silico* analysis of the promoter sequences revealed a number of putative cis elements, including MYBPLANT element (AC-rich motif) known to bind to MYB transcription factors and occur in promoters of genes related to the phenylpropanoid pathway. Biotin-labeled promoter regions for the four genes carrying the putative MYBPLANT element were shown to be strongly retarded by stem nuclear proteins and weakly by leaf nuclear proteins in EMSA experiments, indicating the higher affinity of cis-elements for regulatory proteins, in stem tissue which are more lignified compared to leaf tissue. Partial CCoAOMT promoter (508 bp): : GUS fusion construct was cloned in a promoterless pCAMBIA-1381Z vector and used for genetic transformation of tobacco leaf explants to study its tissue-specific activity in the putative transformants by GUS-histochemical assay. Intense GUS staining was detected in the veins and midrib of the leaf tissue. The stem transverse sections of the transformants showed that promoter conferred GUS expression was present in the vascular bundles (xylem and phloem fibres) and also the pith region. As, a similar study could not be extended for the other three promoters, it may be proposed that functional

conservation of the putative regulatory motifs on all these promoters might confer vascular-tissue specificity to these promoters as well.

The second major part of the study elucidates a functional evidence for the contribution of an R2R3MYB transcription factor in regulating the activity of a few of the above mentioned genes of the pathway. The study involves isolation of a novel R2R3 Myb gene, named *Ll*MYB_SSM from *L. leucocephala* stem tissue cDNA pool using degenerate primer based approach and its functional characterization using various bioinformatic and molecular tools. Phylogenetic analysis supported by MEME software motif search demonstrated that the gene belongs to subgroup 4 of the large R2R3Myb gene family, with AmMYB308 as the closest putative ortholog of the gene forming a separate clade along with other members of subgroup 4 like EgMYB1, ZmMYB31, ZmMYB42 and AtMYB4, which are known to be repressors of phenylpropanoid pathway. To investigate the repressor function of the gene, *in planta* gain-of-function approach was used by over-expressing the gene driven by a strong CaMV35S promoter in tobacco plants. *Ll*MYB_SSM was shown to down-regulate PAL, C4H and 4-CL genes involved in the general phenylpropanoid pathway and up-regulate CCoAOMT gene, with slight reduction in CAD gene transcript levels in the putative transgenics, which were slightly dwarf with respect to the control plants. However, the heterologously expressed *Ll*MYB_SSM protein from *E. coli* did not show any binding with the MYBPLANT element present in the promoters characterized in the first part of the study, in the *in vitro* DNA-binding experiments. This may be reasoned by the absence of post-translational modifications or requirement for partner proteins required for DNA-binding affinity of the transcription factor. Spatio-temporal expression of the gene investigated in different stages of growth of the *L. leucocephala* seedling in leaf, stem and root tissues, showed its peak of expression in 10 and 20 day old stem tissues, which drastically declined from30 to 60 days. The results showed that the gene expression level decreased with increase in lignification, suggesting that repressor function of the gene is absent in more lignified tissues. The functional role of an EAR repressor motif present in the gene was determined by subjecting 15 day *L. leucocephala* seedlings to various stress conditions like UV, salt and jasmonic acid; all of which resulted in different levels of increase in *Ll*MYB_SSM transcript level. The highest fold increase was seen with UV-C stress

suggesting the role of the conserved EAR motif in UV-stress conditions, which in turn influences the stress-induced lignin accumulation in the plant.

The overall results indicate that *Ll*MYB_SSM might play an indirect role in repression of lignin biosynthesis pathway in *L. leucocephala* and hence can serve as a potential candidate gene in manipulating lignin biosynthesis for biotechnological applications.

.

ABBREVIATIONS

%	Percentage
°C	degree Celsius
4CL	4-Coumarate coenzyme A ligase
AA	amino acid
AAP	Abridged Anchor Primer
AldOMT	5-Hydroxyconiferaldehyde O- methyltransferase
BAP	6- Benzylaminopurine
BCIP	5-bromo-4-chloro-3'indolyphosphate
bp	base pairs
BSA	Bovine serum albumin
C3H	Coumarate 3- hydroxylase
C4H	Cinnamate-4-Hydroxylase
CAD	Cinnamyl alcohol dehydrogenase
CaMV	Cauliflower mosaic virus
CCoAOMT	Caffeoyl coenzyme A 3-O- methyltransferase
CCR	Cinnamoyl coenzyme A reductase
cDNA	Complementary DNA
Ci/ mmol	Curie per milli mole
CIAP	Calf Intestinal Alkaline Phosphatase
COMT	Caffeate O-methyltransferase
Cps	Counts per second
Da	Dalton
DEPC	Diethylpyrocarbonate
DNA	Deoxyribose nucleic acid
DTT	Dithiothritol
EDTA	Ethylene Diamine Tetra Acetic acid
ELISA	Enzyme linked immuno sorbent assay
F5H/CAld5H	Ferulate 5-hydroxylase/ coniferaldehyde 5-hydroxylase
g	gram

G	Guaiacyl
g /L	grams per litre
gDNA	Genomic DNA
GSP	Gene Specific Primers
h	Hour(s)
IPTG	Isopropyl β-D-thiogalactoside
Kb	Kilobase pairs
KDa	Kilo Daltons
Kg	Kilogram
L	Litre
MCS	Multiple cloning sites
mg	milligram
min	Minute(s)
mL	millilitre
mM	millimolar
mRNA	messenger RNA
NAA	1-Napthyl aceticacid
NBT	nitro blue tetrazolium
nM	nano molar
nm	nanometer
NUP	Nested Universal Primers
O/N	Overnight
PAL	Phenylalanine ammonia lyase
pg	picogram
pmol	picomole
PMSF	Phenyl methyl sulphonyl fluoride
ppm	Parts per million
RACE	Rapid amplification of cDNA ends
RNA	Ribose nucleic acid
rpm	Rotations per minute
RT	Room temperature

s	second(s)
S	Syringyl
SAD	Sinapyl alcohol dehydrogenase
SDS	Sodium dodecyl sulphate (sodium lauryl sulphate)
SMQ	Sterile Milli Q
Soln	Solution
sp.	Species
TAE	Tris acetic EDTA buffer
TEMED	Tetramethylethylenediamine
U	Units
UDP-GT	UDP-glycosyltransferase
UPM	Universal Primer Mix
UTR	Untranslated Region
UV	Ultraviolet
V	Volt
v/v	volume / volume
w/v	weight / volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
α	Alpha
β	Beta
λ	Lamda
µg	microgram
µg/L	Micrograms per liter
µL	microlitre
µm	micrometer
µM	micromolar

CHAPTER 1

INTRODUCTION

1. Introduction

Vascular plants appeared on earth during the Silurian period around 430 million years ago and one of the features important for their conquest on dry land is the evolution of vascular tissues that solved the problem of fluid transport throughout the plant body .(Raven *et al*,1999) Accompanying this phenomenon was the evolution of the ability for vascular plants to synthesize lignin, a biopolymer which provides mechanical strength and hydrophobicity to the water-conducting tracheary elements and fibres, which are integral components of the vascular tissues. Although lignin is essential for normal plant growth and development, its presence in plant tissues may hinder the uses of plant products by humans. Its presence is undesirable for utilisation of lignocellulosic biomass as an alternative source of biofuel and also decreases the digestibility in animal forage crops. In trees, lignin synthesis is of major importance because of the production of wood (i.e. secondary xylem) which is typically made up of 20–30% lignin on a dry weight basis. As wood is widely used for paper-making, lignin has to be chemically removed during the pulping process, which not only increases the cost of pulping but also leads to production of chemical wastes that cause severe adverse effect on the environment. (Boerjan *et al.*, 2003).

The formation of wood entails the partitioning of a significant proportion of fixed carbon resources into the synthesis of lignin-building blocks through the phenylpropanoid pathway (Amthor, 2003; Boerjan *et al*., 2003). Although the, enzymes and genes of phenylpropanoid and monolignol biosynthetic pathways have been extensively studied (Humphreys and Chapple, 2002; Boerjan *et al*., 2003), still little is known about the mechanisms regulating their expression. Experimental evidences indicate that monolignol synthesis genes are under transcriptional control (Rogers and Campbell, 2004). Among the different classes of transcription factors (TFs) directly or indirectly implicated in lignification, R2R3-MYBs are strong candidates for the regulation of phenylpropanoid enzymes and monolignol biosynthesis (Rogers and Campbell, 2004; Groover and Robischon, 2006). Various MYBs have been shown to be involved in influencing the monolignol biosynthetic genes either by directly interacting with the AC elements found on their promoters or indirectly via intermediate effects.

The research on R2R3-MYBs in lignin biosynthesis has focused on a few model species, which include both trees (*Eucalyptus, Pine* and *Populus*) and herbaceous plants (*Arabidopsis thaliana* L. and *Nicotiana tabacum* L.).*Leucaena leucocephala* is an economically useful tree for paper and pulp industry, the lignin biosynthetic genes for which have been well characterized in our lab. Bioinformatic analysis of promoter sequences for a few genes of the pathway for the plant, showed the presence of cis-regulatory AC-elements known to interact with MYB transcription factors. A novel lignin related R2R3-type MYB transcription factor from the plant has been studied to determine its regulatory role in lignification. This chapter presents a review of literature in relation to the role a leguminous tree species, *Leucaena leucocephal*a in pulp and paper industry and the importance of studying transcriptional regulation of lignin biosynthetic pathway genes. The main emphasis is on the R2R3 type MYB transcription factors and their role in regulating the phenylpropanoid pathway genes.

1.1 *Leucaena leucocephala*

Fig. 1.1: (A) *Leucaena leucocephala* in its bloom showing the young green pods (B) *L.leucocephala* mature brown pods and seeds

Leucaena leucocephala (Lam.) de Wit., popularly known as *subabul* in India (Rao *et al*., 1984) belongs to the Leguminosae family and is one of the fastest-growing, productive and versatile multipurpose leguminous trees. The genus is native to southern Mexico and northern Central America (Belize and Guatemala) (Hughes, 1998) but now naturalized throughout the tropics. The specific name is derived from the Greek words *leuco*, meaning "white", and *cephala*, meaning "head", referring to its flowers. It is essentially a tropical species requiring warm temperatures (25-30^{o}C) for optimum growth and with poor cold tolerance. For optimal growth it is therefore limited to areas 15-25 deg. north or south of the equator. It is arborescent, deciduous, nitrogen fixing species, growing up to 20 m tall.

1.1.1 Classification

Leucaena occupies 2 to 5 million hectares of land worldwide (Brewbaker and Sorenson, 1990). They are recognized as some of the fastest growing and most useful trees in the tropics (NRC, 1984). Members of the genera are vigorous, drought tolerant, highly palatable, high yielding and rich in protein and grow in a wide range of soils (Jones, 1979; Hughes, 1998). However, these attributes are limited by the occurrence of anti-nutritive factors in the fodder, such as tannins and mimosine (Jones, 1979; Hegarty *et al*., 1964b; Hammond *et al*., 1989 a, b).

The genus, *Leucaena* is represented by 22 species. Of these, 6 are intraspecific taxa and 2 are widespread spontaneous hybrids. Most of the species are diploid (2n=52 or 56). However, 4 species (*L. leucocephala, L. diversifolia, L. pallida and L. confertiflora*) are tetraploid 2n = 4X=104 or 112) (Hughes, 1998). *L. leucocephala* is a member of the genus related to the other species within the Mimosoideae sub-family, its subspecies and other related genera. It has been classified as follows:

Kingdom	Plantae
Super division	Tracheobionta
Division	Spermatophyta
Subdivision	Magnoliophyta
Class	Magnoliopsida
Subclass	Rosidae

Order	Fabales
Family	Fabaceae
Subfamily	Mimosoideae
Tribe	Mimoseae
Genus	*Leuceana*
Species	*Leucocephala*
Common name	Lead tree, white popinac ,subabul
Sub species	Glabrata (Rose; S. Zarate); Ixtahuacana (Hughes) and Leucocephala (Benth) Var. Peru and Cunningham
Related genera	*Desmanthus*; *Schleinitzia*; *Calliandropsis*; *Neptunia*; *Alantsilodendron*; *Gagnebina*; *Dichrostachys*; and *Kanaloa*

1.1.2 Growth forms

L. leucocephala has three growth forms that are commonly found in tropics (Hughes, 1998). The first and common type is a shrubby free-shading form known as the Hawaiian Leucaena, which is weedy and low yielding. It is a bushy variety and flowers all year round. The second growth form is giant type, which is a tall tree of about 20 meters, with large leaves, pod and seeds, and almost without branches. It is native of Central America and Mexico, and produces twice the biomass of the common type. It is suitable for timber, wood product and industrial fuel. The last is Peru type, which are a multi-branched, semi-erect, medium height (about 10 meters) and the most productive form. It also produces high quality foliage which is good for forage.

1.1.3 Uses as a multipurpose tree

During the 1970s and 1980s *L.leucocephala* was promoted as a "miracle tree" for its multiple uses (Gutteridge *et al*, 1998). It has also been described as a "conflict tree" in that it is both promoted for forage production and spreads like a weed in some places. However, it is an important crop encouraged under the social forestry schemes in drought-prone areas and semi-arid tracts in India, as it provides useful timber as well as leaves for fuel and energy and feed purposes (Chandrasekhara *et al.*, 1984; Relwani *et al.*, 1981; Azeemoddin *et al.*, 1988; Krishnan *et al.*, 1990 and Torane *et al.*, 1990).Various uses of the tree wood are discussed as follows:

Fodder: *L. leucocephala* is one the highest quality and most palatable fodder trees of the tropics, often being described as the 'alfalfa of the tropics'. The leaf quality compares favourably with alfalfa or lucerne in feed value except for its higher tannin content and mimosine toxicity to non-ruminants. Leaves have a high nutritive value (high palatability, digestibility, intake and crude-protein content), resulting in 70-100% increase in animal live weight gain compared with feeding on pure grass pasture. In addition, it is very persistent over several decades of cutting or grazing, is highly productive and recovers quickly from defoliation. Forage, packed in pellets and cubes, is internationally marketed as animal feed.

Apiculture: *L. leucocephala* is in bloom almost throughout the year, providing constant forage to honey bees.

Fuel: *L. leucocephala* is an excellent firewood species with a high calorific value of 4600 cal/kg. Wood burns steadily with little smoke, few sparks and produces less than 1% ash. The tree makes excellent charcoal with a heating value of 29 mJ/kg and good recovery values (25-30%). Addition of ground *L. leucocephala* to fuel oil for diesel engines was found to involve no harmful agents in the ash.

Fibre: Fibre values are similar to those of other tropical hardwoods, and it produces paper with good printability but low tearing and folding strength; the wood-pulp strength is greater than that of most hardwoods, with almost 50% greater ring crush. Its pulping properties are suitable for both paper and rayon production. Also used for particleboard production.

Timber: *L. leucocephala* has hard heavy wood (about 800 kg/m), with a pale yellow sapwood and light reddish-brown heartwood. The wood is known to be of medium density and to dry without splitting or checking. It is strong, medium textured, close grained and easily workable for a wide variety of carpentry purposes. Sawn timber, mine props, furniture and parquet flooring are among increasingly popular uses. There is growing use of small-dimension sawn wood in a number of industries such as flooring, which might include *L. leucocephala* in the future.

Gum or resin: Gum arises from *Leucaena* stems under ill-defined conditions of injury and disease or from sterile hybrids, especially *L. leucocephala* x *L. esculenta.* The gum has been analysed and found similar to gum arabic and of potential commercial value.

Tannin or dyestuff: Red, brown and black dyes are extracted from the pods, leaves and bark.

1.1.4 Role in pulp and paper industry

India is the fastest growing pulp and paper market in the world with 6% demand growth. Paper consumption in India is expected to grow to 14mt pa by 2015 (Thapar, 2009). The major sources of raw material for our paper industry is accounted by non-wood fibres such as bagasse and wheat straw, which represents about 30% and recycled fibre represents 36% of the resources. However, hardwood species such as *Eucalyptus, Poplar, Casuarina* and *Leucaena* contribute about 30% of the resources. In bamboo growing countries, like India, the proportionate use of bamboos and hardwood species is in the ratio of 15:85. Although all these plant species are of importance to the paper industry, *Leucaena sp.* is extensively used in India and about 25% of hardwood raw material for pulp and paper industry comes from this tree.

Leucaena is one of the few trees from which wood is used for both the industrial and non industrial purposes. Wood of *Leucaena* is generally described as being strong, light in weight, easy to work and able to give attractive finish (Rao, 1984).These qualities make *Leucaena* wood suitable to a wide range of uses, ranging from the traditional small scale use by farmers and small holders to the more recent utilization by large scale industries for pulp and energy generation (Pottinger and Hughes, 1995).

L.leucocephala has porous and fine textured wood structure with longer fiber than that of other hardwoods. It produces pulp that is high in holocellulose, a-cellulose and low lignin, silica, ash, alcohol-benzene soluble, hot water soluble, with xylan type hemicelluloses and a pulp yield of 50 to 52% making it a suitable raw material in pulp and paper industry (Pottinger and Hughes, 1995; Malik *et al.*, 2004)

To meet the increasing demand of high quality wood for paper industry, it is essential to provide designer plant species. However, as a safeguard for the future no plant should

be harvested from areas that may challenge sustainability. It will thus be crucial to raise plantations of the plant species with elite materials and or genetically modified plants that meet the demands of the pulp and the paper industry in economical and sustainable manner. However no study has been done on lignin biosynthesis gene(s) so far in *Leucaena sp.* and study of regulation of these gene(s) will help in understanding the mechanisms controlling lignin biosynthetic pathway in *Leucaena sp*. and its manipulation so as to meet the needs of pulp and paper industry.

1.2 Lignin biosynthesis in nature

The ability to synthesize lignin has been essential in the evolutionary adaptation of plants from an aquatic environment to land and provides crucial structural integrity to the cell wall and stiffness and strength of the stem (Chabannes *et al.,* 2001) Lignin biosynthesis is temporally and spatially controlled, and is closely associated with the differentiation of sclerenchyma cells during normal plant growth and development. Its biosynthesis can also be induced in response to wounding or infection, a process independent of developmentally regulated lignification.

1.2.1 Lignin

Lignin, a complex phenylpropanoid heteropolymer, is the second most abundant natural product after cellulose, accounting for nearly 30% of the organic carbon in plant biomass. It is deposited mainly in cell walls of supporting and conducting tissues, such as fibers and tracheary elements.

The lignin polymer is formed by the oxidative polymerization of hydroxycinnamoyl alcohols, or 'monolignols'(*p*-coumaryl, coniferyl, and sinapyl alcohol),that differ in the degree of methoxylation at the C3 and C5 positions of the aromatic ring (Fig1.2) under the action of peroxidases and/or laccases (Gavnohlt *et al*, 2002; Ranocha *et al*, 2002). When incorporated into lignin, these alcohols are called the *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units of the polymer, respectively. With some notable exceptions (Novo Uzal *et al*., 2009), lignins from gymnosperms (softwoods) are composed of G-units only (with minor amounts of H-units), whereas angiosperm (hardwoods) dicot lignins are composed of G- and S-units. H-units are

elevated in softwood compression wood and may be slightly higher in grasses (Boerjan *et al*., 2003).

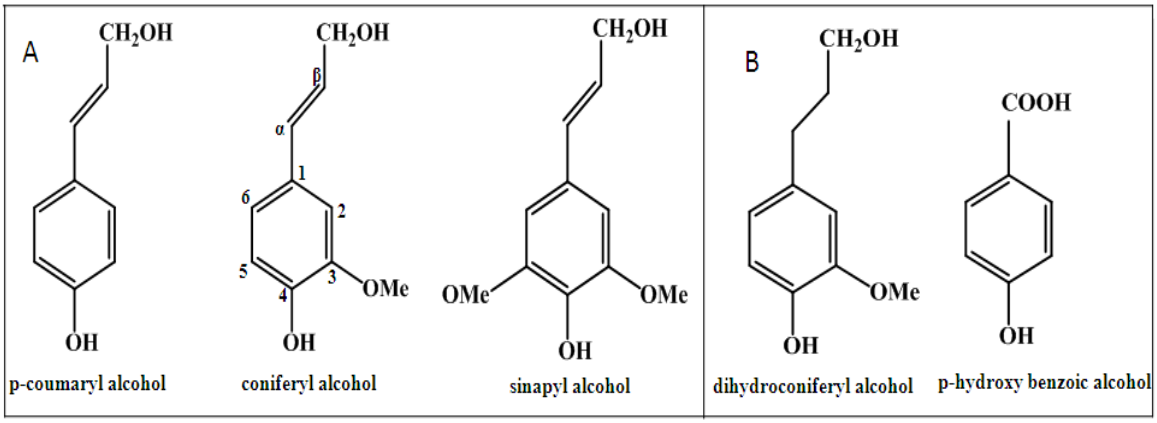


Fig. 1.2: Structures of the three traditional lignin precursors (A) *p*-coumaryl, coniferyl and sinapyl alcohol, and (B) dihydroconiferyl alcohol and *p*-hydroxybenzoic acid that are incorporated to a lesser extent into lignin in gymnosperm and *Populus* species, respectively

1.2.2 Process of lignification

Lignification process occurs in three steps: Monolignol biosynthesis, Transport and storage of monolignols and polymerization of monolignols.

Monolignol biosynthesis: The cinnamate/monolignol pathway provides precursors for various phenylpropanoid compounds including lignins, lignans, neolignans, p-hydroxycinnamate esters, coumarins, suberins, flavonoids, stilbenes and so on. Therefore, the pathway plays the central role in plant secondary metabolism. The pathway is connected to the primary metabolism through the shikimate/chorismate pathway that produces phenylalanine, which is the entry compound of the carbon skeletons into the phenylpropanoid metabolism (Lewis *et al*., 1999).

The lignin monomers or monolignols (p-coumaryl, coniferyl, and sinapyl alcohol), which are main building blocks for lignin, are one of the end products produced intracellularly through this pathway, through a series of enzyme catalyzed hydroxylation, methylation and reduction reactions (Fig. 1.3).

Starting from phenylalanine, monolignol biosynthesis is sequentially catalyzed by about 10 enzymes , namely phenylalanine ammonia lyase (PAL), cinnamic acid 4-hydroxylase (C4H), 4-cinnamoyl CoA ligase (4CL), cinnamoyl CoA:shikimate

hydroxycinnamoyl transferase (HCT), *p*-coumaroylshikimate 3′-hydroxylase (C3′H), caffeoyl CoA *O*methyltransferase (CCoAOMT), cinnamoyl CoA reductase (CCR), ferulic acid 5-hydroxylase (F5H), caffeic acid/5-hydroxyferulic acid *O*-methyltransferase (COMT), and cinnamyl alcohol dehydrogenase (CAD) (Boerjan *et al*., 2003). The activities of these enzymes lead to deamination, hydroxylation, transacylation, methylation, and reduction of phenylalanine, and transform it to the lignin monomeric precursors, *p*-coumaryl, coniferyl, and sinapyl alcohols (*i.e.*, monolignols) (Boerjan *et al.*, 2003; Li *et al.*, 2008).

Among these enzymes, three cytochrome P450 proteins, *i.e.*, cinnamic acid 4-hydroxylase (C4H), *p*-coumaroylshikimate 3′-hydroxylase (C3′H), and ferulic acid 5-hydroxylase (F5H) are membrane bound proteins and associate with the outer surface of the endoplasmic reticulum by virtue of their N-terminal membrane anchor (Li *et al*, 2008). Nevertheless, most other enzymes, such as PAL, 4-CL, COMT, CCoAOMT and CAD in diverse species are soluble and localize within cytosol that were revealed by biochemical and immuno-cytochemical studies, although a few studies suggested that PAL and CAD may have different types of isoforms, and some of them may associate to the ER-Golgi derived vesicles (Takabe *et al.*, 1985; Nakashima *et al.*, 1997).

The cytosolic localization of those biosynthetic enzymes suggests that monolignols are synthesized in the cytoplasm. As many enzymes of the pathway are encoded for by small gene families, spatial and temporal associations with lignification have been used as markers for the involvement of a specific gene/enzyme in lignin biosynthesis.

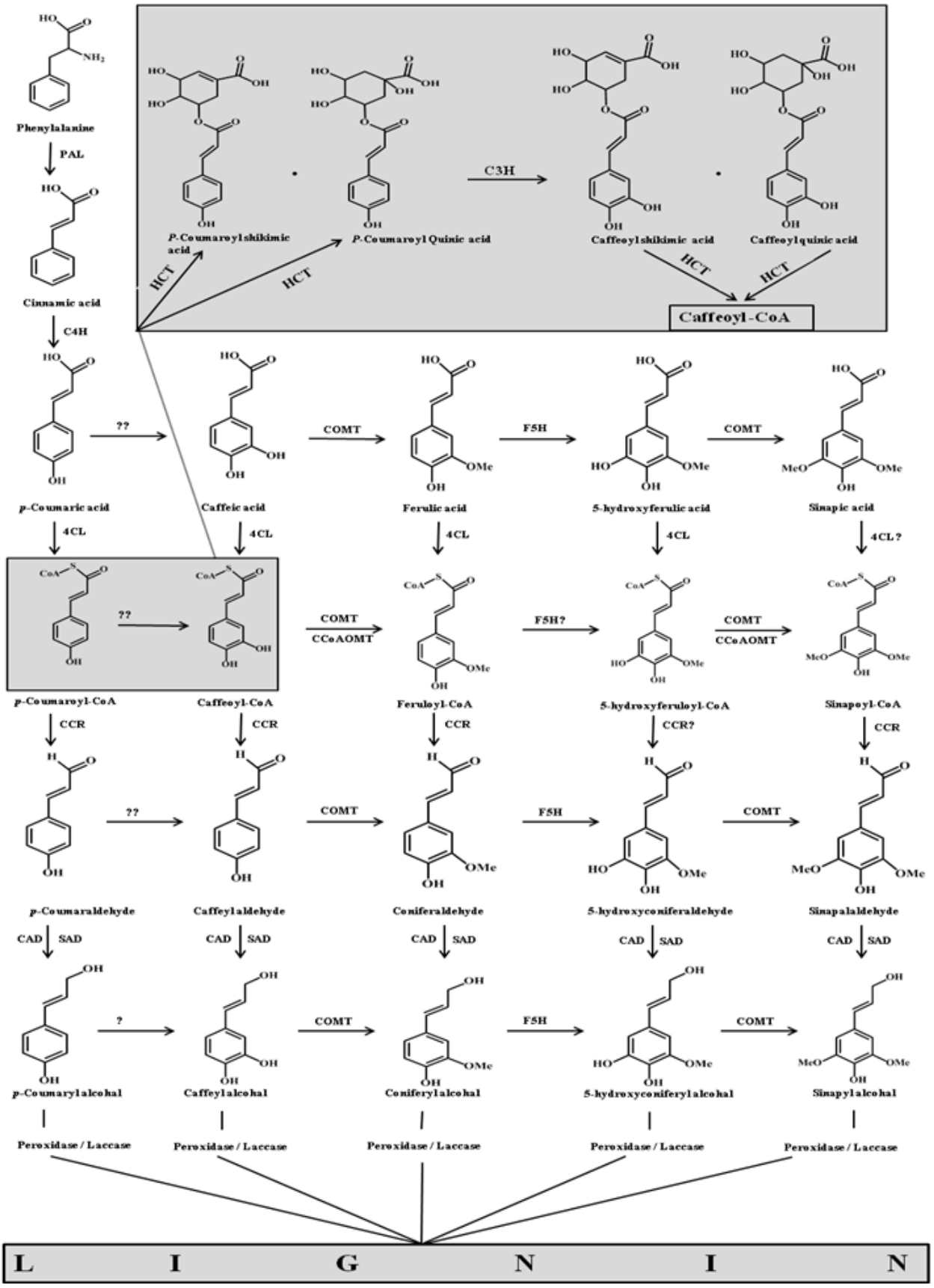

Phenylalanine
PAL
Cinnamic acid
C4H
p-Coumaric acid
4CL
p-Coumaroyl-CoA
CCR
p-Coumaraldehyde
CAD SAD
p-Coumarylalcohal
Peroxidase / Laccase
??
Caffeic acid
4CL
Caffeoyl-CoA
CCR
Caffeylaldehyde
CAD SAD
Caffeylalcohal
Peroxidase / Laccase
P-Coumaroylshikimic acid
P-Coumaroyl Quinic acid
HCT
HCT
C3H
Caffeoylshikimic acid
Caffeoylquinic acid
HCT
HCT
Caffeoyl-CoA
COMT
Ferulic acid
4CL
Feruloyl-CoA
CCR
Coniferaldehyde
CAD SAD
Coniferylalcohal
Peroxidase / Laccase
COMT
CCoAOMT
F5H
5-hydroxyferulic acid
4CL
5-hydroxyferuloyl-CoA
CCR?
5-hydroxyconiferaldehyde
CAD SAD
5-hydroxyconiferylalcohal
Peroxidase / Laccase
COMT
CCoAOMT
Sinapic acid
4CL?
Sinapoyl-CoA
CCR
Sinapalaldehyde
CAD SAD
Sinapylalcohal
Peroxidase / Laccase
LIGNIN

Fig. 1.3: The main biosynthetic route toward the monolignols p-coumaryl, coniferyl, and sinapyl alcohol (Boerjan *et al*, 2003). PAL: phenylalanine ammonia lyase ; **C4H**: cinnamic acid 4-hydroxylase ; **4-CL**: 4-cinnamoyl CoA ligase ; **HCT**: hydroxycinnamoyl CoA:shikimate hydroxycinnamoyl transferase ;**CCoAOMT**: caffeoyl CoAmethyltransferase; **CCR**: cinnamoyl CoA reductase; **F5H:** ferulic acid 5-hydroxylase ;**COMT**: caffeic acid/5-hydroxyferulic acid *O*-methyltransferase; **CAD**: cinnamyl alcohol dehydrogenase

Storage and transport of monolignols: Monolignols are relatively toxic, and unstable. After their synthesis, they are either stored, putatively in the vacuole, as more hydrophilic and less toxic conjugates, or transported into apoplast (Boerjan *et al*., 2003).Glucosylation of small molecule compounds is known to reduce their lipophilicity, thus preventing any further possibility of free diffusion across the lipid bilayer (Bowles *et al*,2006).

In gymnosperms and some angiosperm species, monolignols are often glycosylated to form 4-*O*-β-D-glucosides, namely, coniferin and syringin (Whetten and Sederoff, 1995).Monolignol glycosylation is catalyzed by soluble UDP-glucose: coniferyl alcohol or sinapyl alcohol glucosyltransferase (Steeves *et al*.,2001), which have been detected in crude homogenates of all gymnosperms tested and woody angiosperm species as preferred to the herbaceous ones (Ibrahim *et al*.,1977).On the other hand, specific β-glucosidases, which have been characterized in both the species(Marcinowski & Grisebach,1978; Hösel *et al*.,1978; Hösel & Todenhagen,1980; Leinhos *et al*.,1994) are proposed to release monolignols from the glucoconjugates in cell wall of differentiated xylem tissues for polymerization into lignin. Thus a UDPG :coniferyl alcohol glucosyltransferase/coniferin β-glucosidase (CG) system was proposed to regulate the storage and mobilization of monolignols for lignin biosynthesis and plant lignification (Dharmawardhana *et al*., 1995,1999; Samuels *et al*.,2002; Escamilla-Trevin˜o.,2006).To demonstrate their role in lignification, down regulation studies of these genes was carried out in *Arabidopsis* plants(Lanot,2006). However, no significant lignin phenotype was observed in the plants down- or up-regulated with UDPG-glycosyltransferase gene, although corresponding reduction or accumulation of the soluble monolignol glucosides was observed respectively. This data supported by radiotracer experiments (Kaneda *et al*., 2008) argue against the role of monolignol glucosylation in the export of monolignols

across the plasma membrane and imply that monolignol aglycone may be the chemical form for the transport. Therefore, the role of monolignol glucosylation may be to convert the highly active, unstable lignin precursors into the "storage form", as precursor reservoirs shielded them in a particular compartment.

Three transport mechanisms have been suggested for monolignols. Firstly, vesicular secretion which involves exocytosic transport via the Endoplasmic Reticulum-Golgi derived vesicles. Early autoradiographic, immunocytochemical, and ultrastructural studies suggested that transport of lignin precursors was also potentially via a similar ER-Golgi route as does the secretion of the wall matrix polysaccharides. However, a recent study, suggests that ER-Golgi-vesicle mediate exocytosis is unlikely to play a major role in monolignol transport (Kaneda *et al*., 2008).

Passive diffusion is another proposed mechanism for monolignol transport based on the plasticity of lignin composition. Accommodation of alternative monomers in lignifcation infers a potential non-specific passive diffusion of lignin precursors across the plasmamembrane (Vanholme *et al*., 2008).Although supported by experimental results (Boija *et al*., 2007, 2008) the mechanism is not well understood.

Another possibility is the ATP-binding cassette (ABC) transporters that carry small-molecular-weight compounds across membranes (Yazaki, 2006). Several ABC transporters were co-regulated with lignin biosynthetic genes during the development of *Arabidopsis* inflorescence stem (Ehlting *et al*., 2005). To clarify whether the transport of lignin precursor is the membrane transporter - mediated active process, exploration was recently conducted by the *in vitro* uptake assay that incubates the isolated plasma and vacuolar membrane vesicles from both *Arabidopsis* young rosette leaves and the roots of poplar with monolignols and/or their glucosides (Liu *et al*., 2011). The studies demonstrated that the transport of lignin precursors across both plasma and vacuolar membranes is largely dependent on the presence of ATP. The specific ABC-type transporter inhibitors largely reduced the transport activity of plasma or vacuolar membrane vesicles to lignin precursors. Furthermore, in the presence of ATP, the vacuolar membrane vesicles prepared from *Arabidopsis* rosette leaves displayed considerable activity in sequestering the glucoconjugates coniferin and syringin, but a very limited activity to monolignol aglycones. In contrast, the plasma membrane vesicles

were inactive to the glucoconjugated monolignols and preferred for the aglycones (Fig. 1.4). These data suggest that glucosylation of monolignols is a prerequisite for their vacuolar sequestration while the aglycones are required for the direct transport into cell walls of *Arabidopsis.* Another observation in the above study was the promoiscious active transport across plasma membranes due to its lesser selectivity, also conveying non-classic lignin precursors for deposition on the cell wall, thus explaining the plasticity of lignin composition.

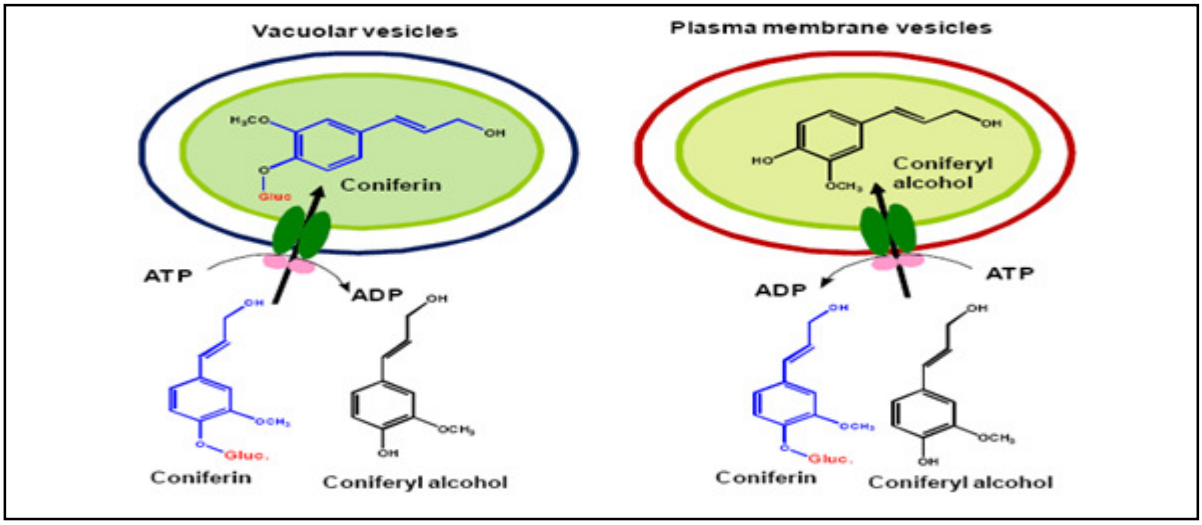


Fig. 1.4: ATP-dependent, selective uptake of monolignols or their glucosides by vacuolar vesicles and the inside-out plasma membrane vesicles (Liu *et al*., 2011)

Polymerisation : Lignin polymerization occurs via oxidative radicalization of phenols, followed by combinatorial radical coupling. According to the latest lignification model, in the first step, the monolignol phenol is oxidized (i.e. dehydrogenated) to generate a phenolic radical. Monolignol dehydrogenation involves peroxidases and/or laccases. Whereas peroxidases use hydrogen peroxide as a substrate, laccases use oxygen to oxidize their metal centers to enable catalytical phenol oxidation. The second step involves, coupling between two monomer radicals to form a (dehydro) dimer, thereby establishing a covalent bond between both subunits .This radical-radical coupling occurs in a chemical combinatorial fashion; thus, the ratio of each of the possible coupling products depends largely on the chemical nature of each of the monomers and the conditions in the cell wall (Ralph *et al*., 2004). Then, the dimer needs to be dehydrogenated, again to a phenolic radical, before it can couple with another monomer

radical. This mode of action, in which a monomer (radical) adds to the growing polymer, is termed endwise coupling: the polymer grows one unit at a time. Coupling of two lignin oligomers is rare in S/G lignins but relatively common in G-lignins, where 5-5 coupling accounts for approximately 4% of the linkages (Argyropoulos *et al*., 2002; Wagner *et al*., 2009). During each coupling reaction, two radicals are "consumed" (in a so-called "termination reaction") as each single electron contributes to the newly formed bond. The average length of a linear lignin chain in poplar is estimated to be between 13 and 20 units (Reale *et al*., 2004; Stewart *et al*., 2009).

Polymerisation process requires a direct oxidation of the growing polymer by peroxidases. The mechanism requires that peroxidases directly oxidize both lignin macromolecules and monolignols, and that a phenolic radical on the lignin macromolecule be coupled to a monolignol radical. It is unlikely, however, that cell wall bound peroxidases can oxidize directly three-dimensional lignin macromolecules as substrates because of the limited freedom of their movements. The model explains that, although monolignols might be dehydrogenated via direct interaction with an electron-removing (oxidizing) enzyme (peroxidases and/or laccases), the radicals might alternatively be generated by radical transfer.

Peroxidases (EC 1.11.1.7) which play a key role in radical generation may differ in their substrate specificities; whereas some almost exclusively accept coniferyl alcohol, others are highly specific toward sinapyl alcohol (Marjamaa *et al*., 2006; Go´mez Ros *et al*., 2007).Because the structure of lignin depends on the availability of monolignol radicals, peroxidase specificity may determine in part the structure of the final lignin polymers, opening possibilities for altering lignin structure by modified expression of specific peroxidase isoforms. Several reports on peroxidase activity or gene expression in lignin-forming tissues have appeared, but only a few isoenzymes or genes have been specifically associated with lignification (Marjamaa *et al.,* 2006). Peroxidases from *Populus alba* (L.) and *Z. elegans* were shown to oxidize Sinapyl alcohol efficiently; (Gabaldon *et al.,* 2005).

In conclusion, the current model of lignification involves peroxidases and/or laccases to provide the oxidative capacity in the cell wall. All phenolic compounds that enter this region will eventually have the potential to radicalize and incorporate into the

lignin polymer, subject to simple chemical oxidation and (cross-) coupling propensities. This model also nicely explains why many other phenolic molecules can be integrated into the growing lignin polymer and opens up the possibility of tailoring lignins for industrial applications by regulating the influx and types of monolignols into the cell wall (Ralph, 2006; Grabber *et al*., 2008).

1.3 Regulation of lignin biosynthesis

Understanding lignin biosynthesis has long been an area of interest to plant biologists because of its applications in tree biotechnology. As there are a number of genes involved in the pathway, it needs to be coordinately turned on to make lignin. In addition there is enormous variation in lignin content and composition among plant species, tissues, cell types, and even developmental stages and environmental conditions play a role. Currently, the biosynthetic pathway of lignin has been well defined and a number of recent reviews cover the topic of lignin biosynthesis and its genetic modification (Vanholme *et al.*, 2008; Li *et al.*, 2008). In contrast, our understanding on the transcriptional regulation of the lignin biosynthetic pathway is still limited. This section focuses on regulatory mechanisms and transcription factors involved in the coordinated activation of lignin biosynthetic genes.

1.3.1Control of lignification

The spatial and temporal control of lignification is critical in plant development to ensure appropriate timing and localization of lignin which is crucial for plant support, water transport and disease resistance. Several reports suggest that the timing of lignin biosynthesis is controlled by carbon supply, which is directly linked to light and photosynthetic activity. The carbon-nutrient balance hypothesis proposes that secondary metabolism is directed towards carbon-rich compounds in nitrogen-limited plants and to nitrogen rich compounds in carbon-limited plants (Coley *et al*., 1985).In support of the hypothesis increased levels of phenylpropanoids including lignin accumulation was observed under limited nitrogen supply (e.g., Gebauer *et al*., 1997; Fritz *et al*., 2006). Thus various physiological and chemical factors are known to control lignin biosynthesis. Other mechanisms of control of lignification are discussed as below:

Metabolic control:

Control of metabolite flux into and through the lignin biosynthetic pathway plays an important role in determining the variation in lignin quality, quantity and distribution seen in nature. Flux into the pathway is likely to be affected by entry-point enzymes, whereas flux through the pathway may be influenced by levels of enzyme activity and by metabolic channeling of substrates and products (Kacser and Burns, 1973).Amongst the early pathway genes, a feedback regulation was observed between PAL and C4H enzymes, sensing the cinnamic acid pool in tobacco, which resulted in an alteration of flux into the phenylpropanoid pathway (Bolwell *et al*., 1986; Blount *et al*., 2000).CCR was another enzyme which was proposed to channel metabolites from general phenylpropanoid biosynthesis into the monoIignol biosynthetic pathway, by catalyzing the conversion of hydroxycinnamoyl-COA esters into the corresponding aldehydes (Goffner *et al.*,1994).The studies demonstrate that entry point enzymes serve as good targets for the directed manipulation of lignin content.

Hormonal Control: There is scarce study regarding hormonal control of lignification till date. Although ,auxin and cytokinin have been shown to control the growth and differentiation of cambial cells (Uggla *et al*.,1998; Mähönen *et al*., 2006) and the transdifferentiation of *Z. elegans* mesophyll cells into tracheary elements (Fukuda and Komamine, 1980; 1982),but the mechanism is still not clear and their influence on secondary cell wall synthesis or lignification is not yet demonstrated. However, it has been shown that auxin and gibberellins (GA) have opposite effects on the quantity of G and S units in lignin in the phloem fibres of *Coleus blumei* (Benth.) (Aloni *et al*., 1990). A role for GA in lignin biosynthesis is supported in other studies as well. For example, transgenic hybrid aspen with upregulated GA20oxidase, a GA biosynthetic enzyme, contained more S lignin and the expression of F5H, COMT and SAD was upregulated. Increased growth and fibre length were also observed (Eriksson *et al*., 2000; Israelsson *et al*., 2003). Similarly, in *Arabidopsis* that overexpressed the GA20oxidasegene, biomass production was up-regulated and the stems contained more lignin due to the induction of monolignol biosynthetic genes (Biemelt *et al*., 2004).

Up-regulation of jasmonic acid and ethylene-related pathways in *Arabidopsis* resulted in aberrant lignification without concomitant secondary cell wall synthesis and ectopic lignification (Schumacher *et al*., 1999; Caño-Delgado *et al*., 2000, 2003; Zhong *et al*., 2000; Ellis *et al*., 2002). It was suggested that the impaired cell wall structure triggers defence responses via jasmonate and ethylene-dependent pathways, leading to ectopic lignification through stress-induced gene family members (Ellis *et al*., 2002; Caño-Delgado *et al*., 2003).

Transcriptional Control:

Control at the transcriptional level is most important in the regulation of lignification, where transcriptional activation/repression plays a key role. The co-ordinated mechanism of control is brought about by the interaction between the highly conserved cis-elements called AC elements found in the promoters of many lignin biosynthetic genes and transcriptional regulators which act as trans- activators/repressors during lignification.

The xylem expression of bean PAL genes was shown to depend on AC-rich elements in their promoters (Leyva *et al*., 1992). These elements have been observed in the promoters of many monolignol biosynthetic genes in several species (Hauffe *et al*., 1991; Ye *et al*., 1994; Feuillet *et al*., 1995; Lacombe *et al*., 2000; Raes *et al*., 2003). AC elements are recognized by MYB type transcription factors, generally belonging to the R2R3-type, where R2 and R3 represent a characteristic N-terminal sequence repeat responsible for DNA binding (Sablowski *et al*., 1994; Patzlaff *et al*., 2003).

AC elements are also recognised by zinc finger transcription factors that contain a LIM domain. NtLIM1 from tobacco was shown to be a positive regulator of lignin biosynthesis, as antisense expression in tobacco decreased lignin content and repressed the expression of PAL, CL and CAD genes (Kawaoka *et al*., 2000).The other AC element-binding transcription factor, ACBF, was isolated but has not been investigated for its possible role in regulation of lignin biosynthesis (Séguin *et al*., 1997).

KNOX type homeodomain transcription factors BREVIPEDICELLUS (BP) from *Arabidopsis* (Mele *et al*., 2003) or ARBORKNOX1 (ARK1) from *Populus* (Groover *et al*., 2006) are another group of regulators which have been shown to affect lignin biosynthesis.

A role in the regulation of lignin biosynthesis has also been proposed for *Arabidopsis* MADS box transcription factors SHATTERPROOF (SHP) 1 and 2, specifically in the dehiscence zone of seed pods (Liljegren *et al*., 2000), and NAC domain transcription factors SECONDARYWALL ASSOCIATED NACDOMAIN PROTEIN 1 (SND1) and NAC SECONDARY WALL THICKENING PROMOTING FACTOR (NST) 1 and 2 in the interfascicular fibres (Zhong *et al*., 2006; Mitsuda *et al*., 2007). The defects must be partly due to the uninitiated or aberrant cell differentiation, but a role as a master regulator of the whole process of differentiation and formation of lignified secondary cell walls cannot be excluded.

1.3.2 Transcription factors regulating lignification and secondary wall biosynthesis

Recent studies on the transcriptional regulation of secondary wall biosynthesis in *Arabidopsis* have revealed that the transcriptional regulation of lignin biosynthesis is under the control of the common transcriptional network activating the entire secondary wall biosynthetic program (Fig. 1.5).

Characterization of the functions of transcription factors is quite complicated attributed to their functional redundancy, which is overcome by the use of chimeric repressor silencing technology (Hiratsu *et al.* 2004).Other supportive data are obtained using over-expression, knock-out mutants and trans-activation studies in plants.

Studies have revealed that, the transcriptional network involves a regulatory cascade of upstream transcription factors that control the formation of secondary walls by activating a range of other transcription factors. Some of these downstream transcription factors are then able to induce the expression of genes of the lignin biosynthetic pathway (Zhong and Ye, 2007).

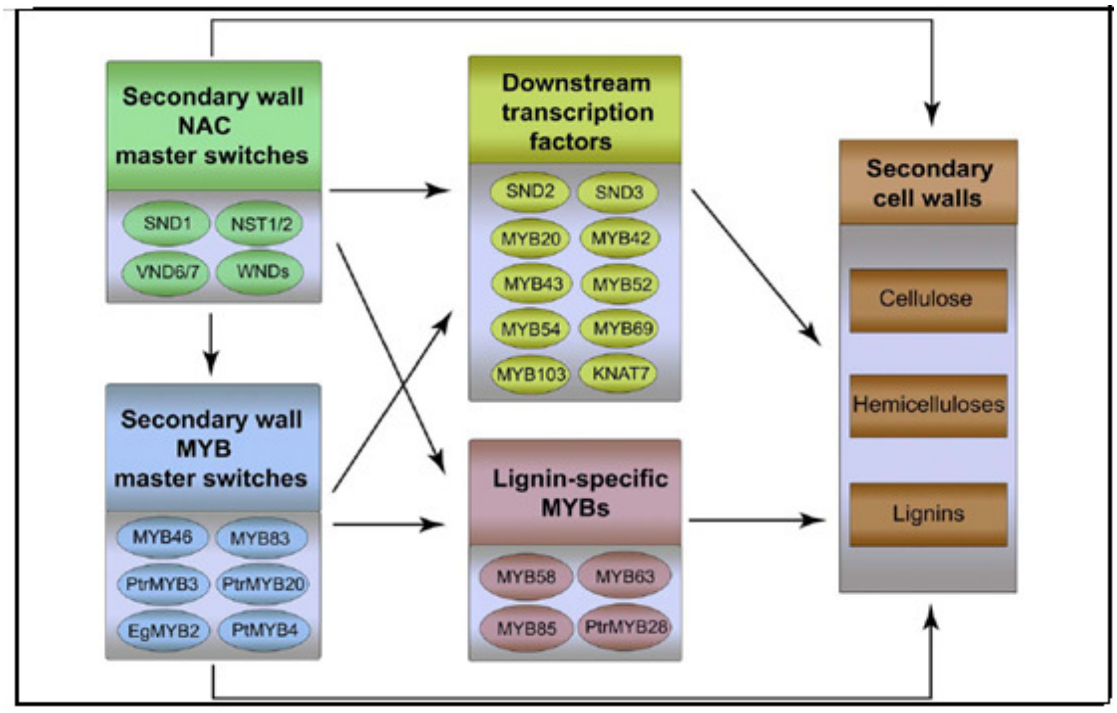


Figure 1.5: Transcriptional regulatory network controlling secondary wall biosynthesis (Zhong *et al*., 2010)

In this network, the NAC domain transcription factors, SND1 and its close homologs, NST1, NST2, VND6 and VND7, act as master switches leading to the activation of the biosynthetic genes of cellulose, xylan and lignin (Zhong and Ye 2007; Demura and Fukuda 2007; Yang *et al*. 2007; Zhong *et al*., 2006, 2007, 2008; Mitsuda *et al*.,2007; Yamaguchi *et al*.,2008). Simultaneous knockout of SND1 and NST1 completely blocks secondary wall thickening and lignin deposition in fibers, indicating that the biosynthesis of lignin together with cellulose and xylan is under the transcriptional control of the same master switches SND1 and NST1. SND1 and its homologs regulate a number of downstream transcription factors involved in secondary wall biosynthesis (Zhong *et al.,* 2008). Among them, MYB46, SND3, MYB103 and KNAT7 were found to be SND1 direct targets and MYB46 acts as another level of master switch able to activate the entire secondary wall biosynthetic program (Zhong *et al*., 2007). Therefore, SND1 and MYB46 function at the top of the transcriptional network leading to regulation of biosynthesis of lignin together with cellulose and xylan. The finding that the direct transcriptional activators of lignin biosynthesis, MYB58 and MYB63, are downstream targets of SND1 and MYB46 (Zhou *et al*.,2009), further

indicates that transcription factors directly regulating lignin biosynthesis are part of the SND1- and MYB46-mediated transcriptional network regulating secondary wall biosynthesis.

The regulatory cascade also explains why several of the currently described transcription factors lead to enhanced or reduced lignification when mis-expressed in plants while they do not directly regulate the lignin biosynthetic genes by binding to their promoters (Zhong *et al*., 2006, 2008).

1.3.3 Cis-elements involved in transcriptional regulation of lignin biosynthesis enzymes

Characterization of promoter regions of most of the lignin biosynthetic pathway genes has revealed specific regions responsible for their expression in lignifying tissues. Among the lignin gene promoters analyzed, the promoter of phenylalanine ammonia lyase (PAL) that catalyzes the first step of phenylpropanoid pathway has been best characterized. It was shown that multiple regions of the bean *PAL2* promoter were able to drive the GUS reporter gene expression in lignifying xylem tissues in transgenic tobacco and that another region was involved in suppression of expression in phloem (Leyva *et al*., 1992). Detailed analyses of the *PAL2* promoter by footprinting and electrophoretic mobility shift assay identified three AC elements together with a G-box involved in xylem-specific expression (Hatton *et al*., 1995). AC elements are called so, due to the fact that they are rich in adenosine and cytosine residues. AC elements are also known as H-boxes, as they lack G residues. Mutations of either the AC-I (ACCTACC) or AC-II (ACCAACC) element resulted in a decrease of xylem- associated expression but a gain of phloem-associated expression, whereas simultaneous mutations of both the AC-I and AC-II elements led to a complete loss of xylem-specific expression. Mutation of the AC-III (ACCTAAC) element caused a slight decrease of xylem-associated expression. It was concluded that the combined activity of AC elements determines xylem-specific expression (Hatton *et al*., 1995). The AC elements, also known as PAL-box and H-box, were first identified in the promoter of parsley PAL1 as one of the two motifs involved in the response to UV light and elicitor treatments (Lois *et al*., 1995). The hypothesis that AC elements are required and sufficient for directing xylem-specific expression was

further supported by the finding that the AC-II heptamer linked with the cauliflower mosaic virus 35S minimal promoter was able to drive the specific expression of the GUS reporter gene in xylem (Séguin *et al*., 1997).

The AC elements found in *PAL* promoters have also been identified in the promoters of other genes encoding enzymes implicated in lignin biosynthesis. For example, the gene that encodes C4H has both AC-I and AC-II elements present in its promoter (Mizutani *et al*., 1997). Not only does the Arabidopsis *C4H* promoter contain *cis* -acting elements that are conserved with the *PAL* genes, but the *C4H* promoter was also able to direct very high gene expression in differentiating xylem cells (Bell-Lelong *et al* 1997). Again, it is likely that this activity is related to the requirement for C4H in cells that are synthesising lignin precursors. Similarly, two AC elements have been implicated in xylem-specific expression directed by the promoter of the parsley 4-CL gene (Hauffe *et al*., 1991;1993).They were also found to be present on the promoter of the gene encoding *CCoAOMT* (Grimmig & Matern, 1997). In a key study assessing the role of the AC elements in the transcriptional control of lignin biosynthetic genes, analysis of the promoter of the *Eucalyptus gunnii* gene encoding CCR determined that AC elements were necessary and sufficient to drive xylem-localised gene expression (Lacombe *et al*., 2000). Analysis of the promoter of the gene encoding the monolignol-synthesising isoform of *CAD2* from *Eucalyptus gunnii* also uncovered a role for AC elements in the regulation of this gene (Lauvergeat *et al*., 2002).

A combination of bioinformatic analysis coupled with detailed analysis of the transcript abundance of all the postulated lignin biosynthetic genes in *Arabidopsis* identified the 'molecular toolbox' required for lignin biosynthesis in the plant (Raes *et al*., 2003). The study showed the presence of atleast one AC element within 1000 bp of 5' non-coding sequence upstream of the start codon of every lignin biosynthetic gene. The stringent matrix-based approach to identify these motifs showed that the occurrence of these elements in lignin biosynthetic genes was not merely due to chance, but was likely to have functional significance .Together, these analyses indicate that conserved AC elements serve as common cis regulatory elements and may be regulated by common transcriptional regulators driving the coordinated expression of lignin biosynthetic genes in lignifying tissues. The candidate transcription factors known to interact with these

elements have been already mentioned in section 1.3.1 under transcriptional control. R2R3-MYB transcription factors are the potential ones, known to regulate the lignin biosynthetic genes by interacting with AC elements. They are also known to regulate lignification without directly interacting with these elements. The role of R2R3-type MYBs in lignification has been discussed in Section 1.4.3.

1.4 MYB transcription factors in plants

Transcription factors are proteins that recognize DNA in a sequence-specific manner and regulate the frequency of initiation of transcription upon binding to specific sites in the promoter of target genes. They can be activators, repressors or both and contain related structural motifs for recognition, namely the DNA-binding domain. Based on the similarities in the DNA-binding domain, transcription factors have been classified into families: helix-turn-helix (HTH) proteins, the homeodomains, zinc finger proteins, the steroid receptors, leucine zipper proteins, and the helix-loop-helix proteins (Pabo and Sauer, 1992). MYB transcription factors which belong to the HTH family, comprise one of the largest families in plants (Romero *et al*., 1998; Riechmann *et al*., 2000).

1.4.1 Characteristics and classification of Myb transcription factors

Discovery

The Myb gene was first identified in the form of the v-Myb oncogene of the avian myeloblastosis virus (Klempnauer *et al*., 1982).Subsequently, members of the Myb gene family were found existing widely in both plants and animals (Rosinski and Atchley, 1998).A small number of Myb genes have also been found in fungi (Lipsick, 1996), mycetozoa (Braun and Grotewold, 1999; Kranz *et al*.,2000), and microsporidia(Jiang *et al*.,2003). There are no Myb genes recognized in prokaryotes so far.

Structural features

MYB transcription factors can be structurally dissected into a highly conserved N-terminal DNA-binding domain and a C-terminal transcriptional activation domain. The DNA-binding domain also known as the MYB domain consists of a region of 50-53 amino acids with constantly spaced tryptophan residues that binds to DNA in a sequence-specific manner (Lipsick, 1996). Multiple copies (up to four) of the non-identical Myb

domains are frequently present as tandem repeats within a single protein (termed R0,R1, R2 and R3). Interestingly, each of the Myb repeats is more closely related to other members of the same family than to other repeats within the same protein (Lipsick, 1996).The second and third helices of each repeat build a helix–turn–helix (HTH) structure with three regularly spaced tryptophan (or hydrophobic) residues, forming a hydrophobic core in the 3D HTH structure (Ogata *et al.*,1996).The third helix of each repeat is the ‘‘recognition helix’’ that makes direct contact with DNA and intercalates in the major groove (Robinowicz *et al.*,1999). During DNA contact, two MYB repeats are closely packed in the major groove, so that the two recognition helices bind cooperatively to the specific DNA sequence motif.

Classification

MYB proteins can be divided into different classes depending on the number of adjacent imperfect repeats (one, two, three or four) found in the DNA-binding domain. The three repeats of the prototypic MYB protein c-Myb, known to regulate proliferation and differentiation of hematopoietic cells in vertebrates (Duprey and Boettiher, 1985) are referred to as R1, R2 and R3.The repeats from other MYB proteins are named according to their similarity to R1, R2 or R3 of c-Myb. Plants, represent the taxon with the highest diversity of MYB proteins as all the four classes of MYB proteins are found in them i.e.**1R, R2R3, 3R** and **4R type** (Fig 1.6)

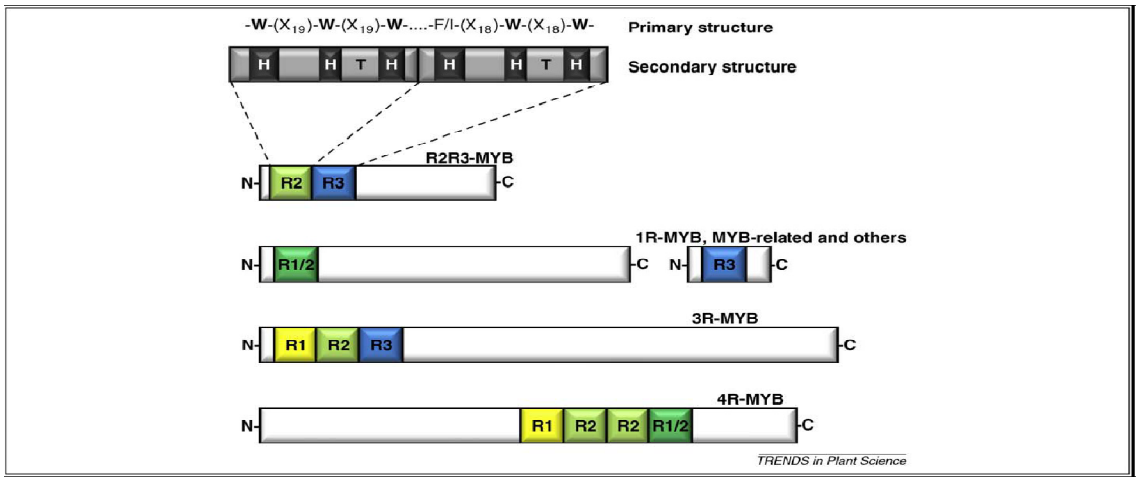


Figure 1.6: Plant MYB transcription factor classes. Illustration showing different plant MYB protein classes, depending on the number of adjacent MYB repeats (R). The primary and secondary structures of a typical R2R3-MYB are indicated. **H**, helix; **T**, turn; **W**, tryptophan; **X**, amino acid (X) (Dubos *et al.*, 2010)

4R-MYB group is the smallest class, whose members contain four R1/R2-like repeats. A single 4R-MYBprotein is encoded in several plant genomes. Little more is known of these proteins in plants. The second class contains R1R2R3-type MYB (**3R-MYB**) proteins, typically encoded by five genes in higher plant genomes. Genes encoding 3R-MYB proteins have been found in most eukaryotic genomes, so they represent a conserved gene class with roles, albeit divergent, in cell cycle control (Ito, 2005; Haga *et al*., 2007). This indicates a functional conservation among 3R-MYBgenes from plants to humans. The third heterogeneous class comprises proteins with a single or a partial MYB repeat, collectively designated "**Myb-related** or "Myb –like proteins". These proteins are fairly divergent and include factors that bind the consensus sequence of plant telomeric DNA (TTTAGGG) (Yu *et al*., 2000). It has also been shown that **1RMYB** factors (e.g. MYBST1 or StMYB1R1) can act as transcriptional activators (Baranowskij *et al*.,1994) and some are associated closely with the activity of the circadian clock (e.g. CCA1 [CIRCADIAN CLOCK ASSOCIATED 1] and LHY [LATE ELONGATED HYPOCOTYL]) (Schaffer *et al*., 2001).

MYB genes containing two repeats (i.e. **R2R3-MYB**) constitute the largest *MYB* gene family in plants, which are thought to have evolved from an R1R2R3-MYB gene ancestor, by the loss of the sequences encoding the R1 repeat and subsequent expansion of the gene family (Rosinski and Atchley, 1998). R2R3-MYB transcription factors have a modular structure, with an N- terminal DNA-binding domain (the MYB domain) and an activation or repression domain usually located at the C terminus. In contrast to the highly conserved MYB domain, the other regions of R2R3-MYB proteins are highly variable. Based on the conservation of the DNA binding domain and of amino acid motifs in the C terminal domains, R2R3-MYB proteins have been divided into 22 subgroups (Kranz *et al*., 1998; Stracke *et al*., 2001). Most of these subgroups, defined first for the proteins of *Arabidopsis*, are also present, and are sometimes expanded, in other angiosperms .The expansion of the R2R3-MYBgene family in plants (Table 1.1), which typically contain more than 100 members (Dubos *et al*., 2010) fits well with the observation that many (if not all) R2R3-MYB transcription factors play central roles in plant-specific processes (Martin and Paz, 1997).

Table 1.1: Numbers of the members in the four different MYB classes; n.d., not determined (Douglas *et al.,* 2010)

MYB protein classes	Eudicot			Monocot
	A.thaiana	*Populus trichocarpa*	*V.vinifera*	*O.sativa*
R2R2MYB	**126**	**192**	**108**	**109**
1R-MYB, MYB-related	**64**	**n.d.**	**n.d.**	**70**
3R-MYB	**5**	**5**	**5**	**5**
4R-MYB	**1**	**-**	**1**	**1**

1.4.2 Multifunction of plant MYB transcription factors in plants

In higher plants, numerous R2R3-MYB proteins have been characterized by genetic approaches and found to be involved in the control of a wide variety of plant-specific processes including

(i) cell fate and identity, (ii) developmental processes and (iii) responses to biotic and abiotic stresses(iv) primary and secondary metabolism,

(i) Control of cell fate and identity

The control of cell morphogenesis and pattern formation, and the enormous improvement in this study is attributed to abundant researches of *Arabidopsis* root hair and epidermis. Trichome development is a well characterized model for the study of plant cell differentiation. Extensive molecular studies have revealed that a complex of transcription factors is involved in the determination of trichome fate. The determination of epidermal cell type involves several R2R3-MYB proteins encoded by *AtMYB*0/GL1, *AtMYB*23 and *AtMYB*66/WER (subgroup 15).*AtMYB*0 and *AtMYB*23 control trichome initiation in shoots, *AtMYB*66 controls root hair patterning. In roots, *AtMYB*23 is positively regulated by *AtMYB*66 and participates in a positive feedback loop to reinforce the cell fate establishment process (Kang *et al*., 2009) AtMYB23 regulates trichome extension and branching in combination with AtMYB5 (Kirik *et al*., 2005; Li *et al*., 2009). AtMYB5 also regulates outer seed coat differentiation. Stomatal differentiation and patterning are strictly coordinated in time and space. Two closely related R2R3-MYBs, AtMYB88 and AtMYB124/FLP, act in this process by restricting divisions late in the stomatal cell lineage and inducing terminal differentiation (Lai *et al*., 2005), by regulating genes involved directly in cell-cycle progression (Xie *et al*., 2010). AtMYB98 regulates

synergid cell differentiation during female gametophyte development, pollen-tube guidance and the formation of the filiform apparatus (Punwani *et al*., 2008).

(ii) Regulation of plant development

Various developmental processes like anther development, axillary meristem formation, inflorescence development, embryogenesis and root development are known to be regulated by R2R3 Myb proteins in combination with other transcription factors. In *Arabidopsis*, AtMYB21, AtMYB24 (subgroup 19), AtMYB57, AtMYB108/BOS1, AtMYB35/TDF1, AtMYB80 (formerly AtMYB103) and AtMYB99 (Cheng *et al*., 2009; Mandaokar *et al*., 2009) are shown to control anther development and functionality.AtMYB37/RAX1,AtMYB38/RAX2/BIT1 and AtMYB84/ RAX3 (subgroup 14) are partially redundant regulators of axillary meristem formation. (Muller *et al*., 2006; Keller *et al*., 2006).AtMYB38 and AtMYB18/LAF1 (subgroup 16) regulate hypocotyl elongation in response to blue (Hong *et al*., 2008) and far-red light (Yang *et al*., 2009), respectively. AtMYB115 and AtMYB118/PGA37 (subgroup 25)have been proposed to play roles in embryogenesis (Wang *et al*., 2008). AtMYB59 regulates root development through the control of cell-cycle progression at the root tips (Mu *et al*., 2009) and AtMYB77 (subgroup 22) regulates lateral root formation by modulating the expression of auxin-inducible genes (Shin *et al*., 2007).

In other plants, evidence of tissue -specific developmental processes is reported as well, such as *GhMYB109* is specifically expressed in cotton fiber initial cells as well as elongating fibers (Suo *et al*., 2003). *GhMYB7/9* are expressed in flowers and fibers, and their expression in fibers is developmentally regulated (Chuan *et al*., 2005). Furthermore, some studies have suggested that GAMYB may be involved in floral initiation, stem elongation, anther development, and seed development (Woodger *et al*., 2003).

(iii) Responses to biotic and abiotic stresses

Plants respond to environmental changes with a number of physiological and developmental changes to tolerate stresses. Drought stress is one such condition, and it affects almost all plant functions including growth and development. AtMYB60 and AtMYB96 act through the ABA signaling cascade to regulate stomatal movement

(Cominelli *et al*., 2005), drought stress and disease resistance (Seo *et al*., 2009; Seo and Park, 2010), respectively. Three members of subgroup 20 are implicated in stress responses: AtMYB2 controls the ABA induction of salt and dehydration responsive genes (Abe *et al*., 2003), AtMYB62 is involved in the response to phosphate starvation (Devaiah *et al*., 2009) and AtMYB108 in both biotic and abiotic stress responses (Mengiste *et al*., 2003). *BcMYB1* isolated from *Boea crassifolia* was strongly induced by drought stress as well (Chen *et al*., 2005).

Many MYB proteins have been reported to display light inducible expression. For example, the expression of *C1* was inhibited by far red light, suggesting a possible involvement of phytochrome (Paz *et al*., 1987). The AtMYB102 and AtMYB21 genes isolated from *Arabidopsis* were induced by light (Quaedvlieg *et al*., 1996; Li *et al*., 2006). Expression of MYB_p1 increased 10-fold in the red relative to the green form of *Perilla frutescens*, which indicates that this gene also is induced by light (Gong *et al*., 2004).

Recently, several cold-inducible genes were reported as well. For example, the rice *OsMYB4* gene encodes a MYB transcription factor involved in cold acclimation whose constitutive expression in *Arabidopsis* resulted in improved cold and freezing tolerance (Vannini *et al*., 2004) and the *Arabidopsis HOS10* gene also is reported to be essential for cold acclimation and may affect dehydration stress tolerance in plants by controlling stress induced ABA biosynthesis (Zhu *et al*., 2005). AtMYB68 is modulated by temperature, and loss of AtMYB68 reduces the ability of *myb68* plants to compensate their growth at higher temperatures (Feng *et al*., 2004).

AtMYB30 encodes an activator of the hypersensitive cell death program in response to pathogen attack, acting through the regulation of very-long-chain fatty acids synthesis. In seedlings, AtMYB30 has also been shown to act in the brassinosteroid pathway controlling hypocotyl cell elongation (Li *et al*., 2009; Raffaele *et al*., 2008). AtMYB102/AtM4 and AtMYB41 (subgroup 11) contribute to plant resistance against insects and probably affect dehydration after wounding (De Vos *et al*., 2006) and osmotic stress responses (Cominelli *et al*., 2008; Lippold *et al*., 2009), respectively.

(iv) Regulation of primary and secondary metabolism

Several R2R3-MYBs are involved in the regulation of different branches of the primary metabolism and secondary metabolism, mainly the phenylpropanoid pathway, which in turn leads to the production of flavonoids and lignin.

*ATR*1 gene from Arabidopsis *thaliana* was the first reported to be an activator of tryptophan gene synthesis (Bender and Fink, 1998). The R2R3-MYB proteins of subgroup 12 regulate glucosinolate biosynthesis and AtMYB28/HAG1/PMG1, AtMYB29/HAG3/PMG2 and AtMYB76/HAG2 regulate the biosynthesis of aliphatic glucosinolates in aerial tissues (Gigolashvili *et al*., 2008; Gigolashvili *et al.*, 2007). However, AtMYB34/ATR1, AtMYB51/HIG1 and AtMYB122 regulate the production of indolic glucosinolates in roots and late stage rosette leaves (Gigolashvili *et al*., 2007).

Phenylpropanoid metabolism is one of the three main types of secondary metabolism involving modification of compounds derived initially from phenylalanine, which is now well understood. As the first step, phenylalanine is deaminated to yield cinnamic acid by the action of phenylalanine ammonia lyase (PAL). Cinnamic acid is hydroxylated by C4H to 4-coumaric acid, which is then activated to 4-coumaroyl-coenzyme A (CoA) by the action of 4-CL.Then it is divided into two major pathways—the flavonoid biosynthesis pathway and the lignin biosynthetic pathway. To date most of the R2R3-MYB genes are known to regulate the phenyl propanoid pathway.

The flavonoid biosynthetic pathway has now been almost completely elucidated. Many R2R3-MYB TFs have been identified from several model plants, such as maize, *Antirrhinum*, tobacco, *Petunia*, and *Arabidopsis*, which are involved in the regulation of different branches of flavonoid biosynthesis metabolism. For instance, the *Antirrhinum majus* genes AmMYB305 and AmMYB340 both can activate the gene encoding the first enzyme of phenylpropanoid metabolism, PAL, and also could activate other two enzymes of flavonol synthesis, CHI and F3H (Moyano *et al*., 1996). The sorghum MYB protein *y1* gene could regulate the expression of *CHS*, *CHI*, and *DFR* genes that are required for biosynthesis of deoxyflavonoids (Duthie *et al*., 2000). AtMYB11, AtMYB12 and AtMYB111 (subgroup 7) control flavonol biosynthesis in all tissues (Stracke *et al*., 2007), AtMYB75/PAP1, AtMYB90/PAP2, AtMYB113 and AtMYB114 (subgroup 6) control anthocyanin biosynthesis in vegetative tissues (Gonzalez *et al*., 2008) and

AtMYB123/TT2 (subgroup 2) control the biosynthesis of proanthocyanidins (PAs) in the seed coat of *Arabidopsis* (Lepiniec *et al*., 2006)

Anthocyanins are pigments derived from a specialized branch of phenylpropanoid metabolism. The first characterized plant R2R3-MYB was *C1* from maize, which regulates genes encoding enzymes of the anthocyanin biosynthetic pathway (Paz-Ares *et al*, 1987). Similar to *C1*, some R2R3-MYB family members control anthocyanin biosynthesis, albeit in a different developmental context (Cocciolone *et al*., 2001). For instance, over-expression of Kyoho grape *MYBA* gene leads to the induction of reddish-purple spots and *UFGT* gene expression in non-colored embryos, which indicated that this gene was involved in the regulation of anthocyanin biosynthesis in the grape via expression of the *UFGT* gene (Kobayashi *et al*., 2002). The strong correlation between the expression of the apple *MdMYB10* gene and apple anthocyanin levels during fruit development suggests that this transcription factor is responsible for controlling anthocyanin biosynthesis in apple fruit in the red fleshed cultivar and in the skin of other varieties.

Another important branch of lignin biosynthetic pathway is also known to be regulated by R2R3-MYB transcription factors in various plants. Recent studies have elucidated their roles as both activators and repressors of the pathway, which have been broadly discussed in the following section 1.4.3.

1.4.3 Role in control of monolignol pathway and lignin biosynthesis

Lignin biosynthesis involves the phenylpropanoid pathway, which converts phenylalanine to *p*-coumaroyl-CoA, the precursor of a wide range of phenolic compounds. The subsequent hydroxylation and ethylation steps have recently been shown to occur at the level of hydroxyl-cinnamic acid esters and their corresponding aldehydes and/or alcohols (Humphreys and Chapple, 2002). In the last decade several R2R3-MYB genes belonging to various subgroups have been isolated and characterized to be regulators of lignification.

The first line of genetic evidence on the possible involvement of MYBs in the regulation of lignin biosynthesis came from the study of two MYB proteins, AmMYB308 and AmMYB330, from *Antirrhinum*, which were overexpressed in *Nicotiana tobacum*.

Both the genes belong to subgroup 4 and acted as repressors of the pathway genes (Tamagnone *et al*., 1998). Subsequently, AtMYB4 an *A.thaliana* ortholog of AmMYB308, was shown to negatively regulate C4H gene and positively regulate CCoAOMT gene in *A.thaliana* plants mutant for *At*MYB4 (Jin *et al*. 2000). Since then several MYBs from *Arabidopsis* have been shown to alter the expression of lignin biosynthesis when overexpressed, *e.g. At*MYB58, *At*MYB63 (subgroup3) and AtMYB85 activate lignin biosynthesis in fibers and/or vessels (Zhou *et al*., 2009; Zhong *et al*, 2008), whereas AtMYB68 negatively regulates lignin deposition in roots (Feng *et al*., 2004). AtMYB46 is a positive regulator of lignin biosynthesis in fibers and vessels and also regulates cellulose and xylan deposition (Zhong *et al*., 2007). AtMYB26/ MS35 controls secondary wall deposition in anthers (Yang *et al*., 2007). AtMYB52, AtMYB54 and AtMYB69 (subgroup21) and AtMYB103 are positive regulators dedicated to cell wall thickening in fiber cells.AtMYB52, AtMYB54 and AtMYB69 are proposed to regulate lignin, xylan and cellulose biosynthesis, and AtMYB103, cellulose biosynthesis (Zhong *et al*, 2008). AtMYB61 (subgroup 13) plays a pleiotropic role including influencing lignin deposition (Newman *et al*., 2004).

Several MYBs from *Pinus* and *Eucalyptus* have been shown to be strong candidates as regulators of lignin biosynthesis (Table1.2). The pine PtMYB1 and PtMYB4 (Patzlaff *et al*.,2003 a,b) and the *Eucalyptus* EgMYB2 (Goicoechea *et al*., 2005) bind to the AC elements and their genes are expressed in developing wood that undergoes secondary wall thickening and lignin biosynthesis. When over expressed in tobacco plants, PtMYB4 and EgMYB2 induced the expression of some of the lignin biosynthetic genes and led to ectopic lignin deposition or increased secondary wall thickening. Phylogenetic and experimental findings indicate that EgMYB2 and perhaps also PtMYB4 are orthologs of AtMYB46 and they regulate the entire secondary wall biosynthetic program during wood formation. Other R2R3-MYBs like *Pt*MYB1 and PtMYB8 were also studied using transformants of another gymnosperm *Picea glauca* (spruce) (Bomal *et al*., 2008). Their over expression in *P. glauca* caused ectopic secondary wall deposition and increased lignin content. On the other hand, EgMYB1, another R2R3-MYB from Eucalyptus, phylogenetically related to subgroup4 was preferentially expressed in the differentiating xylem tissues in stems and roots but repressed the transcription of CAD

and CCR gene (Legay *et al*., 2007). Hence it may be a negative regulator of lignin biosynthesis.

In the hybrid aspen *P. tremula* x *P. tremuloides*, an R2R3-MYB designated as PttMYB21a was expressed strongly in xylem tissues. Transgenic aspen plants expressing the MYB gene in an antisense orientation showed higher transcript levels of CCoAOMT in the phloem, suggesting that PttMYB21a might work as a repressor (Karpinska *et al*. 2004).

One of the orthologs of AtMYB58 and AtMYB63 in Poplar, PtrMYB28, is predominantly expressed in developing wood undergoing secondary wall thickening and lignification and known to induce the expression of 4-CL gene (Zhong *et al*., 2009).

An R2R3-MYB family member from *Vitis vinifera* cv. Cabernet Sauvignon (grape), VvMYB5a, was expressed in skin, flesh, and seeds, mainly during the early stages of berry development and its overexpression in *N. tabacum* resulted in decreased expression of CCoAOMT in stamens (Deluc *et al*. 2006).

Fornale´ *et al*. (2006) characterized two R2R3-MYB transcription factors ZmMYB31 and ZmMYB42 from *Zea mays* (maize).Their overexpression phenotypes in *A.thaliana* showed altered expression of monolignol pathway genes and reduced lignin content. (Table 1.2)

Table 1.2 MYB transcription factors with a role in regulation of lignin biosynthetic genes

↑: Upregulated, ↓: Downregulated or = : not affected target genes

MYB gene	Effect of gene over expression on monolignol pathway genes										Method of determinat-ion	Reference
	PAL	*C4H*	*4CL*	*HCT*	*C3H*	*CCoAOMT*	*CCR*	*F5H*	*COMT*	*CAD*		
Antirrhinum *Am*MYB330			↓								RNA blot	Tamagnone *et al*.,1998
*Am*MYB308	=	↓	↓							↓	Y1H, RNA blot	Tamagnone *et al*.,1998
Arabidopsis *At*MYB*4*	=	↓	↓			↑		=	=		qPCR, TEA	Jin *et al.*, 2000
*At*MYB*46*	↑	↑	↑		↑	↑				↑	qPCR, TEA	Zhong *et al*., 2007; Ko *et* al., 2009
*At*MYB58	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	qPCR, TEA, EMSA	Zhou *et al*., 2009
*At*MYB63	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	qPCR,TEA, EMSA	Zhou *et al*., 2009
Eucalyptus *Eg*MYB1							↓			↓	TEA, EMSA	Legay *et al*., 2007
*Eg*MYB2	=	=	↑	↑	↑	↑	↑	↑	↑	↑	qPCR,EMSA	Goichoeche a *et al*., 2005
Pinus *Pt*MYB4	↓	=	=		↑	↑	↑		↑	↑	RNA blot, EMSA	Patzaff et al., 2003
*Pt*MYB1	↑	↑	↑		↑	↑	↑		↑↓	↑	qPCR	Bomal *et al*., 2008
*Pt*MYB8	↑	↑	↑		↑	=	↑		↑	↑	qPCR	Bomal *et al*., 2008
Poplar *Ptt*MYB21	=	=	=			↓			=	=	qPCR	Karpinska *et al*., 2004
*Ptr*MYB3			↑			↑					qPCR, TEA	McCarthy *et al*., 2010
*Ptr*MYB20			↑			↑					qPCR, TEA	McCarthy *et al*., 2010
Zea mays *Zm*MYB31		=	↓		↓	=		↓	↓	↑	qPCR, EMSA	Fornale *et al.*, 2006;2010
*Zm*MYB42		↓	↓			=			↓	↓	qPCR	Fornale *et al*., 2006

qPCR refers to quantitative RT-PCR results from over expressing lines**; EMSA,** Electrophoretic mobility shift assay with AC-promoter elements; **TEA,** Protoplast transient expression assay;**Y1H**, Yeast one-hybrid assay

1.5 Rationale of the thesis

Leucaena leucocephala is an economically important tree species, largely utilized in India for its pulp in the paper industry as well as a source of fodder. Currently there is a great interest in modifying the content and/or the composition of lignin in the plant as it constitutes a negative value for the end-uses of its wood for either of the two applications.

The biosynthetic pathway of lignin and genetic modification of the pathway enzymes has been well defined in various plant species (Li *et al*., 2008; Vanholme *et al*., 2008).However, engineering pathway enzymes have resulted in some or the other adverse effects on plant development due to alteration in the metabolic flux. Even in case of *Leucaena,* antisense down regulation of a few pathway enzymes resulted either in severely stunted growth or poorly developed plant stems thus, being detrimental to normal plant development. Hence, understanding the transcriptional regulation of lignin biosynthesis genes would have important implications in tree biotechnology. Studies in the last few decades have established that the expression of the pathway genes is coordinated by the interaction between common regulatory AC-elements found on their promoters with specific transcription factors. Emerging evidence indicates the role of R2R3-type MYB transcription factors in regulating the monolignol pathway enzymes, many of which are activators of lignin biosynthesis (Patzaff *et al*., 2003 a,b; Karpinska *et al*.,).However, MYB factors of subgroup 4 have been shown to act as repressors of the pathway genes(Tamagnone *et al*.,1998; Legay *et al*., 2007).These transcriptional activators or repressors can act as master-switches to alter lignification in the plant by channeling the metabolic flux through the pathway in an orchestrated manner, as they interact with common cis elements found on the promoters of these genes.

The present study envisages understanding the regulation of the lignin biosynthetic pathway genes in *Leucaena leucocephala*, by characterization of the AC rich -elements found on the promoters of few lignin biosynthetic genes (C4H, CCoAOMT, CCR and CAD).A novel R2R3-type MYB transcription factor gene named as *Ll*MYB_SSM, belonging to subgroup 4 R2R3-MYBs, was isolated, cloned and characterized from the plant, and a putative regulatory function of the gene was determined in lignification by bioinformatic, spatio-temporal and gel retardation studies.

CHAPTER 2

MATERIALS and METHODS

This chapter describes the biological strains, glass wares and chemicals utilized for the experiments. It also deals with the general laboratory techniques and protocols routinely followed during the course of work. Other important specific methodologies followed have been discussed separately in the respective chapters.

2.1 Plant material

2.1.1. *Leucaena leucocephala*

Leucaena leucocephala cultivar K-636 was used for the studies and it was a selection from Hawaii (Bray *et. al*. 1998). The cultivar produces erect boles suitable for timber production.

For genomic DNA isolation, young *Leucaena* leaves were harvested from garden growing plantlets and washed thoroughly with water, dried and then crushed with liquid nitrogen.

For RNA isolation, xylem tissue of mature healthy plants was harvested as and when required. Outer bark was scraped to expose the xylem tissue, which was ground using liquid nitrogen and RNA was isolated for normal cDNA preparations, RACE and RT PCR.

For obtaining sterile *Leucaena* seedlings for spatio-temporal studies, *Leucaena* pods containing seeds were obtained and seed sterilization was carried out. Potentially good seeds were selected and treated with conc. H_2SO_4 for 2-3 min and then washed extensively with tap water. The scarified seeds were surface sterilized with 0.1% (w/v) $HgCl_2$ for 10 min followed by five rinses with sterile deionized water. The seeds were soaked in sterile water for 24-48 h until the small radicals appeared and then transferred to ½ MS basal medium (Murashige and Skoog, 1962) supplemented with 2% sucrose and 1.5% glucose for germination. The medium was solidified with 0.8% agar. The pH of the medium was adjusted to 5.8 prior to autoclaving. The culture bottles were incubated at 25±2 °C and 60% relative humidity under 16h photoperiod and 70% relative humidity, with a light intensity of 24.4 µmol/m^2/s. The day of inoculation was considered as day zero. Root, shoot and leaves were harvested from 5, 10, 15 and 20 day old seedlings and used for further experiments.

2.1.2. *Nicotiana tabacum*

Tobacco seeds (*Nicotiana tabacum* var. Anand 119) were germinated on wet sterile paper towel. Germinated seeds were transferred to ½ MS basal medium (Murashige and Skoog, 1962) supplemented with 2% sucrose and 1.5% glucose for germination. The medium was solidified with 0.8% agar. The pH of the medium was adjusted to 5.8 prior to autoclaving. The culture bottles were incubated at 25±2 ºC and 70% relative humidity under 16 h photoperiod. One month old axenic cultured plant leaves were used as the explant for further transformation experiments.

2.2 Bacterial strains

Strain	**Genotype**
Escherichia coli XL-10	*endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tet^R F'[proAB lacI^{q}ZΔM15 Tn10(Tet^R Amy Cm^R)]* (Invitrogen, USA).
Escherichia coli XL-1 Blue	*RecA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proABlacIqZΔM15 Tn10 (Tetr)* (Stratagene, USA)
TOPO 10	*F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1araD139 Δ(araleu) 7697 galU galK rpsL (StrR) endA1 nupG*
E.coli BL21(DE3)	*F– ompT gal dcm lon hsdSB(rB – mB -) _(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])* (Invitrogen, USA)
Agrobacterium tumefaciens GV2260	*C58, Rifr, pGV2260 (pTiB6S3_T-DNA), Carbr, Octopine type*

2.3 Plasmid vectors used

Plasmid construct	**Important features (reference or source)**
pGEM-T Easy Vector	Cloning vector (Promega, USA)
pCAMBIA 1301	Binary cloning plant transformation vector (CAMBIA)
pCAMBIA 1381Z	Promoter cloning binary plant transformation vector(CAMBIA)
pET 28b(+)	Expression vector (Novagen)
pET 41a(+)	Expression vector carrying GST tag (Novagen)

2.4 Glassware

Glassware used in all the experiments were procured from "Borosil", India. Test tubes (25 mm x 150 mm), glass bottles (70 mm x 125 mm), petridishes (85 mm x 15 mm; 55 mm x 15 mm), conical flasks (100, 250 & 500 mL; 1, 2 & 5 L capacity) and pipettes (1, 2, 5, 10 and 25 mL capacity) were used during the course of study.

2.4.1 Preparation of Glassware

Glassware used for all the experiments were cleaned by boiling in a saturated solution of Sodium bicarbonate for 1 h followed by repeated washing in tap water. Thereafter, it was immersed in 30% HNO_3 solution for 30 min followed by repeated washing in tap water and rinsed with distilled water. Washed glassware was thereafter dried at room temperature. Test tubes and flasks were plugged with absorbent cotton (Mamta Surgical Cotton Industries Ltd., Rajasthan, India). Autoclaving of the glassware and above items was done at 121 ºC and 15 psi for 1 h.

2.5 Plastic ware

Sterile disposable filter sterilization units (0.22 µm) and petridishes (55 mm and 85 mm diameter) were procured from "Laxbro", India. Microfuge tubes (1.5 mL and 2 mL capacity), microtips (10, 200 and 1000 µL capacity) and PCR tubes (0.2 mL and 0.5 mL capacity) were obtained from "Tarsons" and "Axygen", India and EIA/RIA medium size flat bottom, 96 well plates from "Costar".

2.6 Chemicals

Tris, IPTG, X-gal, SDS, BSA, EDTA, PMSF, PIPES, DTT, PVPP, APS, Imidazole, Urea and Ethidium bromide were purchased from Sigma-Aldrich (USA), Bioworld (USA). Agarose, restriction enzymes, T4 DNA ligase, RNase A and lysozyme were obtained from NEB (USA), Promega (USA), Bioenzymes (USA) and Amersham (UK). Different kits were purchased from BD CLONETECH (JAPAN). Invitrogen (USA), Promega (USA) and Sigma-Aldrich (USA). Taq DNA polymerase was obtained from Sigma-Aldrich (USA) and Bangalore Genei (India). Plasmid vectors, pGEM-T Easy Vector, pET28b (+)and pET41a(+) were purchased from Promega (USA) and Novagen (USA)

respectively. Megaprime labeling kit and Hybond-N+ membrane were obtained from Amersham (UK). [α-^{32}P]-dCTP were obtained from Bhabha Atomic Research Centre (BARC), India. X-ray films were obtained from Konica (Japan) or Kodak (USA). All other chemicals and solvents of analytical grade were purchased from HiMedia, Qualigens Fine Chemicals and E-Merck Laboratories, India. All chemicals used in the tissue culture study were of analytical grade (AR) and were obtained from "Qualigens", "S.D. Fine Chemicals" or "HiMedia", India. The Sucrose, glucose and agar-agar were obtained from "Hi- Media". Bacto-Agar for microbial work was obtained from "DIFCO" laboratories, USA. Biotin-DNA labeling kit, Chemiluminescent detection kit and the EMSA kits were from "Pierce". The oligonucleotides/ primers were synthesized by "Eurofins India Limited" and the quantitative real time PCR master-mix and probe were obtained from "Eurogentec".

2.7 Equipments used for the study

S.No.	Equipment	Make
1	Balances	Contech/ Sartorious
2	Water bath	Fisher Scientific
3	Dry Bath	Eppendorf
4	Incubator	New Brunswick
5	Centrifuge	Sorvall/eppendorf
6	Gel Documentation system	BioRad
7	Thermo Cycler PCR machine	BioRad
8	Spectrophotometer	Perkin Elmer
9	Power pack	Bio-Rad
10	Agarose Gel Electrophoresis Units	Bangalore GeNei/ Bio-Rad
11	Protein Gel Electrophoresis Units	GE HealthCare
12	Sonicator	Misonix
13	pH-Meter	Digital corp.
14	Water purification system	Millipore Unit (Milli RO/ MilliQ)
15	Microwave oven	Electrolux
16	Fridge/ Deep freezer	Vestfrost/Godrej
17	Magnetic rotator	REMI
18	Laminar Air Flow	Microfilt India
19	iBlot Gel Transfer System	Invitrogen
20	Typhoon trio+	GE Healthcare

21	Icematic	Saksham Technologies
22	AKTA Explorer	GE Healthcare
23	ELISA Plate Reader	BioRad
24	Real Time PCR machine	Stratagene
25	Speed Vac concentrator	Eppendorf
26	Hybridisation incubator	SciGene
27	Light microscope	Axioplan2(Carl Zeiss)

2.8 Buffers and Solutions

2.8.1 Buffers and Solutions for DNA Electrophoresis

Name	Ingredients	Preparation and storage
50xTAE buffer	2 M Tris Acetic acid 0.05 M EDTA	pH was adjusted to 8.0 with NaOH and stored at room temperature
10xTBE buffer	90 mM Tris 90 mM Boric acid 2 mM EDTA	Stored at room temperature
DNA loading buffer	0.25 g Xylene cyanol 0.25 g Bromophenol blue 0.25 g Ficoll 400 1.46 g EDTA make up the volume to 100 ml with H_2O	The solutions were filter sterilized using 0.22 micron (μ) filter and stored at room temperature.

2.8.2 Buffers and Solutions for gDNA isolation, Southern blotting and hybridization

Name	Ingredients	Preparation and storage
Extraction buffer	100 mM Tris-HCl (pH 8.0) 20 mM Na EDTA (pH 8.0) 1.4 M NaCl 2.0% (w/v) CTAB Add β-mercaptoethanol to 0.2 % before use.	Room temperature
Depurination buffer	0.25 N HCL	Freshly prepared
Denaturation buffer	1.5 M NaCl 0.5 M NaOH	Room temperature

Neutralization buffer	1.5 M NaCl 1.0 M Tris HCl (pH 7.4)	Room temperature
20 X SSC	3 M NaCl 0.3 M Sodium citrate (pH 7.0)	Room temperature
Hybridization buffer	1% BSA 1.0 mM EDTA, pH 8.0 0.5 M Sodium phosphate pH 7.2 7% SDS	Room temperature
Low stringency wash Buffer	6 X SSC 0.1% SDS	Room temperature
Moderate stringency wash Buffer	2 X SSC 0.1% SDS	Room temperature
High stringency wash Buffer	0.2 X SSC 0.1% SDS	Room temperature

2.8.3 Buffers and Solutions for Plasmid isolation (Alkaline lysis method)

Name	**Ingredients**	**Preparation and storage**
Solution I or Resuspension buffer	50 mM Glucose, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0)	Store at 4 ºC
Solution II or Lysis buffer	0.2 N NaOH and 1% SDS	Freshly prepared
Solution III or Neutralisation buffer	3 M Potassium acetate (pH 4.8)	Store at 4 ºC
RNase A	10 mg/mL	Store at -20 ºC
Other solutions or Reagents	Chloroform, Absolute ethanol, 3.0 M Sodium acetate, 70% ethanol & Deionized sterile water	Room temperature

2.8.4 Buffers and Solutions for Protein Gel Electrophoresis (PAGE)

Name	**Components**	**Preparation and storage**
Monomer solution	29.2% acrylamide 0.8% bis-acrylamide in water	Store 4 ºC (in darkness)
Stacking gel	Distilled water 3.40 mL 1 M Tris-HCl (pH 6.8) 00.63 mL Acrylamide/bis 30%, 0.83 mL 10% SDS 0.05 mL 10% (w/v) APS 0.05 mL TEMED 04 µL	Freshly prepared

Separating gel	Distilled water 3.3 mL 1.5 M Tris-HCl (pH 8.8) 2.0 mL Acrylamide/Bis 30% 4.0 mL 10% SDS (SDS-PAGE) 0.1 mL 10% (w/v) APS 0.1 mL TEMED 07 µL	Freshly prepared
5 X Protein loading dye	Distilled water 2.7 mL 0.5 M Tris-HCl (pH 6.8) 1.0 mL Glycerol, 2.0 mL 10% SDS(SDS-PAGE) 3.3 mL β-Mercaptoethanol 0.5 mL 0.5% Bromophenol blue 0.5 mL	Room temperature
10x SDS-electrode buffer	Tris base 15.1 g Glycine 94.0 g SDS 0.5 g Adjust pH-8.3. Make up the volume to 500 mL with distilled water.	Room temperature
Staining solution	Coomassie-blue R 250, 0.25 g Methanol, 40 mL Acetic acid, 10 mL Make up volume to100 mL with distilled water	Room temperature
Destaining solution	Methanol, 40 mL Acetic acid, 10 mL Make up volume to 100 mL with distilled water	Room temperature
Silver staining Fixer solution	40% Methanol, (150 mL) 10% acetic acid, (50 mL) Make up volume to 100 mL with distilled water	Room temperature
Sensitising solution	0.2% $Na_2S_2O_3$	Room temperature
Silver stain solution	0.2% silver nitrate (0.6 g) 0.01 % formaldehyde (225 µL) Make up volume to 300 mL with distilled water	Freshly prepared in dark
Developing solution	6% Na_2CO_3 (18 g) 0.02% formaldehyde (150 µL) Make up volume to 300 mL with distilled water	Freshly prepared

2.8.5 Buffers for protein extraction and purification under denaturing conditions

Name	Components	Preparation and storage
Lysis buffer	20 mM Tris-HCl (pH 8.0) 1 mM EDTA 10 mM $MgCl_2$ 0.4 mM NaCl 10% Glycerol Lysozyme 100 µg/mL (Added freshly)	Stored at 4 ºC
Dispersion buffer / Binding buffer	10 mM Tris-HCl (pH 8.0) 100 mM NaH_2PO_4 300 mM NaCl 8 M Urea	Room temperature
Wash buffer	10 mM Tris (pH 8.0) 100 mM NaH_2PO_4 300 mM NaCl 8 M Urea 30 mM imidazole	Room temperature
Elution buffer	10 mM Tris (pH 8.0) 100 mM NaH_2PO_4 300 mM NaCl 8 M Urea 250 mM imidazole	Room temperature

2.8.6 Buffers for protein extraction and purification under native conditions

Name	Components	Preparation and storage
Lysis buffer (GST purification)	50 mM Tris-HCl (pH 7.4) 300 mM NaCl 1 mM EDTA 10% Glycerol 1 mM PMSF (freshly prepared) 1: 100 ratio of Protease Inhibitor Cocktail (CalBiochem) Lysozyme 0.2 mg/mL (Added freshly)	Stored at 4 ºC
Wash buffer (GST purification)	50 mM Tris-HCl (pH 7.4) 300 mM NaCl 1 mM EDTA 1 mM PMSF (freshly prepared)	Stored at 4 ºC

Elution buffer (GST purification)	10 mM Glutathione in 50 mM Tris (pH 8.0)	Freshly prepared
HEPES buffer (Nuclear protein dialysis)	20 mM Hepes (pH 7.9) 1.5 mM MgCl2 0.2 mM EDTA 10 mM KCl 25% (v/v) glycerol 1 mM DTT	Stored at 4 ºC

2.8.7. Different media and buffers used for bacterial studies

Name	Ingredients	Preparation and storage
Luria Bertani Broth (LB)	1% Bactotryptone 0.5% Yeast extract 1% NaCl	pH adjusted to 7.0 with NaOH, store at room temperature or at +4 ºC
SOB media	2% Bactotryptone 0.5% Yeast extract 10 mM NaCl 10 mM $MgCl_2.6H_2O$ 2 mM KCl	pH adjusted to 6.8 with NaOH, store at room temperature or at +4ºC
TB buffer	10 mM PIPES 15 mM CaCl2 250 mM KCl	pH was adjusted 6.8 with KOH. $MnCl_2$ was added to final concentration of 55mM and filter sterilized
YEP	0.5% Beef extract 0.1% Yeast extract 0.5% Peptone 0.5% Sucrose 0.049% $MgSO_4.7H_2O$	Store at room temperature

2.8.8. Buffers and Solutions for ELISA/Immunocytolocalization/ GUS Assay

Name	Ingredients	Preparation and storage
1XPBS	1.44 g Na_2HPO_4 0.24 g KH_2PO_4 0.2 g KCl 8 g NaCl in 1000mL of distilled water.	Room temperature

Antibody dilution Buffer	PBS with 0.25% BSA	Store at 4 ºC
Substrate Buffer	200 mM TrisCl pH 9.5 0.5 mM $MgCl_2$	Freshly prepared
PBS-T	1.44 g Na_2HPO_4 0.24 g KH_2PO_4 0.2 g KCl 8 g NaCl 0.05% v/v Tween-20 in 1000mL of distilled water.	Store at 4 ºC
0.5 X SSC	75 mM NaCl 7.5 mM Na Citrate	Room temperature
BCIP/NBT mix	0.577 mM BCIP 0.122 mM NBT	Store at 4 ºC
Other reagents used	Polyvinyl alcohol, Ethanol, Tertiary butanol, Paraffin, Xylene, Glycerol	Room temperature
X-GluC	1 mM X-Gluc (5-bromo,4-chloro,3-indolyl-β-D-glucuronide:Cyclohexylammonium salt (X-GlcA) in dimethylformamide 100 mM sodium dihydrogen phosphate dihydrate $NaH_2PO_4.2H_2O$ 0.5% Tween-20	Store at 4 ºC
Phloroglucinol	2% Phloroglucinol in 95% ethanol	Store at 4 ºC

2.8.9 Components of Murasighe and Skoog media /hormone and antibiotic stock solutions:

Name	**Components**	**Concentration in mM**	**Storage**
Major component	NH_4NO_3	20.61	Store at 4 ºC
	KNO_3	18.75	
	$CaCl_2.2H2O$	2.99	
	$MgSO_4.7H2O$	1.5	
	KH_2PO_4	1.24	
Minor component	$MnSO_4$	0.147	Store at 4 ºC
	$ZnSO_4$	5.3×10^{-2}	
	$CuSO_4$	1.56×10^{-4}	
	$CoCl_2.6H_2O$	1.05×10^{-4}	

	KI	4.99×10^{-3}	
	H_3BO_4	0.1	
	$Na_2Mo_4.2H_2O$	1.03×10^{-3}	
Vitamins	Myoionsitol	5.55×10^{-2}	Store at 4 ºC
	Nicotinic acid	4.06×10^{-3}	
	Pyridoxine HCl	2.43×10^{-3}	
	Thymine HCl	2.96×10^{-4}	
	Glycine	2.66×10^{-2}	
Iron	FeSO4.7H20	0.1	Store at 4 ºC
	Na2EDTA	0.1	
BAP	Dissolve 20 mg BAP in 1mL 1N NaOH and make up to 50mL with sterile MQ water	1.776 mM	Store at -20 ºC
NAA	Dissolve 20mg in 50mL sterile MQ water	2.148 mM	Store at -20 ºC
Hygromycin	Dissolved in sterile MQ water	50 mg/mL	Filter sterilize and Store at -20 ºC
Cefotaxime	Dissolved in sterile MQ water	250 mg/mL	Filter sterilize and Store at -20 ºC
Acetosyringone	Dissolved in DMSO	200 mM	Filter sterilize and Store at 4 ºC

2.8.10 Different media used for *Nicotiana tabacum* tissue culture

Name	Components for (1000mL)	Storage
Shoot Induction Medium (SIM)	Major (40x)-25 mL Minor (100x)-10 mL Iron (100x)-10 mL Vitamins (200x)-5 mL BAP (4.4 µM))-2.5 mL NAA (5.37 µM)-250 µL Glucose-1.5% Sucrose-2.0% pH-5.6 to 5.8 Agar-0.8%	Store at 4 ºC
Resuspension medium (RSM)	Major(40x)-25 mL Minor(100x)-10 mL Iron (100x)-10 mL	Store at 4 ºC

	Vitamins (200x)-5 mL BAP(4.4µM)-2.5 mL NAA(5.37µM)-250 µL Glucose-1.5% Sucrose-2.0% pH-5.6 to 5.8 Acetosyringone-200 µM $MgSO_4$-40mM	
Selection medium	SIM Hygromycin- 7.09 µM Cefotaxime-200 µM	Store at 4 ºC
Root induction medium (RIM)	Major(40x)-25 mL Minor(100x)-10 mL Iron (100x)-10 mL Vitamins (200x)-5 mL NAA(5.37 µM)-250 µL Glucose-1.5% Sucrose-2.0% pH-5.6 to 5.8 Agar-0.8%	Store at 4 ºC

2.8.11 DNA and protein markers used

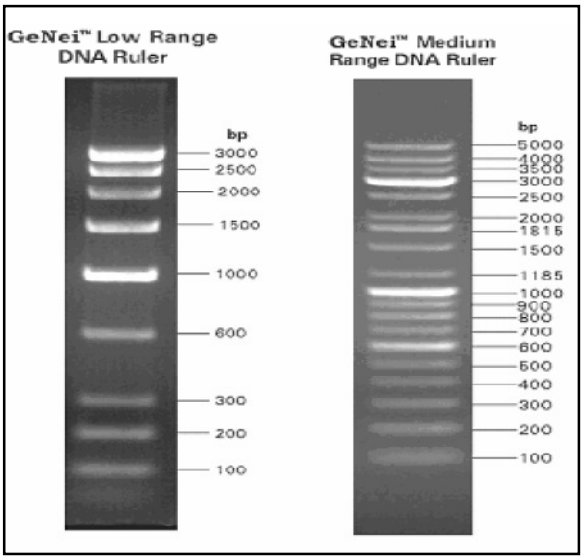


Fig 2.1: DNA Marker (Bangalore Genei)

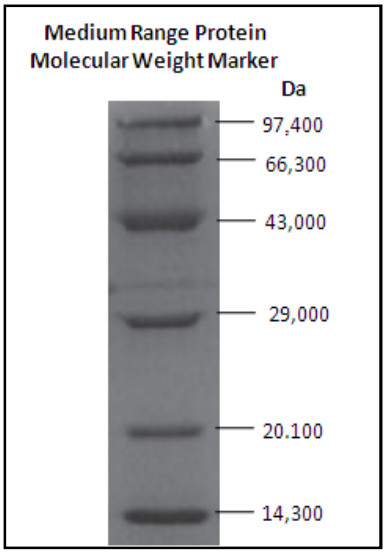


Fig 2.1: Protein Molecular Weight Marker (Bangalore Genei)

2.9 Methods

2.9.1 Bacterial culture conditions

E. coli was grown at 37 ºC with shaking at 200 rpm in Luria Bertani (LB) broth/ SOB broth and maintained on LB/ SOB plates with 1.5% agar (Sambrook *et al.*, 1989). For plasmid DNA preparation recombinant *E. coli* was grown in LB medium supplemented with appropriate antibiotic(s).

2.9.2 Bacterial transformation

2.9.2.1 Preparation of competent cells using TB buffer

A single colony of *E. coli* XL1 Blue was inoculated in 5 mL of LB medium containing tetracycline (12.5 mg / L) and grown overnight at 37 ^{o}C in incubator shaker at 200 rpm. Overnight grown culture (500 µL) was added to 50 mL of SOB medium and grown for 2-3 h at 37 ^{o}C in incubator shaker at 200 RPM. Cells were kept on ice for 10 min and harvested by centrifugation at 5,000 RPM for 10 min at 4 ^{o}C. The cell pellet was suspended in 5 mL ice-cold TB buffer containing 7% DMSO. This was then dispensed into aliquots of 100 µl in 1.5 mL microfuge tubes, frozen in liquid nitrogen and stored at -80 ^{o}C.

2.9.2.2 Preparation of competent cells using $CaCl_2$

A single colony of *E. coli* XL1 Blue was inoculated in 5 mL of LB medium containing tetracycline (12.5 mg / L). LB medium (50 mL) was inoculated with 1% of the overnight grown *E. coli* culture and allowed to grow till A_{600} was 0.5. The cells were harvested by centrifugation at 5,000 g for 10 min at 4 ºC, suspended in ice-cold 100 mM $CaCl_2$ and kept on ice for 30 min. Cells were centrifuged, the pellet suspended in 1 mL of ice-cold 100 mM $CaCl_2$ and stored as aliquots of 200 µL at 4 ºC or in -70 ºC after adding 15% glycerol or 7% DMSO.

2.9.2.3 *E. coli* transformation

The competent *E. coli* cells were transformed according to Sambrook *et al.,* (1989). Gently, DNA (~50 ng in 10 µL or less) was added to the (200 µL) competent *E. coli* cells, mixed and kept on ice for 30 min. The cells were then incubated at 42 ºC for 1.5 min

(heat shock) and immediately kept back on ice, 800 µL of LB broth was added and further incubated at 37 ºC for 1 h. Cells were pelleted by centrifugation at 5,000 rpm for 5 min. The supernatant media was discarded and pellet was dissolved in 100 µL of LB broth. The resuspended cells were plated on LB agar medium with appropriate antibiotics, IPTG and X-gal as per need (Sambrook *et al.*, 1989).

Solutions	**Stock**	**Final concentration**
IPTG	200 mg mL^{-1}in sterile distilled water	40 µg mL^{-1}
X-gal	20 mg mL^{-1} in dimethylformamide	40 µg mL^{-1}

2.9.2.4 *Agrobacterium tumefaciens* transformation and selection

The method adopted for the preparation of *A. tumefaciens* competent cells was essentially as reported by An (1987). A single colony of *A. tumefaciens* (GV2260) was inoculated in 50 mL LB broth containing Rifampicin 150 µg $mL^{-1}$and incubated at 28 ºC with shaking at 200 rpm till A_{600} was 0.5. Cells were centrifuged at 5,000 g for 10 min at 4 ºC and washed twice with ice cold 150 mM $CaCl_2$. The cells were pelleted and resuspended in 1mL of ice cold 20 mM $CaCl_2$. Aliquots of 200 µL were made. For transformation 1µg DNA of the desired binary plasmid vector was added to an aliquot of the competent *A. tumefaciens* cells and incubated on ice for 30 min. The cells were then snap frozen in liquid nitrogen and allowed to thaw at 37 ºC. After thawing 1mL LB medium was added and the tubes incubated at 28 ºC for 2 h with gentle shaking. The cells were centrifuged at 4,000 g for 5 min and 100 µL supernatant was retained. The cells were resuspended in the 100 µL supernatant and plated on LB agar medium with appropriate antibiotic(s). The plates were incubated at 28 ºC for two days to allow the transformed colonies to grow for further processing.

2.9.2.5 Colony Screening by PCR

This method bypasses DNA purification, and relies on the selectivity of PCR amplification to determine whether a bacterial colony of interest does indeed contain the desired DNA. Simply adding a small portion of a bacterial colony to a PCR master mix will introduce enough templates DNA for amplification. A single bacterial colony was

picked up from the agar plate containing transformants with the help of microtip and added to 1.5 mL microfuge tube containing 25 µL of sterile miliQ water. The microtip was agitated in the water to suspend the bacterial cells. Subsequently, 2-3 µL of this suspension was added into 15 µL of PCR reaction. The remaining components were added to the PCR reaction and subjected to normal cycling parameters for the particular primers. If insert orientation, as well as presence, needs to be determined, utilization of a forward vector-specific primer and a reverse insert-specific primer, or vice versa, allows such determination. If only the presence of the insert needs to be determined, then two insert-specific primers can be used. An additional 5 min denaturation step at 95 ºC before the amplification cycles will aid the bacterial lysis to enhance PCR product amplification success. The resulting PCR products were checked on an agarose gel for the presence of the amplicon of expected size.

2.9.3 Preservation of bacteria

Bacterial cultures containing plasmids with target insert were stored in glycerol. In a microfuge tube, 800 µL of bacterial culture were mixed with 200 µL of glycerol (sterilized by autoclaving). The mixture was vortexed to ensure that the glycerol is dispersed. The tube was stored at -80 °C for long-term storage. To recover the bacteria, the frozen surface of the culture was scraped with a sterile inoculating needle and the bacteria adhered to the needle were immediately streaked on the surface of a LB agar plate containing the appropriate antibiotic. The plates were incubated overnight at 37 °C.

2.9.4 Isolation of Nucleic Acids and Polymerase Chain Reaction (PCR)

2.9.4.1 Isolation of plasmid DNA from *E. coli* cells

The alkaline lysis method of Sambrook *et al*., (1989) was improvised upon so that 12-24 samples could be processed conveniently for plasmid DNA extraction within 3 h, with yields of 5-30 µg per 1.5 mL culture depending on the host strain and the plasmid vector. An important feature of this protocol was the use of PEG for purification, which resulted in precipitation of high quality super-coiled plasmid DNA free of contamination. The bacterial cultures were grown overnight with shaking (200 rpm) at 37 ºC in LB broth, with appropriate antibiotic(s). About 1.5 to 3 mL culture was centrifuged for 1 min at

7000 g to pellet the bacterial cells. The pellet was resuspended in 100 µL of Soln.I (section: 2.8.3) by vigorous pipetting, 200 µL of Soln. II (section: 2.8.3) was added, mixed by inversion till the solution becomes clear normally for 2-3 min. The cell lysate was neutralized by addition of 150 µL of Soln. III (section: 2.8.3), mixed well and incubated on ice for 5 min. The cell debris was removed by centrifugation for 10 min at 12000 g at 4 ºC. The supernatant was transferred to another microfuge tube, RNase A was added to a final concentration of 20 µg/mL (Sambrook *et al*., 1989) and incubated at 37 ºC for 20 min. To the above solution 400 µL of chloroform was added, vortexed for 2 min and centrifuged for 5 min at 12, 000 g at room temperature. The upper aqueous layer was transferred to a clean tube, 1/10th volume sodium acetate and one volume absolute ethanol was added with gentle mixing and kept at -20 ºC for 1-2 h. The sample was centrifuged at 12,000 g for 10 min at 4 ºC. The pellet was washed thrice with 70% ethanol and dried under vacuum. The dried pellet was dissolved in 40 µL of deionized water and 40 µL of PEG/NaCl solution (20% PEG 8000 in 2.5 M NaCl) was added. The mixture was incubated on ice for 20 min and the plasmid DNA pelleted by centrifugation at 12,000 g for 15 min at 4 ºC. The supernatant was aspirated carefully, the pellet washed with 70% ethanol and air-dried. The dried pellet was resuspended in 20 µL TE buffer and stored at -20 ºC.

2.9.4.2 Isolation of plant Genomic DNA

Genomic DNA was isolated by using the protocol of Lodhi *et al*. (1994). Fresh young leaves were collected, frozen in liquid nitrogen and crushed to a fine powder. About 1g of ground tissue was extracted with 10 mL extraction buffer (section 2.8.2). The slurry was poured into a clean, autoclaved 50 mL centrifuge tube and 100 mg insoluble polyvinylpolypyrrolidone (PVPP) as well as 20-40 µL of β-mercaptoethanol was also added. The tube was inverted several times but very gently to thoroughly mix the slurry, incubated at 65 ºC for 30 min and then allowed to cool down to room temperature. 12 mL of chloroform: isoamylalcohol mixture was added and the contents mixed by inverting the tube gently till an emulsion formed. The mixture was then centrifuged at 6,000 g for 15 min at room temperature. Supernatant was carefully collected in a fresh tube and chloroform: isoamylalcohol (24:1) extraction step repeated. To the clear supernatant 0.5

volume of 5 M NaCl was added and mixed gently and two volumes of cold (-20 ºC) 95% ethanol was added and the sample kept at 4 ºC until DNA strands appeared. The tube was centrifuged at 3,000 g for 3 min and then at 5,000 g for next 3 min. The supernatant was poured off, and DNA pellet washed with cold (4 ºC) 70% ethanol and air-dried. DNA was dissolved in 400 µL of TE buffer. The DNA solution was treated with 1 µL RNase A (10 mg mL-1) per 100 µL DNA and incubated at 37 ºC for 30 min. The sample was extracted with chloroform: isoamyl alcohol to remove RNase A. DNA was reprecipitated and dissolved in 40-100 µL TE buffer. Purity of DNA was checked spectrophotometrically by measuring the absorbance ratio (A_{260}/A_{280}) and also by visualization on 0.8 % agarose gel with 1X TAE. Genomic DNA was stored at 4 ºC.

2.9.4.3 Restriction digestion of DNA

Plasmid and genomic DNA restriction digestion was set up as per restriction enzymes manufacturer recommendations.

2.9.4.4 Extraction and purification of DNA from agarose gels

The restriction digested DNA or PCR amplified products were run on an agarose gel in 1X TAE buffer. The gel was stained with ethidium bromide (0.5 µg mL^{-1}) and viewed using a hand held long wavelength UV illuminator. The fragment of interest was excised from the gel and weighed. A gel slice 100 µg (upto maximum 400 µg) was transferred to a 1.5 mL micro centrifuge tube and 3 three volumes i.e. 300 µL of Gel Solubilisation Buffer (L3) (Purelink™ Quick Gel extraction kit, Invitrogen) added. The tube was incubated at 50 ºC dry bath for 5 to 10 min with intermittent mixing until the gel slice was completely dissolved. For optimal yield of eluted DNA, 1 gel volume of isopropanol was added to the dissolved gel solution and mixed well. The above molten agarose was put into Quick gel extraction column and placed into 2 mL microfuge collection tube. The assembly was centrifuged at 12,000 g for 1 min and filtrate was discarded. 500-700 µL of wash buffer 1 (provided by Invitrogen) was added and centrifuged at 12,000 g for 30 s, filtrate was discarded. The empty column was centrifuged at maximum speed for 2-3 min to ensure complete removal of residual salts or ethanol from the column. The column was transferred into a fresh 1.5 mL microfuge recovery tube and 25-30 µL of

elution buffer was added to the centre of the membrane and kept it as such for 1 min at room temperature. Then it was centrifuged at 12000 g for 1 min. The eluted DNA was stored at 4 ºC. This eluted PCR product or any DNA fragments are of good quality and can be visualized on 1% agarose gel by taking an aliquots of 3-4 µL. The eluted DNA /PCR product was stored at -20 ^{o}C and was used for subsequent reactions.

2.9.4.5 RNA Extraction

RNase free environment was created and maintained as described by Blumberg (1987). All glass and plasticware were DEPC (0.1% in water) treated overnight and autoclaved. The pestle and mortar were also DEPC treated and then baked at 300 ºC for 6 h. All materials were dried in a vacuum oven. Total RNA from different plant tissues was isolated using TRIzol reagent. The plant tissue was collected, washed with DEPC treated water, frozen in liquid nitrogen and crushed to a fine powder. To 100 mg of the fine powder, 1 mL TRIzol reagent was added and mixed thoroughly using a vortimix. Chloroform: isoamyl alcohol (300 µL) was added and mixed thoroughly using vortimix. The tubes were centrifuged at 13,000 g for 15 min at 4 ºC. The supernatant was transferred to 1.5 mL tubes and the chloroform: isoamyl alcohol step was repeated. The aqueous phase was transferred to 1.5 mL tubes and 0.6 volume isopropanol added. It was mixed thoroughly and kept for RNA precipitation for 1 h at room temperature. Total RNA was pelleted by centrifugation at 13,000 g for 15 min at 4 ºC. The RNA pellet was washed with 70% ethanol twice and dried in a Speedvac centrifugal concentrator. RNA pellet was dissolved in 40 µL of DEPC treated water and stored at -80 ºC in aliquots. Purity of RNA was confirmed by measuring A_{260}/A_{280} ratio and also by visualization on 1.5% TAE Agarose gel.

2.9.4.6 mRNA Purification

Total RNA was quantified spectrophotometrically as well as with the Bioanalyser system. The amount of RNA was in the range of 1-3 mg. Appropriate amount of OBB Buffer and Oligotex suspension were added according to manual instruction. The sample was incubated for 3 min at 70 ºC in a heating block. Sample was removed from the heating block, and placed at 20 – 30 ºC for 10 min. Oligotex: mRNA complex was pelleted down

by centrifugation for 2 min at maximum speed (14,000–18,000 g) and carefully removed the supernatant by pipetting. Oligotex: mRNA pellet was resuspended in the appropriate amount of OW2 buffer by vortexing and pipetted onto a small spin column placed in a 1.5 mL microcentrifuge tube and centrifuged for 1 min at maximum speed. Spin column was transferred to a new RNase-free 1.5 mL microcentrifuge tube, appropriate amount of OW2 buffer added to the column and again centrifuged for 1 min at 12,000 g, flowthrough was discarded. Spin column was transferred to a new RNase free 1.5 mL microcentrifuge tube. Appropriate amount of hot (70 ºC) OEB buffer was pipetted onto the column and resuspended by pipetting up and down; spin it down for 1 min at 12,000 g. The Eluted mRNA sample was quantified spectrophotometrically and used for further downstream processes.

2.9.4.7 Spectrophotometric determination of nucleic acid concentration

DNA concentration was determined by measurement of the absorption at 260 nm. A Lambda 25 Perkin Elmer Photometer was used to determine the concentration of 1:50 diluted RNA or DNA samples in a volume of 1ml in a 1 cm light path quartz cuvette. Absorbance value (A_{260}) should fall between 0.1-1.0 to be accurate. Sample dilution was adjusted accordingly. An absorbance of 1.0 at 260 nm corresponds to 50 µg DNA/ mL. The Bioanalyser system (Agilent 2100) was also used to check the quantity of the nucleic acids (according to manufacturer's instructions). Some other useful parameters like gel electrophoresis pattern of the samples (Nucleic acids and protein), the integrity of the RNA and DNA, concentration of the individual bands can also be analyzed simultaneously with this system. The basis of the system to perform multiple analysis simultaneously is that, one has to load the samples, standards, suitable ladder and the gel matrix in the different wells of a single microchip provided with the system and then the system primarily performs electrophoretic separation of all the samples loaded through the inbuilt micro-capillaries of the chip and then after analyses the various parameters in one go.

2.9.4.8 First strand cDNA synthesis by Reverse Transcription

Complementary DNA (cDNA) was synthesized from a mature mRNA template in a reaction catalyzed by the enzyme reverse transcriptase. The resulting molecule is a DNA-RNA hybrid and the process is called as cDNA 1st strand synthesis. For DNA double strand synthesis this hybrid molecule is digested with RNase H (specific for degrading RNA strand in a DNA-RNA hybrid), DNA second strand is synthesized using DNA polymerase I (Kimmel and Berger, 1987). In the present study cDNA first strand was synthesized using ImPromII™ Reverse Transcription System (Promega, USA). The reactions were set up as per the manufacturer's guidelines. Briefly, reverse transcription reactions of up to 1 µg of total RNA performed in 20 µL reactions comprised of components of the ImPromII Reverse Transcription System. Experimental RNA was combined with the oligo (dT)15 primer. The primer/template mixture was isothermally denatured at 70 ºC for 5 min and snap chilled on ice. A reverse transcription reaction mix was assembled on ice to contain nuclease-free water, reaction buffer, reverse transcriptase, magnesium chloride, dNTPs and ribonuclease inhibitor RNasin®. As a final step, the template-primer combination was added to the reaction mix on ice. Following an initial annealing at 25 for 5 min, the reaction was incubated at 42 ºC for up to 1-1.5 h. The cDNA synthesized was directly added to amplification reactions. The first strand reaction was set up as follows:

Reagent	**Volume**
Experimental RNA (1µg)	1.0 µL
Primer [Oligo(dT)15 or Random (10 pmol)	1.0 µL
DEPC treated	Water 3.0 µL
Final volume	5.0 µL

The tubes were incubated at 70 ºC for 5 min and then chilled in ice-water for 5 min. Tubes were briefly spun in a microcentrifuge to collect the condensate and maintain the original volume. The tubes were kept closed on ice until addition of the reverse transcription reaction mix. The reverse transcription reaction mix was prepared by adding the following components of the ImProm-II Reverse Transcription System in a sterile 1.5 mL microcentrifuge tube on ice.

Reagents	Volume
ImProm-II. 5X Reaction Buffer	4.0 µL
$MgCl_2$ (15 mM)	2.0 µL
dNTP Mix (7.5 mM)	1.0 µL
RNasin® Ribonuclease Inhibitor (40 U/ µl)	0.5 µL
ImProm-II Reverse Transcriptase	1.0 µL
Nuclease-free water	6.5 µL
Final volume	15.0 µL

An aliquot of 1.0 µg total RNA and oligo (dT)15 or Random hexamer primer (10 pmol) mix total vol 5 µl was added to the above reaction for a final reaction volume of 20 µL per tube. The reaction mixure was incubated at 25 ºC for 5 min for primer annealing and then at 42 ºC for 1 h for cDNA first strand synthesis. Reverse transcriptase was thermally inactivated by incubation at 70 ºC for 15 min prior to proceeding with PCR amplification (Chumakov, 1994).

2.9.4.9 Polymerase Chain Reaction (PCR)

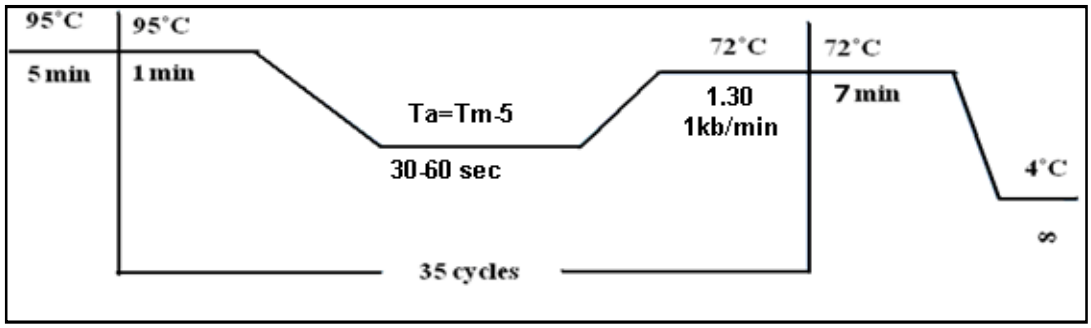
PCR is a powerful technique to amplify a desired nucleotide sequence using sequence specific primers. This amplification may be either of and from a single template or of a template from a mixture of templates (Saiki *et al*., 1985; Mullis and Faloona, 1987; Saiki *et al*., 1988; Arnheim and Erlich, 1992). This technique has been successfully used for various purposes like fishing out gene(s) from genomic DNA or from cDNA population (Todd *et al*., 1987), introducing restriction sites of interest in the amplified product for directional cloning (Scharf *et al*., 1986), creating sequence mismatch/ deletion/ addition resulting in mutant version of a gene or nucleotide sequence (Goodenow *et al*., 1989), differentiating between two alleles (Kwok *et al*., 1990) etc. In the present study applications of PCR were exploited for a few of the above specified applications. The PCR reaction mixture and cycling conditions used were as follows:

Reaction mixture

Reagent	Volume
Sterile deionized water	6.2 µL
Template (50 ng/ µL)	1.0 µL
Forward primer (6 pmol)	1.0 µL

Reverse primer (6 pmol)	1.0 µL
dNTPs (0.2 mM)	4.0 µL
10 x Buffer (Mg^{+2} 1.5 mM)	1.5 µL
Taq DNA Polymerase (1 U/µL)	0.3 µL
Total volume	**15.0** µL

PCR cycle conditions



2.9.4.10 Rapid amplification of cDNA ends (RACE)

Generally, using reverse transcription PCR, either partial cDNA fragments (both 5' and 3' ends missing) or cDNA with full 5'end missing are amplified from total cDNA. If a partial cDNA sequence is known, unknown sequences to the 5' and 3' of the known sequence can be reverse transcribed from RNA, amplified by PCR using RACE (Frohman *et al*., 1988).

An important feature in performing the cDNA synthesis using total RNA was the use of SuperScript® III Reverse Transcriptase, a version of M-MLV RT that has been engineered to reduce RNase H activity and provide increased thermal stability (Kotewicz et al., 1985; Gerard et al., 1986). The enzyme can synthesize cDNA at a temperature range of 45–60 °C, providing increased specificity, higher yields of cDNA, and more full-length product than other reverse transcriptases. Because SuperScript® III RT is not significantly inhibited by ribosomal and transfer RNA, it was used to synthesize cDNA from total RNA.

Preparation of 3' RACE cDNA

RACE Ready cDNA Kit (Invitrogen, USA) was used to perform 3'RACE. The reaction was set up as per the manufacturer's guidelines with minor modification. The prerequisite to begin the 3'RACE was to have a good quality RNA or mRNA, which was verified

using Nanodrop spectrophotometer and the integrity checked on 1% agarose gel. 1-5 µg total RNA or 50-250 ng purified mRNA in DEPC water was used for the 3'RACE reaction**.** The reaction mix including following components was prepared:

RNA	X µL
Oligo dT primer	1 µL
dNTP Mix	1 µL
Sterile , distilled water	11-X µL
Total volume	13 µL

The reaction mix was heated at 65 °C for 5 min to remove any RNA secondary structure, chilled on ice for 1min and centrifuged briefly. Following reagents were added to the 13-µl ligated RNA and primer mixture:

5X First Strand Buffer	4 µl
0.1 M DTT	1 µl
RNaseOut™ (40 U/µl)	1 µl
SuperScript™ III RT (200 U/µl)	1 µl
Total Volume	20 µl

All the above reaction components were mixed by gentle pipeting up and down, centrifuged briefly and incubated at 50 ^{o}C for 50 min. The RT reaction was inactivated by incubation at 80 °C for 10-15 min, chilled on ice for 2 min and centrifuged briefly at maximum speed. 1 µl of RNase H (2 U) was added to the reaction mix and incubated at 37 °C for 20 min, to chew-up the RNA from DNA-RNA hybrids generated during the RT reaction. The prepared cDNA referred to as 3' RACE ready cDNA was stored at -20 °C and used for PCR.

Following sets of primers were used for 3' RACE reaction:

Name	**Size**	**Sequence**
GeneRacer™ OligodT Primer	60 bases	5′-GCTGTCAACGATACGCTACGTAACGGCATGACA GTG(T) $_{24}$-3′
GeneRacer™ 3′ Primer	25 bases	Position 1-25 of GeneRacer™ Oligo dT Primer , Tm- 76°C
GeneRacer™ 3′ Nested Primer	23 bases	Position 14-36 of GeneRacer™ Oligo dT Primer, Tm-72°C

Preparation of 5' RACE cDNA

Briefly, the 5'RACE technique is based on oligo-capping and RNA ligase-mediated (RLM) RACE methods (Maruyama and Sugano, 1994; Vollo Ch *et al*., 1994). The GeneRacer method involves selectively ligating an RNA oligonucleotide (GeneRacer RNA Oligo) to the full-length 5' ends of decapped mRNA using T4 RNA ligase. However, it's a time consuming procedure involving multiple steps and hence a different approach of anchored PCR was adopted for preparing 5'RACE cDNA, which used Abridged Anchor Primer (AAP). The AAP has been specifically engineered to permit optimal amplification of oligo-dC tail. This primer does **NOT** contain the dUMP sequences as found in the original 5′ RACE Anchor Primer, but does contain three restriction sites (*Sal* I, *Spe* I, *Mlu* I) and a *Not* I half-site which may be used for molecular cloning or for excision of cloned insert from a cloning vector. The Abridged Anchor Primer has been used successfully in PCR at annealing temperatures up to 68 °C.

The first strand for 5' RACE cDNA was prepared using GSP-STOP primer as the reverse primer instead of the oligo-dT primer for 3'RACE (rest of the components remaining same as for 3'RACE cDNA preparation), using the purified total RNA pool. The cDNA was prepared using SuperscriptIII RT, which is capable of synthesizing full length cDNA transcripts upto 10-12 kb. The cDNA was first purified using PCR purification kit (Sigma, USA), to remove any unspent reaction components. The purified cDNA, which was enriched for the gene specific transcripts was poly-C tailed at its 3'-end using terminal deoxynucleotidyl transferase (TdT) enzyme (Promega) by incubating the following reaction mixture at 37 °C for 15min.

Purified cDNA	14 µl
dCTP(10mM)	1 µl
TdT reaction buffer(5X)	4 µl
TdT enzyme	1 µl
Total volume	**20 µl**

The TdT enzyme was inactivated by heat denaturing at 70°C for 10min.

The poly-C tailed cDNA was used as template for 5' RACE PCR reaction where, the Abridged Anchor Primer (AAP), which is a universal primer, was used as the 5'forward primer and the GSP primer designed with appropriate Tm nested internal to the STOP primer as the reverse primer.

Following primer was used for 5' RACE reaction:

Name	Sequence
Abridged Anchor Primer (AAP)	5´- GGCCACGCGTCGACTAGTACG GGI IGGGIIGGGIIG -3´

5' and 3' RACE PCR reaction setup

Reactions were setup to amplify either the 5′ end or the 3′ end of gene of interest. The reaction was set up as follows:

Reagent and concentration	5′ RACE	3′ RACE
AAP 10 µM	3.0 µL	-
Reverse GSP 10 µM	1.0 µL	-
GeneRacer 3' Primer10 µM	-	3.0 µL
Forward GSP 10 µM	-	1.0 µL
RT Template	1.0 µL	1.0 µL
10X High Fidelity PCR Buffer	5.0 µL	5.0 µL
dNTP Solution (10 mM each)	1.0 µL	1.0 µL
Platinum® Taq DNA Polymerase or High Fidelity, 5 U/ µL	0.5 µL	0.5 µL
MgSO4, 50 mM	2.0 µL	2.0 µL
Sterile Water	36.5 µL	36.5 µL
Total Volume	**50.0 µL**	**50.0 µL**

PCR cycling parameters:

Temperature	Time	Cycles
94 ºC	2 min	1
94 ºC	30 S	5
72 ºC 1 min/ 1 kb	2 min	
94 ºC	30 S	5
70 ºC 1 min/ 1 kb	2 min	
94 ºC	30 S	25
68 ºC	30 S	
72 ºC	2 min	
72 ºC	10 min	1

5' RACE product was obtained directly after the primary PCR and was gel purified, cloned and sequenced.

Nested PCR

Nested PCR was done to increase the specificity and sensitivity of 3'RACE products for the 3′ ends of gene. 1 µL of the original amplification reaction was used as template for nested PCR. Reactions were set up as described below:

Reagent and concentration	3′ RACE
GeneRacer 3' Nested Primer10 µM	3.0 µL
Forward Nested GSP 10 µM	1.0 µL
Primary PCR product	1.0 µL
10X High Fidelity PCR Buffer	5.0 µL
dNTP Solution (10 mM each)	1.0 µL
Platinum® Taq DNA Polymerase or High Fidelity, 5 U/ µL	0.5 µL
$MgSO_4$, 50 mM	2.0 µL
Sterile Water	38.5 µL
Total Volume	**50.0 µL**

PCR cycling conditions for nested PCR:

Temperature	Time	Cycles
94 ºC	2 min	1
94 ºC	30 s	30
66 ºC	45 s	
72 ºC	2 min	
72 ºC	10 min	1

10-20 µL of nested PCR product was analyzed on a 1% agarose /ethidium bromide gel. Nested PCR products were shorter by the number of bases between the original primers and the nested primers. The largest product was representing the most full-length message. The band(s) were excised, cloned and sequenced.

2.9.5 Isolation of nuclear proteins

The separation of enriched nuclei from plants is the prerequisite to obtain the nuclear proteome required for various molecular studies. To overcome the technical difficulties associated with relatively pure nuclear protein isolation from a tree species, like *Leucaena leucocephala* a protocol based on the CelLytic PN extraction kit from Sigma was used for the present study. The kit has been previously used for isolation of nuclei

from tobacco, tomato, spinach, *Arabidopsis* (Gonzalez *et al*., 2007) , *maize* (Hernandez *et al*., 2007) and *Xerophyta viscose* (Abdalla *et al*., 2009).The nuclear protein extraction method includes cell wall breakage, followed by disruption of the cell membrane using a detergent. The nuclei can be isolated and separated in different degrees of purity: crude, semipure on a sucrose cushion, and highly-pure on a Percoll®/sucrose layer. The nuclear proteins are subsequently extracted from the nuclei with a high salt buffer.

Kit components:

Nuclei Isolation Buffer 4X (NIB)

Percoll®

Sucrose - 2.3 M

TRITON- X-100 - 10% solution

Extraction Buffer

Nuclei PURE Storage Buffer

Filter Mesh 100 micron pore size

Protocol: Young *L. leucocephala* plant tissues (leaf, stem or root) were used for isolation of crude nuclear protein, with slight modifications in the kit protocol. To isolate nuclei, the plant cell wall was disrupted by grinding frozen tissue material (2.0 g) in liquid nitrogen and 3 ml/g 1X NIB with 1mM DTT was added. After mixing, the suspension was passed through a 100-µm filter mesh and the resultant liquid fraction passed another time through double-layered 100-µm filter mesh. The organelles were collected after centrifugation at 1260 x *g* for 10 min and the supernatant (homogenized NIB supernatant, HBS) was removed. The pellet (organelle fraction) was resuspended in 0.75 ml 1X NIBA buffer [NIB buffer containing protease inhibitor cocktail in 1:100 (v/v) ratio]. The organellar membranes were differentially lysed by the addition of 10% Triton X-100 to a final concentration of 0.3%, 0.5% and 1% for leaf, stem and root respectively and mixed moderately for 2 min using a vortex, and then incubated on ice for 15 min. The Triton X-100 removes only the outer of the two nuclear membranes (Watson *et al*., 1986). To obtain a crude preparation of nuclei, the organelle lysate, obtained after Triton X-100 addition, was centrifuged at 12,000 x *g* for 10 min at 4 °C. The supernatant (membrane lysis supernatant, MLS) was removed and the nuclei pellet washed once with 2 ml NIBA buffer by centrifugation at 12,000 *g* for 5 min at 4 °C. To obtain a semipure preparation

of nuclei, a 1.7 M sucrose cushion was preparedin 1X NIBA onto which the organelle lysate was applied. After centrifugation at 12,000g for 10 min at 4 °C the sucrose cushion and the supernatant (membrane lysis supernatant)
were removed and the semipure nuclei pellet was washed twice in 2 ml NIBA buffer with centrifugation at 12,000 x *g* for 5 min at 4 °C. The nuclei pellets obtained in either preparation were stored in nuclei storage buffer (Sigma) or used immediately for protein extraction.

To extract the nuclear proteins, working extraction buffer was prepared freshly by adding DTT to the kit supplied extraction buffer to a final concentration of 5 mM and adding 1:100 (v/v) Protease Inhibitor Cocktail. Nuclei obtained from the crude and semi-pure preparations (g wet weight) were thoroughly resuspended in the working extraction buffer (2/3 of the pellet volume) and vortexed at medium-high speed for 15-30 min at 4 °C. The suspension was centrifuged for 10 min at 12,000 *g* and the supernatant mainly consistent of the crude nuclear protein transferred to a clean chilled test tube. For storage, aliquots of the supernatant were snap-frozen in liquid nitrogen and stored at -70 °C.

2.9.6 Electrophoretic Mobility Shift Assay (EMSA)

The electrophoretic mobility shift assay (EMSA) has been used extensively for studying DNA-protein interactions (Revzin, 1989; Hendrickson, 1985). This technique is based on the fact that DNA-protein complexes migrate slower than non-bound DNA in a native polyacrylamide or agarose gel, resulting in a “shift” in migration of the labeled DNA band.

In the present study, the Thermo Scientific LightShift Chemiluminescent EMSA Kit was used, which is based on a nonisotopic method to detect DNA-protein interactions. Biotin end-labeled DNA containing the binding site of interest is incubated with a nuclear extract or purified factor or protein. This reaction is then subjected to gel electrophoresis on a native polyacrylamide gel and transferred to a nylon membrane. The biotin end-labeled DNA is detected using the Streptavidin-Horseradish Peroxidase Conjugate and the Chemiluminescent Substrate.

Kit components:

Component	Composition	Storage
10X Binding Buffer	100 mM Tris, 500 mM KCl, 10 mM DTT; pH 7.5,	Stored at -20°C
Poly (dI•dC	1 µg/µL in 10 mM Tris, 1mM EDTA; pH 7.5	Stored at -20 °C
50% Glycerol	-	Stored at -20 °C
1% NP-40	-	Stored at -20 °C
1 M KCl	-	Stored at -20 °C
100mM $MgCl_2$	-	Stored at -20 °C
200mM EDTA pH 8.0	-	Stored at -20 °C
5X Loading Buffer	-	Stored at -20 °C

Protocol: To begin with the binding reactions, the target DNA fragments were biotin-end labeled at 3' or 5' ends first as discussed in Chapter 3. Nuclear protein extracts prepared as described in section 2.9.5 were used as source of target protein. The nuclear protein was dialysed with HEPES buffer (section 2.8.6) using Slide-A-Lyzer® MINI Dialysis Unit from Thermo Scientific to remove excess salts in the extract and hence make it favourable for EMSA. As the specific binding conditions were not already known, only minimal reaction components; e.g., 10X binding buffer and Poly (dI•dC), together with the biotin-labeled target DNA, protein extract and unlabeled DNA of the target fragment were used to set up the binding reaction. Poly (dI•dC), included in the kit, is the nonspecific competitor DNA of choice for most systems and was hence used in the reaction. Binding reaction components were added in the following order in order to maintain the specificity of DNA-protein complexes:

Component	Final amount	Reaction	
		#1	#2
Ultrapure Water		15 µL	(15-X) µL
10X Binding Buffer	1X	2 µL	2 µL
1µg/µL Poly (dI•dC)	50 ng/µL	1 µL	1 µL
Unlabeled Target DNA	4 pmol	-	-
Protein Extract	As per requirement	-	X µL
Biotin End-Labeled Target DNA	20 fmol	1 µL	1 µL

The binding reactions were kept for incubation at room temperature for 30 mins. The reaction was stopped before loading by adding 5 µl of 5X Loading Buffer to each binding reaction, pipetting up and down several times to mix. In the mean time, a 6% native polyacrylamide gel of 8 × 8 × 0.1cm dimensions was prepared in 0.5 X TBE and pre-run for 30-60 mins in cold room at 100 V with 0.5 X TBE buffer. The binding reactions were loaded carefully after flushing the wells and electrophoresed at 100V until the bromophenol blue dye had migrated approximately 3/4 down the length of the gel.

A positively charged nylon membrane (Whatman) was cut to the gel dimensions and soaked in 0.5 X TBE for at least 10 min. The gel and nylon membrane were sandwiched in a clean electrophoretic transfer unit (I-Blot, Invitrogen) and transferred according to the manufacturer's instructions. When the transfer was complete, the membrane was placed with the bromophenol blue side up on a dry paper towel. Without allowing it to dry, the membrane was cross linked for 10-15 min with the membrane face down on a transilluminator equipped with 312 nm bulbs.

2.9.7 Detection of Biotin-labeled DNA by Chemiluminescence

Following kit components were provided for detection of biotin-labeled DNA by chemiluminescence from Thermo Scientific:

Component	**Storage**
Stabilized Streptavidin- Horseradish Peroxidase Conjugate	Stored at 4 °C
Chemiluminescent Substrate	Stored at 4 °C
Luminol/Enhancer Solution	Stored at 4 °C
Stable Peroxide Solution	Stored at 4 °C
Blocking Buffer	Stored at 4 °C
4X Wash Buffer	Stored at 4 °C
Substrate Equilibration Buffer	Stored at 4 °C

Protocol:

The Blocking Buffer and the 4X Wash Buffer were gently warmed to 37-50 °C in a water bath until all particulate was dissolved. To block the membrane prepared in section 2.9.6, 10-20 mL of Blocking Buffer was added and incubated for 15 min with gentle shaking on an orbital shaker. Blocking buffer was decanted from the membrane and replaced with

the conjugate/blocking solution freshly prepared (1:300 dilution of Stabilized Streptavidin-Horseradish Peroxidase in Blocking Buffer). The membrane was incubated in the conjugate/blocking buffer solution for 15 min with gentle shaking. The membrane was transferred to a new container and rinsed briefly with 20 mL of 1X wash solution. The washes were repeated for four times of 5 min each in 1X wash solution with gentle shaking. The membrane was then transferred to a new container and incubated with 30 mL of Substrate Equilibration Buffer for 5 min with gentle shaking. Substrate Working Solution was prepared freshly by adding 6 mL Luminol/Enhancer Solution to 6 mL Stable Peroxide Solution (1:1 ratio). The membrane was carefully removed from the substrate equilibration buffer and excess buffer was removed by blotting an edge of the membrane on a paper towel. Substrate Working Solution was poured onto the membrane so that it completely covered the surface and incubated for 5 min without shaking preferably in dark. The membrane was removed from the Working Solution and excess buffer was blotted out through an edge on a paper towel. Without allowing the membrane to become dry, it was wrapped in a plastic wrap, avoiding bubbles and wrinkles and subsequently exposed to the phosphorimager to obtain the desired signal.

2.9.8 Quantitative real time PCR (QRT-PCR)

Quantitative Real Time PCR is a powerful tool for gene expression analysis and was first demonstrated by Higuchi *et al*., (1992, 1993). Q-PCR quantitates the initial amount of the template most specifically, sensitively and reproducibly, and is a preferred alternative to other forms of quantitative reverse transcriptase PCR that detect the amount of final amplified product at the end-point (Freeman *et al*., 1999; Raeymaekers, 2000). QPCR monitors the fluorescence emitted during the reaction as an indicator of amplicon production during each PCR cycle (i.e., in real time) as opposed to the endpoint detection (Higuchi *et al*., 1992, 1993). The QPCR system is based on the detection and quantitation of a fluorescent reporter (Lee *et al*., 1993; Livak *et al*., 1995). There are three main fluorescence-monitoring systems for DNA amplification (Wittwer *et al*., 1997a): (1) hydrolysis probes, (2) hybridizing probes and (3) DNA-binding agents (Wittwer *et al*., 1997 b, Vander Velden *et al*., 2003). Most commonly used are the hydrolysis probes, which include TaqMan probes (Heid *et al*., 1996), molecular beacons (Tan *et al*., 2004;

Vet and Marras, 2005) and scorpions (Saha *et al*., 2001; Terry *et al*., 2002). They use the 5' exonuclease activity of Taq polymerase to measure the amount of target sequences in cDNA samples. The use of fluorescent probe technologies reduces the risk of sample contamination while maintaining convenience, speed and high throughput screening capabilities (Reischl *et al*., 2002). In the present study the Brilliant® II QPCR Master Mix (Stratagene, USA) was used. This kit supports quantitative amplification and detection with multiplex capability, and shows consistent high performance with various fluorescent detection systems, including molecular beacons and TaqMan® probes. The kit supports PCR amplifications and detection of a variety of DNA targets, including genomic DNA, plasmid DNA, and cDNA. The Brilliant II QPCR master mix includes SureStart® Taq DNA polymerase, a modified version of Taq2000™ DNA polymerase with hot start capability. A passive reference dye (an optional reaction component) is provided so as to control the final dye concentration, increasing the flexibility of the reagents for use with different platforms.

2.9.8.1 Pre-protocol considerations

Magnesium Chloride Concentration

Magnesium chloride concentration in the PCR reaction mix affects the specificity of the PCR primers and probe hybridization (Eckert and Kunkel, 1990). The Brilliant® II QPCR Master Mix kit contains $MgCl_2$ at a concentration of 5.5 mM (in the 1 x solution), which is suitable for most targets

Probe Design

Probes are designed in a fashion so as to have a melting temperature 7–10 ºC higher than the annealing temperature of the primers. Lyophilized custom molecular beacon or TaqMan probes are constituted in 5 mM Tris-HCl (pH 8.0) and 0.1 mM EDTA (low TE buffer).

Fluorescence Detection

Fluorescence may be detected either in real-time or at the endpoint of cycling using a real-time spectrofluorometric thermal cycler. For endpoint analysis, PCR reactions can be run on any thermal cycler and then analyzed with a fluorescence plate reader that has

been designed to accommodate PCR tubes and is optimized for the detection of fluorescent PCR reactions. Reading may be acquired before and after PCR for comparison.

Optimal Concentrations for Experimental Probes and Primers

The optimal concentration of the experimental probe is determined empirically. The optimal concentration is the lowest concentration that results in the lowest Ct and an adequate fluorescence for a given target concentration. In present study TaqMan® probes were used.

TaqMan® Probes

The TaqMan probe concentration can be optimized by varying the final concentration from 50 to 300 nM in increments of 50 nM.

PCR Primers

The optimal concentration of the upstream and downstream PCR primers is determined empirically. The optimal concentration is the lowest concentration that results in the lowest Ct and an adequate fluorescence for a given target concentration. The primer concentration for use with TaqMan probes can be optimized by varying the primer concentration from 100 nM to 900 nM. The best concentrations of the upstream and downstream primers are not always of equal molarity.

Reference Dye

A passive reference dye may be added to compensate for non-PCR related variations in fluorescence. Fluorescence from the passive reference dye does not change during the course of the PCR reaction but provides a stable baseline to which samples are normalized. In this way, the reference dye compensates for changes in fluorescence between wells caused by slight volume differences in reaction tubes. The excitation and emission wavelengths of the reference dye are 584 nm and 612 nm respectively.

Data Acquisition with a Spectrofluorometric Thermal Cycler

Real time PCR can be divided into four major phases: the linear ground phase, early exponential phase, log-linear (also known as exponential) phase, and plateau phase. During the linear ground phase (usually the first 10–15 cycles), PCR is just beginning, and fluorescence emission at each cycle has not yet risen above background. Baseline fluorescence is calculated at this time. At the early exponential phase, the amount of

fluorescence has reached a threshold where it is significantly higher (usually 10 times the standard deviation of the baseline) than background levels. The cycle at which this occurs is known as Ct in ABI Prism® literature (Applied Biosystems, Foster City, CA, USA) or crossing point (CP) in LightCycler® literature (Roche Applied Science, Indianapolis, IN, USA). This value is representative of the starting copy number in the original template and is used to calculate experimental results. During the log-linear phase, PCR reaches its optimal amplification period with the PCR product doubling after every cycle in ideal reaction conditions. Finally, the plateau stage is reached when reaction components become limited and the fluorescence intensity is no longer useful for data calculation. In general, lower Ct value indicates higher initial copies. Acquisition of real-time data generated by fluorogenic probes was as recommended by the instrument manufacturer. Data collection was either at the annealing step of each cycle (3-step cycling protocol) or the annealing/extension step (2-step cycling protocol).

2.9.8.2 Preparing the QRT-PCR reactions

The experimental reactions were prepared by adding the following components in order:

Reagent Mixture

Component	**Volume**
2 X master mix	12.5 µL
Experimental probe (optimized concentration)	X µL
Upstream primer (optimized concentration)	X µL
Downstream primer (optimized concentration)	X µL
Diluted reference dye	0.375 µL
Experimental gDNA, cDNA or plasmid	X µL
Final Volume	25 µL

2.9.9 Nucleic Acids Hybridization

For Southern hybridization (Southern, 1975) DNA samples were electrophoresed on an agarose gel in 1X TAE buffer containing 0.5 µg mL^{-1} ethidium bromide. The gel was rinsed with deionized water (D/W) and placed in depurination solution for 15 min. The

gel was rinsed with deionized water and immersed in denaturation solution for 30 min with gentle shaking. The gel was again rinsed with deionized water and transferred to neutralization solution for 45 min. The gel was next set up for capillary transfer of DNA to solid membrane support. A tray was filled with the transfer buffer (20X SSC). A platform was made and covered with a wick, made from 2 sheets of Whatman 3 MM filter paper saturated with transfer buffer and the gel was placed on it. It was surrounded with Saran Wrap to prevent the transfer buffer being absorbed directly by the paper towels stacked above the membrane. A sheet of Hybond-N^+ membrane (Amersham, UK) of the exact gel size was wetted with deionized water followed by transfer buffer (20X SSC) and then placed on top of the gel. A glass rod was rolled over the membrane to remove any trapped air bubbles. One piece of Whatman 3MM paper wetted with 20X SSC was placed on the membrane followed by Whatman 3MM paper pre-wetted in 2X SSC. On this paper another dry Whatman 3MM paper was placed followed by a stack of absorbent paper towels. A glass plate and a ~0.5 kg weight were placed on the top of the paper towels. Transfer of DNA was allowed to proceed for 18 h. The membrane was marked for orientation, removed carefully and washed with 6X SSC. The membrane was air dried and baked for 2 h at 80 ºC to immobilize DNA onto the nylon membrane. Hybridization and autoradiography were carried out as is described in the following section.

2.9.9.1 Southern hybridization

Genomic DNA extracted from *L. leucocephala* was quantified and checked for its digestibility and quality as well. 20 µg of gDNA was digested up to completion with individual restriction enzymes in separate eppendorf tube. The restriction enzyme was selected on the basis of its presence within the gene of interest. The digested DNA was electrophoretically separated on 0.8% agarose gel. For slot blot hybridization DNA or RNA samples were diluted according to experimental requirements. The DNA samples were denatured by adding $1/10^{th}$ volume of 3 M NaOH and incubated at 65 ºC for 10 min. To the denatured sample an equal volume of 6X SSC was added. Two layers of Whatman 3 MM filter paper wetted with sterile deionized water and 6X SSC were placed in the Slot Blot apparatus followed by Hybond-N^+ membrane (Amersham, UK) treated in the

manner as above. The Slot Blot unit was assembled and wells washed with 500 µL of 6X SSC by applying vacuum. After washing, samples prepared earlier were applied in the wells and vacuum applied till whole sample volume passed through the well slit and wells appear dry. The unit was carefully disassembled and the membrane taken out. The membrane was air dried and baked for 2 h at 80 ºC to immobilize DNA. Radiolabeling of probe, hybridization and autoradiography were carried out as described in the following sections.

2.9.9.2 Random Primer Labeling

Random primer labeling of the DNA probes (Feinberg and Vogelstein, 1983) was done using the Megaprime DNA labeling kit (Amersham, UK). Reaction (50 µL) was set up as follows:

Component	**Volume**
25 ng DNA (used as probe)	5.0 µL
Primer solution (Random hexanucleotides) (3.5 A_{260} U)	5.0 µL
Final Volume	**10.0 µL**

Above mixture was heated in a boiling water bath for 10 min and cooled to room temperature facilitating primer annealing to the DNA. This reaction mixture was added to the following reaction mixture:

Component	**Volume**
10 X reaction buffer (500 mM Tris-HCl, pH 8.0; 100 mM MgCl2; 10 mM DTT; 0.5 mg/ mL acetylated BSA)	5.0 µL
dATP (0.5 mM)	4.0 µL
dGTP (0.5 mM)	4.0 µL
dTTP (0.5 mM)	4.0 µL
(α-^{32}P)-dCTP (Sp. activity 3000 Ci mmol)	5.0 µL
Sterile deionized water	16.0 µL
Exonuclease free Klenow fragment (2 U/ µL)	2.0 µL
Total volume	**50.0 µL**

The reaction was carried out at 37 ºC for 1 to 1.5 hr and stopped by incubation at 80 ºC for 10 min and snap chilled on ice.

2.9.10 Hybridization

The blot made as in section 2.9.9 above was pre-hybridized at 60 ºC in 30 mL of hybridization buffer for 6-8 h in a hybridization incubator (SciGene). The buffer was decanted and fresh buffer added with the denatured radiolabelled probe. Hybridization was carried out at 60 ºC for 14-18 h. The solution was discarded and the membrane washed with low stringency buffer at 55-65 ºC for 5 min followed by a high stringency wash at 55-65 ºC for 5 min. The moist blot was wrapped in saran wrap and placed in the developing / intensifying screen provided with the Typhoon trio+ scanner and closed it. After 6-8 h of exposure, the membrane was removed and the screen was scanned with Typhoon trio scanner system. One can increase the exposure time if the band intensity is weak. The membrane scanning is done usually in the resolution mode instead of sensitivity mode of the scanner. Good quality of band pictures can be obtained by scanning at 500 micron resolution with the system. This instrument will provide many more options and functions to get exquisite band pictures of the exposed membrane. After scanning the intensifying screen, the signals can easily be erased from it, upon exposure to a white light (white light lamp is provided with the system) for 10 min and the same screen can be used to develop the other membrane.

2.9.11 Expression and purification of recombinant protein

2.9.11.1Screening for over-expressing recombinant *Ll*MYB-SSM protein

Heterologous expression of *Ll*MYB-SSM was done in *E. coli* (BL21) host cell. The transformants were grown at 37 ºC in Luria-Bertani medium containing kanamycin (50 µg/mL). Screening was done at small scale in glass tubes with 5 mL of LB broth containing kanamycin, which was inoculated with 10% of the overnight grown culture and incubated at 37 ºC and 200 rpm shaking. When the A_{600} reached 0.6-0.8, the expression was induced by the addition of 1 mM IPTG, and cells were further incubated at 37 ºC. For protein expression analysis cells were harvested after IPTG induction at 5h and isolated by centrifugation. The pelleted cells were resuspended in sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) sample buffer and heated at 90 ºC. Proteins were separated by SDS-PAGE on 10% gels and stained with Coomassie

blue. Protein over-expression was evident from SDS- PAGE gel visualization and also the expression level of different clones appeared to be different. The clone showing optimal level of protein expression in the crude extract was used for large scale studies, for determining its expression level in the soluble and insoluble protein fractions.

2.9.11.2 Protein isolation from inclusion body

The transformed *E. coli* colonies screened for the optimal expression level, as mentioned above, harboring recombinant vector were grown overnight in 5 mL LB-kan media. 500 µL culture was transferred to 50 mL LB-kan media ($1/100^{th}$ of cultured volume) and grown at 37 ºC for 1-2 h until the cultures reached A_{600} of 0.4 to 0.5. The culture was then induced with 1 mM IPTG and grown upto 6 h. After 6 h, the culture was centrifuged at 6000 rpm for 10 min at 4 ºC in a 50 mL centrifuge tube. The pellet of cells was suspended in 3 mL lysis buffer (section 2.8.5) and sonicated for 5 min with pulses of 30 s on and 30 s off. 30 µL lysozyme (0.1 mg/mL) was added to the centrifuge tube and incubated at room temperature for 20 min. The above mixture was then centrifuged at 10,000 rpm for 15 min at 4 ºC. The supernatant (lysate) was saved and pellet (inclusion bodies) thus obtained was suspended in 3 mL of 8 M dispersion buffer (section 2.8.5). An aliquot of 20 µL of inclusion bodies and the lysate fraction were checked on 12% SDS PAGE to confirm the protein expression.

2.9.11.3 Affinity Purification of Recombinant Protein Using Ni+ NTA Beads

The recombinant protein, among several other bacterial proteins is loaded on affinity matrix column such as Ni-agarose. This affinity matrix contains bound metal ion nickel, to which the poly histidine-tag binds with micro molar affinity. The matrix is then washed with buffer to remove unbound proteins. This can be achieved either by increasing the imidazole concentration in buffer or by lowering the pH of the washing and elution buffer. The column(s) and buffers were equilibrated to room temperature. The Ni+ beads column was equilibrated with binding buffer for 30 min and then three bed volume of binding buffer was passed through the column followed by addition of soluble protein in dispersion buffer. The column was kept at 4 ºC for 1 h for binding of recombinant protein to Ni+ beads. Flow through was collected in different tubes after 1 h

and column was washed with two bed volume of wash buffer (section 2.8.5). The washing efficiency may be improved by the addition of 20 mM imidazole and histidine-tagged proteins are then usually eluted with 150-300 mM imidazole. (section 2.8.5). 6 x his-tagged protein was eluted in 4 aliquots of elution buffer, 0.5 mL each. Protein elution was monitored by measuring the absorbance at 280 nm of collected fractions. The eluted protein was separated and analyzed by SDS-PAGE.

2.9.11.4 Affinity Purification of pET41a (+) Cloned Recombinant Protein using GST•Bind ™ Resins

A GST-tag is often used to separate and purify proteins that contain the GST-fusion that are expressed in *E. coli* or other prokaryotic expression systems. The GST•Bind™ purification systems are based on the widely recognized affinity of glutathione-S-transferase fusion proteins for immobilized glutathione. Proteins are quickly and easily purified to near homogeneity in a single chromatographic step. Glutathione-resin based purifications require that the GST domain is soluble and properly folded. The gentle elution condition (10 mM reduced glutathione) avoids target protein denaturation. GST•Bind resin utilizes an 11-atom spacer arm to covalently attach reduced glutathione via a sulfide linkage. The high degree of substitution of glutathione ensures a high binding capacity with yields of GST fusion proteins of 5-8 mg/ml settled resin. The purity and amount of protein is assessed by SDS-PAGE**.**

In present study GST•Bind ™ resin from Novagen was used. The soluble fraction of recombinant GST-tagged *Ll*MYB_SSM protein was obtained in the lysate using the lysis buffer (section 2.8.6) followed by lysozyme treatment for 30 min on ice and sonication as mentioned in section 2.9.11.2. Before purification, the GST•Bind resin was equilibrated with binding buffer (lysis buffer), as mentioned in section 2.9.11.3. The column was then kept in cold room (10 ºC) for 2 h for binding of recombinant protein to GST•Bind resin and rest of the protocol (washing and elution) was performed in cold room to rescue the purified protein in active form. Flow through was collected in different tubes after 2 h and column was washed with four bed volume of washing buffer (section 2.8.6). Finally, the GST-tagged protein was eluted in 4 aliquots of elution buffer,

0.5 mL each. Protein elution was monitored by measuring the absorbance at 280 nm of collected fractions. The eluted protein was separated and analyzed by SDS-PAGE.

2.9.11.5 Polyacrylamide gel electrophoresis (PAGE)

PAGE system is the widely used electrophoresis system for protein separations (Laemmeli, 1970). The resolution in a Laemmeli gel is excellent because the treated peptides are stacked in a stacking gel before entering the separating gel.

2.9.11.6 Preparation of the Separating Gel

A vertical slab gel (Hoeffer Scientific, U.S.A.) was assembled using 1.0 mm spacers. In a side armed vacuum flask, 10% separating gel solution was made according to section 2.8.4, except for the addition of ammonium persulfate and TEMED. The solution was degassed by applying vacuum while stirring on a magnetic stirrer. TEMED and ammonium persulfate were added, and mixed gently without generating bubbles. The solution was pipetted into the gel cassette leaving 1.5 cm from the top unfilled. The gel solution was overlaid with water saturated n-butanol to remove trapped air bubbles and to layer evenly across the entire surface. When a sharp liquid-gel inter-surface was observed after the gel polymerization, the slab was tilted to pour off the overlay.

2.9.11.7 Preparation of the Stacking Gel

Stacking gel solution was prepared according to section 2.8.4, excluding ammonium persulfate and TEMED. As in the separating gel, this solution was degassed. TEMED and ammonium per sulfate were added, mixed and overlaid on the separating gel. A comb was inserted taking care not to trap air bubbles beneath the comb teeth. The gel was left to polymerize.

2.9.11.8 Preparation of the sample

Equal parts of the protein sample and the loading buffer were mixed in a microcentrifuge tube and kept in a boiling water bath for 3-5 min for SDS-PAGE. Then the samples were centrifuged at 10,000 g for 10 min, supernatant collected and used for electrophoresis. Gel was run at room temperature at constant ampere 20 mA or 80 V. The protein samples

for native PAGE were mixed with native dye and gel was run at 80 V at low temperature ranging from 4-10 ºC.

2.9.11.9 Loading and Running the polyacrylamide Gel

Once wells have been formed in the gel, flushed it with distilled water and drained off completely. The wells were filled with tank buffer and the samples under laid using a syringe. Lower and upper buffer chambers were filled with tank buffer. Voltage was set between 70-80 units. The run was stopped when the dye reached the bottom of the polyacrylamide gel.

2.9.11.10 Silver staining of the gel Solutions

Protein separated on SDS or native PAGE, was transferred to the fixer solution for 1 h. This was followed by 2 x 20 min washes in 50% ethanol. The gel was transferred to thiosulfate solution for 1 min and rinsed thrice with deionized H_2O (20 s each). The gel was silver stained (HCHO was added to the solution just before use) for 20 min with intermittent shaking in dark. The gel was then rinsed thrice with deionized H_2O (20 s each) and transferred to the developer till the bands developed. The gel was washed with deionized H_2O and stored in fixer.

2.9.12 Histology and Immunocytolocalization

The reagents used for the experiment have been listed in section 2.8.8. Free hand transverse sections were fixed overnight under vacuum in freshly prepared cold 4% buffered formaldehyde (4% paraformaldehyde in 1X PBS). The sections were dehydrated by passages through increasing ethanol: water series (30%, 50%, 70%, 85%, 95% and 100% ethanol) for 30 min each. This was followed by passages through tertiary butanol: ethanol series (25:75, 50:50, 75:25, and 100:0). The sections were rehydrated by treating with 70% and 50% ethanol and 0.5X SSC for 2 min. The rehydrated sections were soaked in two changes of 1X PBS for 10 min each. Next, the sections were washed in 1X PBS containing 0.1% BSA for 5 min and subjected to 30 min of blocking with 10% BSA at room temperature in a humidified chamber. Post blocking washes included three washes of 15 min each with 1X PBS containing 0.1% BSA. Primary antibody incubation

was carried out overnight in a humidified chamber at 4 ºC using 75 µL of diluted antibody (1 µg mL^{-1}) in 1 X PBS containing 0.1% BSA. Negative controls included either the use of pre-immune serum or the omission of both antibody and pre-immune serum. Following the primary antibody incubation, the sections were washed thrice for 15min each in 1X PBS containing 0.1% BSA. A secondary antibody, 0.2% Anti-rabbit-IgG goat alkaline phosphate conjugate antibody (diluted in 1X PBS with 0.1% BSA), was added to the tissue sections at this stage and incubated at 37 ºC in a humidified chamber for 2 h in dark. Post secondary antibody washes were carried out at room temperature using 1X PBS with 0.1% BSA. Color was developed in dark by addition of 60 µL of color development reagent (20 µL BCIP/NBT mix in color development buffer containing 10% polyvinyl alcohol) to the color development buffer (100 mM Tris, pH 9.5, 150 mM NaCl, 50 mM $MgCl_2$) treated slides. In present study we have got the ready to use BCIP/NBT solution therefore the volume from this solution was added sufficiently to submerge the sections in the watch glass. The slides were placed in humidified (color development buffer) chamber at RT in dark for 45 min. Upon color development, 10 mM EDTA was used to stop the reaction, rinsed with water, air dried and cover slip-mounted using glycerol and then it was observed under microscope, and microscopic image captured.

2.9.13 Histochemical Staining

Free hand transverse sections were prepared for histochemistry as described above. Phloroglucinol staining was done by covering pretreated slide with phloroglucinol for half an hour and then concentrated HCl was added (Harris *et al.,* 1980; Speer *et al.,* 1987). Lignified tissue takes red color. A cover slip mounted on the specimen using glycerol was sealed to the slide using nail enamel. The slides were viewed under a light microscope and pictures captured.

2.9.14 ELISA (Enzyme-Linked Immunosorbent Assay)

2.9.14.1 Determination of titre of antibodies

ELISA Buffers listed below were prepared in sterile milliQ water as per the compositions mentioned in section 2.8.8

1X Phosphate buffer saline (PBS)

Wash Buffer (PBST)

Antibody diluting buffer:

Blocking reagent: 1% BSA in PBS

Substrate diluting buffer

ELISA was performed to determine the titre of first, second and third bleed of rabbit serum. Equal quantity of antigen *i.e* 100 ng was coated in triplicates on ELISA plates and kept overnight at 4 °C. Next morning, the plate was washed with 250 µL of PBS-T, three times for 5 min and 300 µL of blocking reagent was added. The plate was wrapped in aluminium foil and kept at 37 °C for 2 h. ELISA plate was washed again as described earlier and challenged to different dilution of serum such as, 1: 1000, 1:5000 , 1: 10000, 1:20000, 1:40000, 1:60000, 1:80000 and 1:100000. All dilutions were in triplicates and plate was kept at 37 °C for 2 h again. Plate was washed as described earlier with PBST buffer and secondary antibodies (anti goat IgG against rabbit IgG) tagged with alkaline phosphatase was added to a dilution of 1:20000. Plate was incubated at 37 °C for two more hours. After 2 h of incubation plate was washed with 250 µL of PBST, three times for 5 min and 100 µL of 1 mg/mL substrate, *p*- Nitro phenyl phosphate (pNPP) was added and incubated for 45 min. Reaction was stopped by adding 10 mM EDTA and readings were taken at 405 nm. Once the antibody titre was determined then a fixed dilution of antibodies was used for rest of the experiments.

2.9.14.2 ELISA of *Ll*MYB_SSM protein in different tissues of *Leucaena leucocephala*

Fresh leaf and stem tissues were collected from mature plants of *L.leucocephala,* frozen in liquid nitrogen and crushed to a fine powder. Crude nuclear protein was isolated as per protocol mentioned in section 2.9.5. The protein was quantified using Bradford reagent. Equal amount of protein was coated on 96 well micro titer plates. Antigen (recombinant protein) was diluted in PBS to an optimal concentration and coated on 96 well microtiter plate (100 µL/ well) in different volumes to generate a standard curve for the same protein concentration determination in unknown samples. Plate was incubated for 2 h at room temperature or over night at 4 ºC and washed twice with PBS after incubation. Non-specific sites were blocked with blocking buffer 300 µL/ well and incubated for 1h

at room temperature. After washing out the unbound blocking agent with PBS, primary antibody (Anti rabbit IgG) was added and incubated for 1 h at room temperature. The unbound primary antibody was washed twice with PBS and secondary antibody conjugated with alkaline phosphatase was added and incubated for 2 h at room temperature. Enzyme specific substrate BCIP/ NBT or pNPP was added and incubated for 45 min for the color development and the absorbance was measured at 405 nm.

2.9.15 *Agrobacterium* mediated tobacco transformation

The tobacco regeneration and transformation protocol is modified from Horch *et al*., (1985). *Agrobacterium tumefaciens* strain GV2260 harboring the binary plasmid vector to be used for tobacco transformation was inoculated in 5 mL YEP media containing rifampicin (250 mg L^{-1}) and kanamycin (50 mg L^{-1}). The culture was allowed to grow overnight at 28 ºC with shaking at 200 rpm. Next day 1 mL inoculum from this tube was taken and added to 50 mL YEP medium in a flask containing appropriate antibiotics. The culture was incubated at 28 ºC for 24 h with shaking at 200 rpm. Fresh leaves were taken from axenic cultures of *Nicotiana tabacum* var. Anand 119 and ~ 6 mm diameter disc were punched out. The leaf discs were inoculated on MS medium agar plates containing 4.4 µM BAP and 5.37 µM NAA (MSBN) and kept for 48 h in dark. *A. tumefaciens* GV2260 cells grown for 24h were centrifuged for 10 min at 5,000 g. Cells were suspended in 20 mL of 50 mM $MgSO_4$ and 200 µM acetosyringone and allowed to stand for 10 min with intermittent shaking. The tobacco leaf discs from axenic tobacco cultures were co-cultivated in this suspension for 10 min with intermittent shaking. The leaf discs were then transferred to MSBN plates without any antibiotics. The adaxial side of the leaf disc was in contact with the medium. The plates were incubated in dark at 28 ºC. After two days the leaf discs were harvested, washed with liquid MSBN and dried on sterile filter paper. Up to 10 leaf discs per plate were transferred to MSBN agar plates containing kanamycin (7.09 µM) and cefotaxime (200 µM). The cultures were incubated at 25±2 ºC with 18/6 h photoperiod in diffused light (24.4 µmol/m^2 s^{-1}). Leaf discs were sub-cultured every week till elongated shoots were excisable. Elongated shoots were excised and transferred to root induction medium. The rooted plants were hardened and further analyzed.

2.9.16 GUS histochemical assay

Bacterial β-glucuronidase (GUS) is the most widely used reporter gene system for evaluating transient and stable transformation in plants because of its high sensitivity and the easy way of performing histochemical as well as quantitative assays. However, endogenous or intrinsic GUS-like activity reported in several untransformed seed plants including tobacco (Hu *et al.*, 1990), has resulted in false positives for interpretation of results. An important property to differentiate between the interfering plant-intrinsic GUS-like activity and the bacterial GUS was found to be their different behaviour during thermal inactivation Whereas the *E.* coli derived GUS is rather resistant to elevated temperature with a half life at 55°C of about 2 h (Jefferson *et al.*, 1987), plant-endogenous GUS showed already after 10 min a 50% inhibition of activity at 50°C (Alwen *et al.*, 1992). Similar results were reported by Hodal *et al.*(1992) for young, non-transgenic and transgenic tobacco plants, and has been used to modify the regular protocol by Jefferson (1987).

The GUS staining solution was prepared by taking 1 mM X-Gluc (5-bromo, 4chloro, 3-indolyl-β-D-glucuronide: Cyclohexylammonium salt (X-GlcA) from a 20 mM stock made in dimethylformamide, 100 mM sodium dihydrogen phosphate dihydrate and 0.5% Tween-20. The pH of the solution was adjusted to 7.0 with 1N NaOH.

Control and transgenic tobacco leaf tissues were preincubated at 55 °C for 1 h before incubation with the GUS staining solution to heat inactivate the endogenous GUS-like activity. After this the tissues were incubated at 37 °C in dark for overnight, dipped in the assay solution. GUS gene expression was distinctly observed in the transformed tissue as compared to the non-transformed one by bleaching in 70% alcohol for 3-4 h, later replaced by absolute ethanol and photographed by using an Axioplan 4 microscope from Carl Zeiss.

2.9.17 BRADFORD PROTEIN ASSAY

Principle

The assay is based on the observation that the absorbance maximum for an acidic solution of Coomassie Brilliant Blue G-250 shifts from 465 nm to 595 nm when binding to protein occurs. Both hydrophobic and ionic interactions stabilize the anionic form of

the dye, causing a visible color change. The assay is useful since the extinction coefficient of a dye-albumin complex solution is constant over a 10-fold concentration range.

Reagent required

1. Bradford reagent: Dissolve 10 mg Coomassie Brilliant Blue G-250 in 5 mL 95% ethanol; add 10 mL 85% (w/v) phosphoric acid. Dilute to 100 mL when the dye has completely dissolved, and filter through Whatman #1 paper just before use.

2. (Optional) 1 M NaOH (to be used if samples are not readily soluble in the color reagent).The Bradford reagent should be a light brown in color. Filtration may have to be repeated to rid the reagent of blue components.

ASSAY

1. Warm up the spectrophotometer 15 min. before use.

2. Dilute samples with buffer to an estimated concentration of 20 to 200 mg/mL.

3. Prepare standards containing a range of 20 to 200 mg protein (albumin or gamma globulin are recommended) to a standard volume (generally 1 mL or less).

4. Prepare unknowns to estimated amounts of 20 to 200 mg protein per tube, same volume as the unknowns.

5. Add 0.25 mL 1 M NaOH (Optional) to each sample and vortex.

6. Add 5 mL dye reagent and incubate 5 min.

7. Measure the absorbance at 590 nm

ANALYSIS

Prepare a standard curve of absorbance versus micrograms protein (or *vice versa*), and determine amounts from the curve. Determine concentrations of original samples from the amount of protein, volume/sample, and dilution factor, if any.

CHAPTER 3:

Cloning and characterization of promoter sequences for lignin biosynthetic pathway genes (C4H, CCoAOMT, CCR and CAD) From *Leucaena leucocephala*

3.1 Introduction

Promoters are the 5' upstream non-coding regions of a gene, carrying various cis-elements which interact with certain trans-regulatory elements, called transcription factors and control the expression of the gene. As already discussed in section 1.3.3, cis-elements found on the promoters of lignin biosynthetic genes play an important role in transcriptional regulation of the pathway. Hence, promoter analysis has been the main tool for studying the regulation of many phenylpropanoid pathway genes. Promoter analysis has been done for PAL (Bevan *et al.,* 1989; Ohl *et al.*, 1990; Hauffe *et al.*, 1991; Levya *et al.*, 1992), C4H (Bell-Lelong *et al.*, 1997), 4CL (Douglas *et al.*, 1991), COMT (Capellades *et al.*, 1996), CCoAOMT (Chen *et al.*, 2000) and CAD (Feuillet *et al.*, 1995). The transient expression assays of parsely CCoAOMT promoter, with sets of its 5'-truncated promoter::GUS fusion showed that significant promoter activity was retained in the 354 bp fragment upstream of the transcription start site (Grimmig and Matern, 1997). All these gene promoters directed the expression of GUS reporter gene in lignifying tissues. Several potential cis-regulatory elements were identified on their promoters including the AC-rich elements like the P, L, G and H boxes (Lois *et al.*, 1989; Logemann *et al.*, 1995). These elements also represent motifs recurring in the promoters of several genes of stress inducible phenylpropanoid pathway. AC elements are shown to serve as binding sites for certain plant MYB-related transcription factors, which are known to coordinate the transcriptional activity of phenylpropanoid genes (Sablowski *et al.*, 1994). Deletion analyses and gel-retardation studies have identified these AC elements as being critical for tissue-specific expression and are found to be conserved in the promoters of most of the phenylpropanoid pathway genes.

As there is no report on the study of regulation of lignin biosynthesis in *Leucaena leucocephala*, it was necessary to gain an insight into the vital cis-elements found on the promoters of the genes involved in the pathway. The present chapter deals with the isolation and cloning of the promoter regions of the monolignol pathway genes like C4H, CCoAOMT, CAD and CCR *from Leucaena leucocephala,* followed by their i*n silico* comparative analysis using bioinformatic tools. One of the primary aims of the chapter is to determine the role of MYBPLANT or AC element in coordinated regulation of these genes. Hence, promoter regions necessarily carrying the MYBPLANT element, along

with other cis-regulatory elements were used for DNA-protein interaction studies with nuclear protein extracts. The ability of partial CCoAOMT promoter in driving GUS expression was also studied in a heterologous system, tobacco, by transforming it with a promoter::*gusA* fusion construct by *Agrobacterium* mediated transformation system.

3.2 Materials & Methods

3.2.1 gDNA isolation from *Leucaena leucocephala*

Good quality of plant gDNA was isolated as described in Chapter 2, section (2.9.4.2) for carrying out PCR or construction of Genome Walker libraries.

3.2.2 Promoter isolation using TAIL PCR

Various PCR based methods have been used for amplification of promoter region for a gene. TAIL (Thermal Asymmetric Interlaced) PCR (Liu and Whittier 1995; Hui *et al.* 2002) is one of the methods to define the 5' end flanking region of a known gene sequence. TAIL-PCR relies on a series of PCR amplifications with gene specific and degenerate primers to reliably amplify the unknown 3'or 5' ends of a gene.

Methodology: A series of 3 PCR reactions (designated primary, secondary and tertiary TAIL) were performed and the AD (Arbitrary degenerate) primers were as follows: (5' to 3'): NGTCGASWGANAWGAA, TGWGNAGSANCASAGA, AGWGNAGWANCAW-AGG, STTGNTASTNCTNTGC, NTCGASTWTSGWGTT and GTGNAGWANCAN-AGA. The GSPs (Gene specific primers) were designed with a length of 21–24 nucleotides, melting temperature of about 64 °C and GC content of 50–60%. The thermal cycling conditions are described in detail in Fig. 3.1

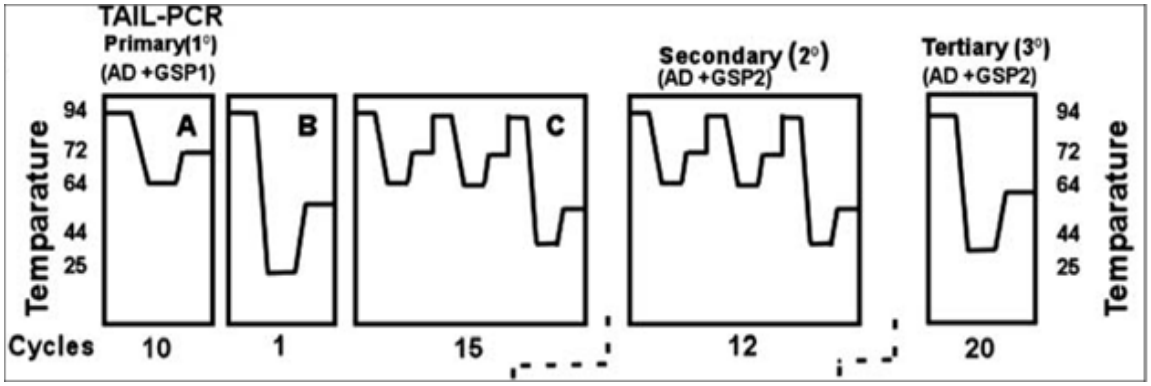


Fig 3.1: Primary, secondary and tertiary TAIL PCR conditions: In the primary TAIL-PCR, the outermost specific primer (GSP1) is used with the degenerate primers (AD); ten high stringency cycles (Primary-A) increase the single-stranded template by extension of GSP1; the low stringency cycle (Primary-B) will generate the complimentary strands: Primary-A and Primary-B together achieve moderate template amplification, providing the template for exponential amplification in Primary-C. Primary-C has 15 ‘‘TAIL’’ cycles or ‘‘super cycles’’ each with two high stringency cycles (annealing temperature 64 °C) during which GSP1 binds to the template and the amplification is linear, followed by a lower stringency cycle (annealing temperature 44 °C) during which AD and GSP bind and amplify the product exponentially. There is a moderate yield of specific products from the primary TAIL reaction, but most of the products are nonspecific from lower stringency binding of GSP to non-target areas in the genome. The product of primary TAIL reaction is diluted 200-fold and used as template for the secondary TAIL reaction, along with the same degenerate primers (AD) and GSP2 which is nested internal to GSP1. Secondary TAIL is similar to Primary-C (two higher stringency and one lower stringency cycle); the use of nested gene specific primer essentially eliminates further amplification of the non-specific products obtained from GSP1 binding to non-template DNA. In the tertiary TAIL reaction, 200-fold diluted product of secondary TAIL reaction is used as template, along with GSP3 (nested internal to GSP2) and the same degenerate primers. Only lower stringency cycles are required given the template is already highly enriched in the specific product.

Thermal cycling conditions for Primary TAIL-PCR were as follows:

Temperature	Time	Cycles
94°C	(5 min)	1
94°C,	(10 s)	10
64°C	(30 s)	
72°C	(3 min)	
94°C	(10 s)	1
25°C	(3 min)	
72°C	(2.5 min)	
94°C	(10 s)	15
64°C	(3 min)	
72°C	(2.5 min)	
94°C	(10 s)	
64°C	(3 min)	
72°C	(2.5 min)	
94°C	10 s	
44°C	1 min	
72°C	2.5 min	

Secondary TAIL-PCR conditions were as follows:

Temperature	Time	Cycles
94 °C	(5 min)	1
94 °C,	(10 s)	12
64 °C	3min	
72 °C	2.5 min	
94 °C	10 s	
64 °C	(3 min)	
72 °C	(2.5 min)	
94°C	10 s	
44°C	1 min	
72°C	2.5 min	

Tertiary TAIL-PCR conditions were as follows:

Temperature	Time	Cycles
94 °C	5 min	1
94 °C	10 s	20
44 °C	1 min	
72 °C	2.5 min	

Following the tertiary PCR, the products were gel purified and sequenced.

3.2.3 Cloning of promoter fragments in pGEM-T vector in XL-10 cells

The promoter fragments amplified using TAIL PCR or fished out using PCR based method were gel purified as discussed in Chapter 2, section 2.9.4.4 and following ligation reaction was set-up with pGEM-T vector and kept at 16 °C overnight :

Ligation reaction:

Vector	0.5 µL
Insert	4 µL
Ligase buffer (2X)	5 µL
T_4 DNA ligase	0.5µL
Total reaction	**10 µL**

The ligation mixture was used for transforming XL10 competent cells as discussed in chapter 2, section 2.9.2.3, followed by screening for the positive colonies for the insert and sequencing the plasmid isolated from the positive clone.

3.2.4 Isolation of nuclear proteins

Crude nuclear protein was extracted from mature leaf and stem tissues of *L. leucocephala* as discussed in chapter 2, section 2.9.5 and quantified using Bradford assay as per section (2.9.17).

3.2.5 Biotinylation of partial promoter regions carrying MYBPLANT element

5'-end biotinylated forward primers and promoter specific non-biotinylated reverse primers were designed for amplifying the promoter regions carrying the MYBPLANT element.

Primer name	Primer sequence
C4H-Bio-F	5'-Bio-AACGGCACCGTCATGATCCT-3'
C4H-pro-R	5'-GGTAGGTAGGTGAGGGAAGATTGAA-3'
C4H-pro-GR	5'-GGAAGCTTGAAGCGATTCCCTCGAA-3'
CCoAOMT-Bio-F	5' Bio-AAAGGCGTGACTTATTTATCTATATAA-3'
CCoAOMT-pro-R	5'-TACTTCTTATCCTTTTTTGTTTCCTCG-3'
CCR-Bio-F	5' Bio-GATTTTGTAATGGGGTTGGTAATTCTG-3'
CCR -pro-R	5'- CATGGTGGTGGGTAGTGTTAATAA- 3'
CAD-Bio-F	5'Bio-TCTCTGATGTCATGTTCTAATGTCTTC-3'
CAD -pro-R	5'-ACATGGAAACTTTTGCCTCGTCTT-3'

3.2.6 Construction of recombinant pCAMBIA-1381Z vector carrying partial CCoAOMT promoter

The partial CCoAOMT promoter was cloned in the binary vector pCAMBIA-1381Z without a promoter for driving the *gusA* reporter gene, for carrying out *Agrobacterium* - mediated transformation of tobacco.

3.2.6.1 Physical and genetic map of pCAMBIA1381Z vector

pCAMBIA-1381Z plasmid belongs to the pCAMBIA series of binary vectors for which the vector backbone is derived from the pPZP vectors (Hajdukiewicz *et al.*, 1994) and is designed for promoter testing *in planta*. This vector features a promoterless version of *gus*A (N358Q) with the catalase intron immediately downstream of a truncated *lacZα* containing the pUC8 polylinker. The presence of intron inside the coding sequence ensures that expression of glucuronidase activity is derived from eukaryotic cells, not from expression by residual *A.tumefaciens* cells. The plasmid has hygromycin as the plant selection gene, and bacterial selection is available with kanamycin. The truncated *lacZα* is functional for blue/white screening of clones in suitable *E.coli* host strains.

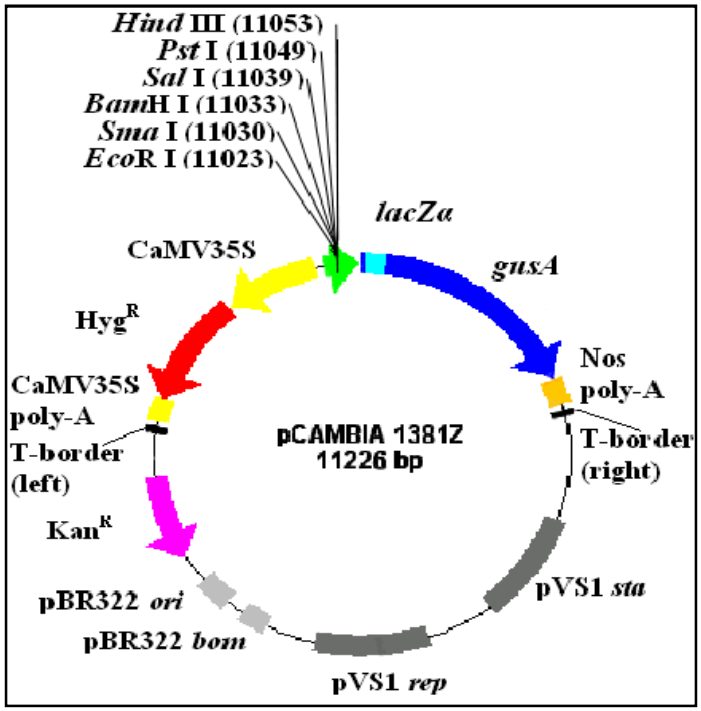


Fig3.2: Physical and genetic map of pCAMBIA1381Z vector

3.2.6.2 Cloning of CCoAOMT promoter in the vector

Based on restriction analysis of the partial CCoAOMT promoter sequence, *Eco*RI and *Hin*dIII sites were chosen to clone the promoter in the vector, pCAMBIA-1381Z to drive the expression of *gus* gene. Following primer sets were designed to reamplify the promoter sequence with the desired flanking sites incorporated at the 5' and 3' ends:

Primer name	**Primer sequence**	**Tm of the oligos**
CpF*Eco*	5'-*GAA TTC* CAACTTCATCCACGTTCA-3'	59.3
CpR*Hin*	5'-*AAGCTT* TACTTCTTATCCTTTTTTGTTTCCTCG-3'	63.3

The reamplified promoter was cloned in pGEM-T Easy vector and sequenced for validation. The modified promoter fragment was then separately cloned in *Eco*RI and *Hin*dIII digested pCAMBIA-1381Z vector and transformed in XL10 cells to stabilize the cloned product.

3.2.7 *Agrobacterium tumefaciens* transformation

The recombinant plasmid **1381Z-CCoAOMTpro**, cloned in XL-10 cells was used for transformation of *Agrobacterium tumefaciens* as discussed in chapter 2, section 2.9.2.4. The positive colonies appeared 24-48 hrs after incubation at 28 ^{o}C were confirmed using colony PCR with CCoAOMT promoter specific primers as discussed in section 2.9.3.

3.2.8 *Agrobacterium* mediated transformation of tobacco

Axenic tobacco explants were prepared for carrying out *Agrobacterium*-mediated transformation of the leaf discs with the promoterless *gus*A construct and CCoAOMTpro::*gus*A constructs cloned in pCAMBIA1381Z vector. The protocol was followed as described in chapter 2, section 2.9.15. After 4 months of subculturing, small shoots were obtained which were directly used for PCR and transient GUS expression assay.

3.2.9 GUS Assay of transformed plants

The GUS histochemical assay was performed as described in chapter 2, section 2.9.16.The protocol was modified to eliminate the endogenous gus activity in tobacco.

3.3 Results and Discussion

3.3.1 Isolation and cloning of C4H and CCoAOMT promoters

Gene specific primers were designed based on the nucleotide sequences for C4H (GenBank: GU183363.1) and CCoAOMT1 (GenBank: DQ517929) genes isolated from *L. leucocephala* available in the NCBI database for 5' genome walking using TAIL PCR. Following set of reverse primers were designed from the 5'coding sequences of the two genes:

Primer name	**Primer sequence**	**Tm of the oligos**
C4H-GSP-R1	5'-GTAACCGCCTCCTCAGCACGATCCCATT-3'	64 °C
C4H-NGSP-R1	5'-AACGGCACCGTCATGATCCTCCTC-3'	64 °C
CCoAOMT R1	5'-GGGTACCTGGTAGAGAGCATCACTCTGC-3'	64 °C
CCoAOMT NR1	5'-CTTGGTGCCTTCCTGCTTCGC-3'	66 °C

Primary, secondary and tertiary TAIL- PCR reactions were set up as discussed in section 3.2.2, with *L. leucocephala* genomic DNA using the reverse Gene Specific Primers (GSPs) and Arbitrary Degenerate Primers (ADP1, 2, 3, 4). The final product after tertiary PCR with ADP 3 for C4H and ADP6 for CCoAOMT genes resulted in an amplicon of ~600bp for C4H gene (Fig. 3.3a) and ~500 bp for CCoAOMT gene (Fig. 3.4a).

The amplicons were gel eluted and cloned in pGEM-T Easy Vector .The positive clones were screened using colony PCR and recombinant plasmids were isolated for restriction digestion with *Eco*RI to confirm for the presence of inserts [Fig 3.3(b) and Fig 3.4(b)] and sequenced. The clones, designated C4H Pro1 and CCoAOMT Pro1 of 605 bp and 548 bp respectively were identified. Both clones showed the presence of the initiation codon **ATG** and the reverse primer. The C4H Pro1 and CCoAOMT Pro1 nucleotide sequences were aligned with the gene sequences. The C4H Pro1 started 195 bp upstream of the translation initiation codon (ATG). The CCoAOMT Pro1 started 508 bp upstream of the ATG. Both the promoter sequences, C4H Pro1 and CCoAOMT Pro1, have been submitted to NCBI GenBank database with designations **C4Hpro1** and **CCoAOMTPro1** and the sequence data is available under accession numbers **JN656956** and **GU13243** respectively.

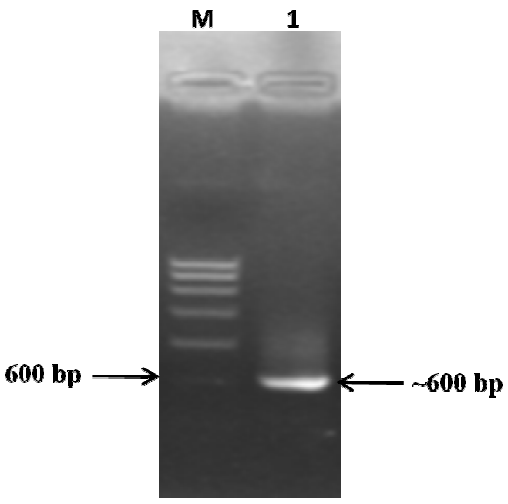


Fig3.3 (a) C4H tertiary TAIL PCR product easy Lanes: M: 100bp DNA ladder; 1: amplicon insert

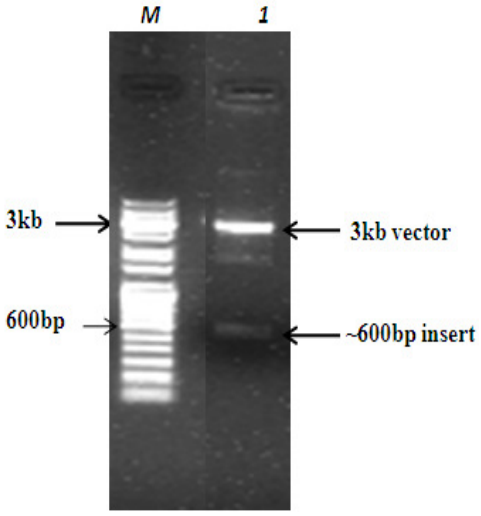


Fig3.3 (b) C4H Pro1 insert in pGEM-T vector. Lanes: M: MR DNA ladder; 1:

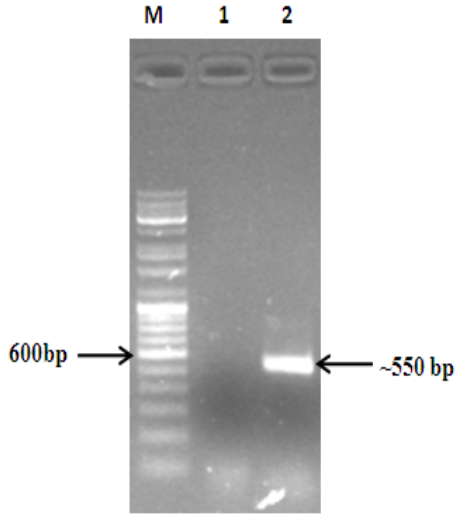


Fig3.4 (a) CCoAOMT tertiary TAIL PCR product Lanes: M: MR DNA ladder; 2: amplicon DNA ladder; 2: insert

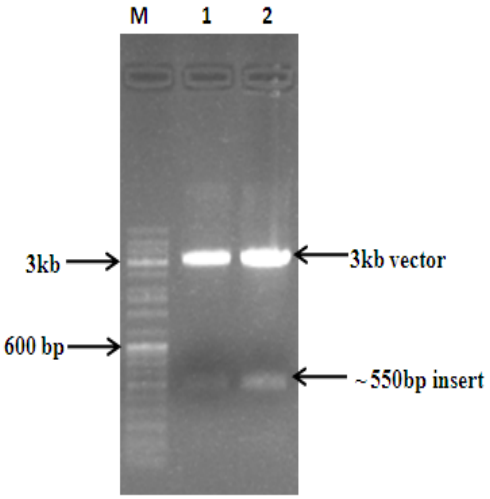


Fig3.4 (b) CCoAOMT Pro1 insert in pGEM-T easy vector, Lanes: M: MR

```
-195 AAACGGCACC GTCATGATCC TCCTCATGCC CCAATATTCC CCTAACGTAT
-145 GCCACTTCAC CACCACACCC GCACCAATTC CCTTCCACCA TAAAACAAGC
 -95 GATCCCTTCT ACCTCTCAAC CAACCTCCAT TTCTGAAATT AATTTTCCTC
 -45 GAACACTTCA TCTTCCTCGA TTCAATCTTC CCTCACCTAC CTACCATGAA
                                                    +1
  +6 TCTCCTCCTC CTGGAGAAGA CCCTGCTCGG CCTCTTCGTT GCCGCTGTGG
 +56 TCGCCATTGT CGTTTCCAAG TTTCGAGGGA ATCGCTTCAA GCTTCCGCCG
+106 GGTCCTCTTC CTGTCCCCAT CTTCGGAAAT TGGCTTCAGG TCGGTGATGA
+156 TCTCAACCAC CGGAATTTGA CCGATTTGGC GAAGAAGTTT GGCGATATCT
+206 TCTTGCTCCG GATGGGGCAG AGGAATCTTG TGGTTGTTTC GTCGCCGGAG
+256 TTGGCGAAGG AGGTGCTGCA CACGCAGGGG GTGGAGTTCG GATCTAGGAC
+306 TCGGAATGTG GTGTTCGACA TCCTCACTGG TAAAGGTCAG GATATGGTGT
+356 TTACGGTGTA CGGGGAGCAC TGGAGGAAGA TGAGGAGGAT CATGACGGTG
+406 CCGTT
```

Fig3.5: Nucleotide sequence of the C4H gene promoter, C4H Pro1. The sequence in italics is the promoter region. First nucleotide of the translation initiation site is underlined and designated **+1** as a reference for upstream and down stream numbering of the remaining sequence. The reverse primer, C4H-NGSP-R (nested) is indicated as shaded portion.

```
-508 CAACTTCATC CACGTTCACC TTCGATTTCC CTAAAAAAAT TTCATATACA
-458 CCAATATTCA TCCCCTTTAG TACTTCCCTG CCTTCTTTTC CTAAGGTAAT
-408 TTTTATAACT AATATGAATA TTCATAAAAA TTAGAAAAAA TATTACAAAT
-358 GAATATAATT ACTAAAAGTT GAAATATTTT TTCTTACAGA ATATTTTGAA
-308 ATTTAAAATT TAAAGGCATG ACTTATTTAT CTATATAGAT AATTATACGA
-258 GTGGATTGGG GTGGTGGTCG AATATTGCTG GCCCACGTTC ACGCCGTACA
-208 AGAACCTCAC CAACCACACC CGGTCGGCAG CCGGTTCAAC ACCACTGCTG
-158 CTTTTGCTCC TCCCATGCCG GTTCAAACCG GACAGTTCCC CCTCCCTATA
-108 TATCATACAA GCCGCCTTCG CCATGTTTAA GGCTCTGGAA GAAAGACTCG
 -58 AAAACAGAGC AGAAGAAAAG AAGAGAGAGT TCGAGGAAAC AAAAAAGGAT
  -8 AAGAAGTAAT GGCGGATCAG AATCAAAGCG AAGCAGGAAG GCACCAAG
          +1
```

Fig 3.6: Nucleotide sequence of the CCoAOMT gene promoter, CCoAOMT Pro1. The sequence in italics is the promoter region. First nucleotide of the translation initiation site is underlined and designated **+1** as a reference for upstream and down stream numbering of the remaining sequence. The reverse primer, CCoAOMT-NR1 (nested) is indicated as shaded portion.

3.3.2 Cloning of CCR and CAD promoters

Gene specific primers were designed based on the promoter sequences available in the NCBI database for CCR (GenBank: **GU984572**) and CAD (GenBank: **EU722904.3**) genes isolated from *L. leucocephala*, to fish out the promoter sequences for the study.

Primer name	Primer sequence	Tm of the oligos
CAD F1	5' CAGAATTATCATCGCTGCATGCAA 3'	58 °C
CAD R1	5' AGTTGTCTCAGAGGCAGAGAGAG 3'	61 °C
CCR F1	5' GATAAGAGAAGTGACAGTGTTTAAACACA 3'	60 °C
CCR R1	5' CATGGTGGTGGGTAGTGTTAATAA 3'	58 °C

An amplification of ~900 bp was obtained for CCR promoter (Fig 3.7) and ~2 kb for CAD promoter (Fig 3.8) which was in agreement to the NCBI data. The two amplicons were gel eluted, cloned in pGEM-T easy vector and sequenced for their further characterization. The CCR promoter was 971bp and CAD promoter was 1961bp, as confirmed by sequencing.

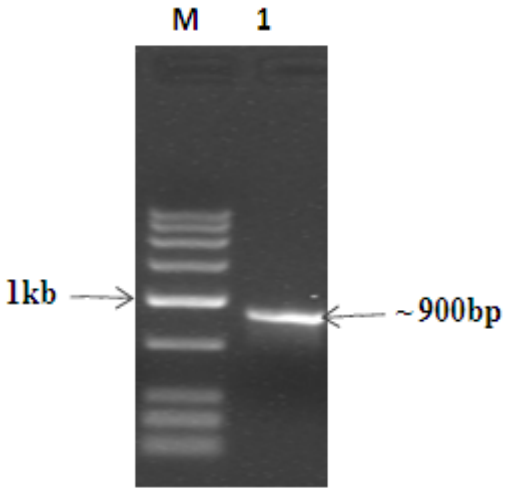


Fig. 3.7 CCR promoter amplicon
Lanes, M: Low range ladder, 1: amplicon amplicon

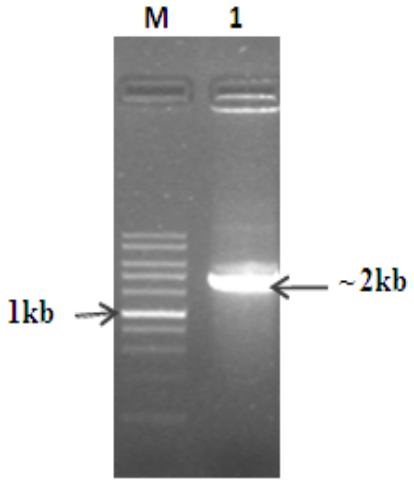


Fig. 3.8 CAD promoter amplicon
Lanes, M: Low range ladder, 1:

3.3.3 Bioinformatic analysis of C4H, CCoAOMT, CAD and CCR promoters

The nucleotide sequences of the promoters were analyzed using signal scan search tool against the PLACE database (Higo *et al,* 1999). Various cis-acting regulatory elements were found on the promoters like TATA box, CAAT box, etc. which have been earlier reported in different plant promoters and play role in gene regulation. Total number of elements identified overall was 124. Number of elements identified on every promoter sequence was as per the size of each promoter. 92 elements were identified on CAD, 70 on CCR, 55 on CCoAOMT and 22 on C4H promoter sequences. Almost 9 elements were found to commonly occur in all the 4 promoter sequences, which were mostly found in 8-10 copies in every sequence and their putative functions and occurrence have been explained in table 3.1. The most important element, MYBPLANT, a cis-acting regulatory element (Ac.No.S000167), having signal sequence MACCWAMC (where M= A/C and W=A/T) responsible for binding of R2R3 type MYB transcription factors was found on all the promoter sequences, which are known to regulate the lignin biosynthetic enzymes. However, its occurrence did not exceed 1-2 even in longer promoter sequences like Cad and CCR.

Table 3.1: Common elements in all the promoter sequences

Where: M=A/C**; W=**A/T**; Y=**C/T**; R=**A/G**; N=**A/T/G/C

S.No.	Name of element	Signal Sequence	Accession number	Times of Occurrence on + and - strand			
				C4H	CCoAO-MT	CCR	CAD
1	CAATBOX1	CAAT	S000028	**5**	**3**	**5**	**21**
2	CACTFTPPCA1	YACT	S000449	**2**	**6**	**10**	**32**
3	DPBFCOREDCDC3	ACACNNG	S000292	**1**	**1**	**2**	**3**
4	GT1CONSENSUS	GRWAAW	S000198	**3**	**7**	**11**	**20**
5	GTGANTG10	GTGA	S000378	**3**	**3**	**6**	**16**
6	**MYBPLANT**	MACCWAMC	S000167	**2**	**1**	**1**	**2**
7	**MYBPZM**	GGATA	S000179	**3**	**1**	**3**	**2**
8	POLLEN1LELAT52	AGAAA	S000245	**2**	**4**	**8**	**16**
9	ROOTMOTIFTAPOX1	ATATT	S000098	**2**	**14**	**16**	**11**

Functions/ Descriptions of the common elements:

1. **CAATBOX1:** "CAAT promoter consensus sequence" found in legA gene of pea
2. **CACTFTPPCA1:** Tetranucleotide (CACT) is a key component of Mem1 (mesophyll expression module 1) found in the cis-regulatory element in the distal region of the phosphoenolpyruvate carboxylase (ppcA1) of the C4 dicot *F. trinervia*.
3. **DPBFCOREDCDC3:** A novel class of bZIP transcription factors, DPBF-1 and 2 (Dc3 promoter-binding factor-1 and 2) binding core sequence; Found in the carrot (*D.c.*) Dc3 gene promoter; Dc3 expression is normally embryo-specific, and also can be induced by ABA.
4. **GT1CONSENSUS:** Consensus GT-1 binding site in many light-regulated genes, e.g., RBCS from many species, PHYA from oat and rice, spinach RCA and PETA, and bean CHS15; R=A/G; W=A/T; GT-1 can stabilize the TFIIA-TBP-DNA (TATA box) complex; The activation mechanism of GT-1 may be achieved through direct interaction between TFIIA and GT-1; Binding of GT-1-like factors to the PR-1a promoter influences the level of SA-inducible gene expression.
5. **GTGANTG10:**"GTGA motif" found in the promoter of the tobacco (*N. tabacum)* late pollen gene *g10* which shows homology to pectate lyase and is the putative homologue of the tomato gene *lat56*.
6. **MYBPLANT:** Plant MYB binding site; Consensus sequence related to box P in promoters of phenylpropanoid biosynthetic genes such as PAL, CHS, CHI, DFR, CL, Bz1; Myb305; M=A/C; W=A/T; The AmMYB308 and AmMYB330 transcription factors from *Antirrhinum majus* regulate phenylpropanoid and lignin biosynthesis in transgenic tobacco.
7. **MYBPZM:** "CCWACC" core of consensus maize P (Myb homolog) binding site; Maize P gene specifies red pigmentation of kernel pericarp, cob, and other floral organs.
8. **POLLEN1LELAT52:** One of two co-dependent regulatory elements responsible for pollen specific activation of tomato (*L.esculentum*) *lat52* gene; Found at -72 to -68 region; AGAAA and TCCACCATA (S000246) are required for pollen specific expression; Also found in the promoter of tomato endo-beta-mannanase gene (LeMAN5) gene .

9. **ROOTMOTIFTAPOX1:** Motif found in promoters of rolD from *A. rhizogenes* showing root specific expression in plants

One of the important elements common to all the promoter sequences is MYBPLANT (AC rich element), which is known to regulate lignin biosynthetic pathway genes by binding to R2R3MYB transcription factors. MYBPZM, another MYB protein binding element known to regulate pigmentation in plants by regulating flavonoid pathway genes was common to the promoter sequences. CAATBOX1 element responsible for proper initiation of transcriptional machinery was discovered on all the promoter sequences. Two of the elements (GTGANTG10, POLLEN1LELAT52) seem to be responsible for pollen specific expression of the genes. An element DPBFCOREDCDC3, known to bind to bZIP transcription factors and a GT1CONSENSUS sequence responsible for light -regulated gene expression were also found on all the promoter sequences. Few other motifs which were not functionally significant have not been discussed here.

Table 3.2: Location of MYBPLANT element on the promoter sequences

S.No.	Promoter name	Size in bps	Position on (+) strand	AC element
1	CCoAOMT	508	-201	CACCAACC
2	C4H	195	-12, -79	AACCAACC, CACCTACC
3	CAD	1961	-1218, -1078	CACCTAAC, CACCAAAC
4	CCR	971	-115	CACCAACC

Table 3.3: Total number and unique elements identified

S.No.	Promoter name	Total number of elements	Unique elements
1	CAD	92	25
2	CCR	70	12
3	C4H	22	2
4	CCoAOMT	55	9

Elements common only between CAD and CCR:

BIHD1OS: Binding site for a homeodomain transcription factor induced in disease resistance response

CTRMCAMV35S: CT-rich motif (inverted GAGA) found in a 60-nucleotide region downstream of the transcription start site of the CaMV 35S RNA known to enhance gene expression.

MYCATERD1, MYCATRD22: MYC recognition sequence found on dehydration responsive genes discovered in *Arabidopsis.*

PREATPRODH: "PRE (Pro- or hypoosmolarity-responsive element) found in the promoter region of proline dehydrogenase (ProDH) gene in *Arabidopsis*; binds with a bZIP subgroup of transcription factors.

SEBFCONSSTPR10A: Binding site of the potato silencing element binding factor (SEBF) gene found in promoter of pathogenesis-related gene (PR-10a), similar to auxin response element.

SURECOREATSULTR11: "GAGAC" Core of sulfur-responsive element (SURE) found in the promoter of SULTR1; 1 high-affinity sulfate transporter gene in *Arabidopsis*; containing auxin response factor (ARF) binding sequence.

WBOXATNPR1: "W" box element recognized specifically by salicylic acid (SA)-induced WRKY DNA binding proteins.

These elements possibly signify their presence in longer stretches of promoter sequences as both CAD and CCR promoter sequences are longer than C4H and CCoAOMT. Most of these elements (BIHD1OS, MYCATERD1, MYCATRD22, PREATPRODH and SURECOREATSULTR11) are associated with disease or different stress responses. An enhancer element (CTRMCAMV35S) and a silencer element (SEBFCONSSTPR10A) were also discovered on the two promoters which might be responsible for regulating the expression level of the two genes. A W-box motif (WBOXATNPR1) regulated by WRKY DNA-binding proteins was also found on the sequences.

Elements common only between C4H and CCoAOMT:

CBFHV: A putative cis- element found on dehydration responsive gene promoters.

PALBOXLPC: Represents one of the three putative cis-acting elements (Boxes, P, A, L) of phenylalanine ammonia-lyase gene in parsley, conferring elicitor or light responsiveness in plants

Elements common only between CAD, CCR and C4H:

ASF1MOTIFCAMV:"TGACG" motifs are found in many promoters and are involved in transcriptional activation of several genes by auxin and/or salicylic acid and may be relevant to light regulation

INRNTPSADB: An initiator element found in TATA-less promoters

SORLIP1AT: "GCCAC" motif containing Sequences Over-Represented in Light-Induced Promoters

WRKY71OS: TGAC-containing "W box" elements responsible for binding of WRKY transcription factors involved in regulating gibberelin or abscisic acid signalling pathways

Elements common between CAD, CCR and CCoAOMT:

GATABOX: A common element responsible for light regulated gene expression.

GT1GMSCAM4: "GT-1 motif" found in the promoter of soybean (*Glycine max*) CaM isoform, SCaM-4 which plays a role in pathogen- and salt-induced gene expression

IBOXCORE: A conserved sequence upstream of light-regulated genes found in both monocots and dicots

TATABOX5: "TTATTT" A functional TATA box found in glutamine synthetase gene from *P.sativum.*

MYBST1: Binding site for single MYB repeat proteins, which act as transcriptional activators in plants.

WBOXNTERF3: "W box" found in the promoter region of a transcriptional repressor ERF3 gene in tobacco

DOFCOREZM: Core site required for binding of DNA-binding (Dof) proteins associated with expression of multiple genes involved in carbon metabolism in maize.

MYCCONSENSUSAT: Also known as E-box with the consensus "CANNTG" with MYC recognition site found on promoters of dehydration or cold responsive genes.

PYRIMIDINEBOXOSRAMY1A: Pyrimidine box found in promoters involved in sugar repression in response to gibberelic acid.

Elements common between CCR, C4H and CCoAOMT:

BOXLCOREDCPAL: Acc. No. (S000492), Signal sequence (ACCWWCC)

This element is the "core" sequence of box-L-like sequences in carrot (*D. carota*) PAL1 promoter region and DcMYB1 bound to these sequences *in vitro* is known to act as a transcriptional activator in response to in response to elicitor treatment, UV-B irradiation and the dilution effect

Elements common between CAD, C4H and CCoAOMT:

CCAATBOX1: "CCAAT box" element commonly found in 5'UTR region of eukaryotic genes, known to act cooperatively with HSEs (Heat Shock Elements) to increase the heat shock promoter activity.

Elements common between CAD and CCoAOMT only:

ANAERO1CONSENSUS: One of the 16 motifs found on the promoters of anaerobically induced genes of different species.

CURECORECR: Copper-responsive elements associated with oxygen-deficiency responsive genes

SREATMSD: Sugar-responsive element discovered in most of the *Arabidopsis* genes associated with axillary meristem formation.

P1BS: Element found on phosphate starvation responsive genes,

TAAAGSTKST1: Element responsible for guard cell-specific gene expression

Elements common between CCR and CCoAOMT only:

ANAERO2CONSENSUS: One of the 16 motifs found on the promoters of anaerobically induced genes of different species.

CPBCSPOR: A novel cis-element involved in Cytokinin-Dependent Protein Binding *in vitro* in the 5'-region of NADPH-Protochlorophyllide Oxidoreductase Gene in Cucumber

WBOXHVISO1: "TGACT", a W-box element in barley, binding site for sugar-signalling pathway related WRKY transcription factors.

Signal scan search engines like PLACE; depend on detection of CARE (cis-acting regulatory elements) in the promoters, which is not self-evident, because such short motifs are statistically expected to occur at random every few hundred base pairs. Therefore, the main problem lies in discriminating "true" from "false" regulatory elements (Blanchette and Sinha, 2001). Therefore, only certain motifs with significant functions have been discussed in the section to show the functional conservation between the four promoter sequences.

Various elements with similar functions were seen conserved among the promoter sequences, especially those associated with dehydration, cold and anaerobic stress responses. Putative binding sites for other transcription factors like WRKY and MYC were also found on the promoters, which might play important roles in regulating the expression of these genes. Due to the smallest size, C4H promoter shared no common elements exclusively with CAD and CCR promoters.

3.3.4 Gel retardation studies of the promoters with nuclear proteins:

3.3.4.1 Mobility of biotinylated promoters on gel

Promoter regions of the genes: C4H, CCoAOMT, CCR and CAD preferably carrying the MYBPLANT element were 5'-end biotin- labeled as discussed in section 3.2.5. The reverse primer used for C4H was C4H-Pro-GR, to check the amplicon's gel mobility. The labeled promoter fragments were run on 6% native PAGE to find out the extent of their mobility on the gel with all the EMSA components but without nuclear protein (Fig. 3.9). The EMSA reaction was carried out as discussed in Chapter 2, section 2.9.6. The loading dye was allowed to run more than 3/4th of the gel, as we were using promoter sequences upto 200-300bp and not simply oligos as provided in the kit.

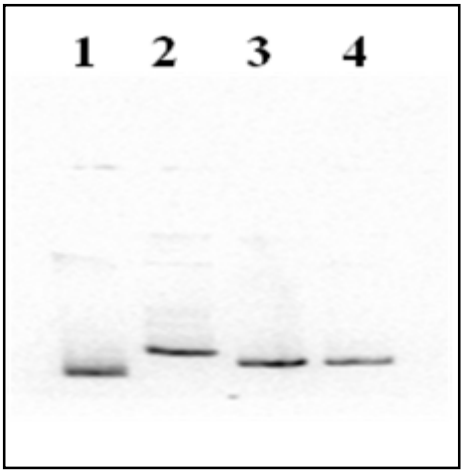


Fig. 3.9: Biotinylated promoters (2 ng) without any proteins run on 6% native polyacrylamide gel and blotted on Nytran suprecharged membrane. Lane 1: CCR promoter (247 bp); Lane 2: CAD promoter (302 bp) ;Lane 3: C4H promoter (295 bp); Lane 4: CCoAOMT promoter (299 bp)

3.3.4.2 Nuclear proteins isolated from leaf and stem tissues

Gel retardation studies were carried out using crude nuclear proteins which were isolated using CelLytic PN extraction kit from Sigma (Chapter 2, section: 2.9.5). The protein profile was checked on 12% SDS gel for leaf [Fig3.10 (a)] and stem [Fig.3.10 (b)] extracts.

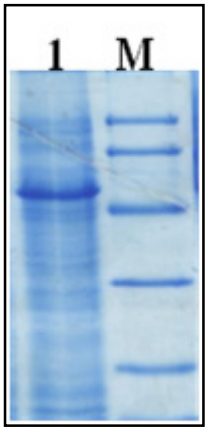


Fig3.10 (a) Leaf nuclear protein
Lane1: Leaf NP; M: Protein marker

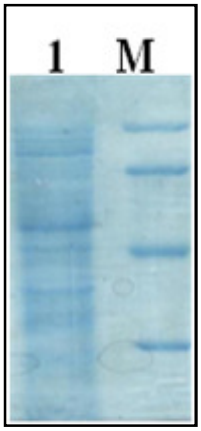


Fig. 3.10(b) Stem nuclear protein
Lane1: Stem NP; M: Protein marker

The crude protein was quantified using Bradford Assay (2.9.17), to be used for EMSA studies. The crude nuclear protein (CNP) from leaf was of higher concentration i.e. 4 µg/µL as compared to stem NP, which was estimated to be 2 µg/µL. The protein extracts were dialysed against HEPES buffer and stem nuclear protein was concentrated using concentrator solution. The final concentrations of the dialysed proteins were approximately 1 µg/µL for both the extracts.

3.3.4.3 Electrophoretic mobility shift Assay of promoters with the nuclear proteins

Gel retardation experiment of one of the promoters, C4Hpro1 (using reverse primer C4H-Pro-R) was carried out with different amount of leaf and stem CNPs, in order to optimize the concentration of nuclear protein required to carry out the retardation. The reverse primer used to amplify the C4H promoter was C4H-pro-R. The binding reactions of 20 µL were set up as per the conditions mentioned in the Pierce kit (chapter 2, section: 2.9.6), but the protein extracts were varied from 2 µg to 10 µg. The binding reactions were kept for 30-40 min at room temperature and then run in 6% native PAGE in cold room until the loading dye ran upto the same extent as mentioned in section 3.3.4.1.

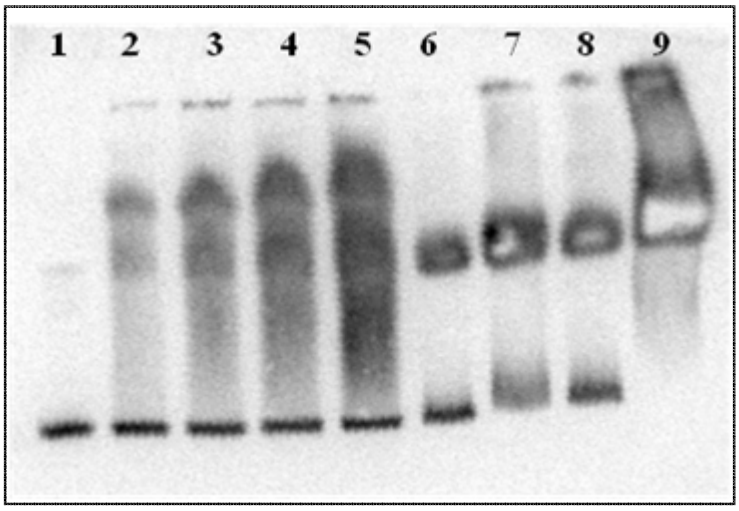


Fig 3.11: EMSA of C4H promoter with leaf and stem crude nuclear proteins. Lane 1: 1 ng of C4H promoter (without protein); Lanes 2, 3, 4 & 5: 1 ng C4H promoter + 2 µg, 4 µg, 6 µg & 10 µg leaf nuclear protein respectively; Lanes 5, 6, 7 & 8: 1ng C4H promoter + 2 µg, 4 µg, 6 µg & 10 µg stem nuclear protein respectively

As seen in Fig.3.11 (Lanes 2-5) C4H promoter formed two types of complexes with the varying concentrations of leaf CNP. The extent of complex formation increased with increasing protein concentrations. In case of stem nuclear proteins (lanes 6-9) C4H appeared to form a single but intense complex with increasing protein concentrations.10 µg of the stem nuclear protein appeared to completely retard the promoter fragment. However, 6µg of both stem and leaf nuclear proteins was chosen to be optimal for carrying out gel retardation studies, to bind with 1ng/µL of biotinylated promoters .

Binding reactions were set up for CCoAOMT, CCR and CAD promoters with 6 µg of leaf and stem CNPs and similar conditions as for C4H promoter EMSA to observe the pattern of retardation.

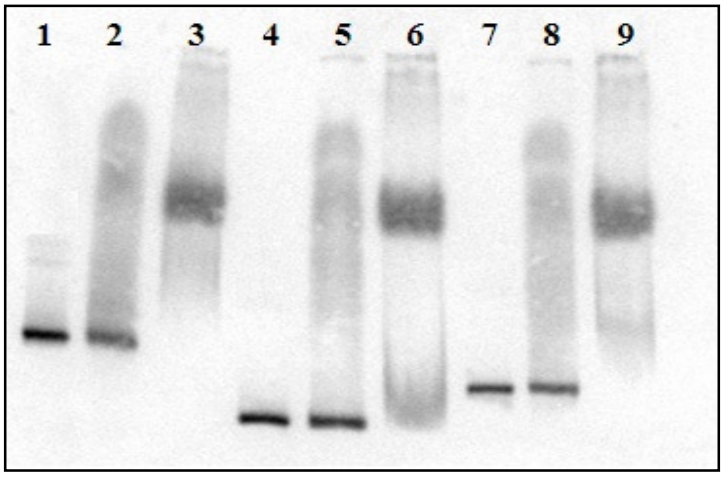


Fig. 3.12: EMSA of CAD, CCR & CCoAOMT promoters with leaf and stem nuclear proteins Lanes1, 4 & 7: 1 ng of CAD, CCR & CCoAOMT promoters (without protein); Lanes2, 5&8: 1ng of promoters in the above order + 6µg leaf nuclear protein respectively Lanes 3, 6 & 9: 1ng of promoters in the above order + 6µg stem nuclear protein respectively

As seen in Fig. 3.12 the pattern of complex formation was similar in each of the promoters. The retardation with leaf CNP appeared to be even weaker than for C4H with 6 µg of protein for all the three promoters. However, a complex of similar intensity was formed with the stem CNP for all the three promoters. CAD promoter was most strongly retarded as compared to CCR and CCoAOMT promoters.

3.3.5 Directional cloning of CCoAOMT promoter in pCAMBIA-1381Z vector

Based on the partial CCoAOMT promoter sequence obtained as described in section 3.3.1 its restriction analysis was done using pDRAW32 software and *Eco*R1 and *Hin*dIII sites were found to be compatible for cloning in pCAMBIA-1381Z vector to drive the expression of *gus* gene. The partial promoter was reamplified using the modified set of forward and reverse primers with flanking restriction sites as listed in section 3.2.7.2, and cloned in pGEM-T vector. The cloned recombinant pGEM-T plasmid construct was digested with *Eco*R1 and *Hin*dIII enzymes to create compatible ends for ligation with pCAMBIA-1381Z vector digested with the same enzymes [Fig. 3.13(a)]. The gel eluted compatible vector and insert were quantified and ligation reaction was set up at16 ^{o}C in dry bath for overnight. The ligation was transformed in XL10 cells and plasmid was isolated from a positive colony. The plasmid was double digested with the two enzymes to confirm for the presence of insert [Fig. 3.13 (b)] and correct orientation of the cloned promoter insert was confirmed by sequencing with the gene specific primers.

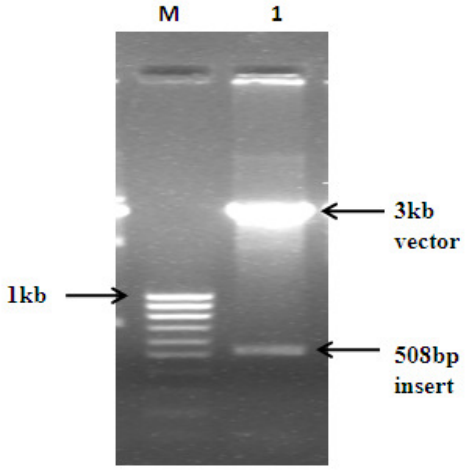


Fig. 3.13 (a) CCoAOMTpro1 insert in pGEM-T easy vector. Lane M: 100 bp ladder. Lane 1: vector backbone and 508 bp vector insert

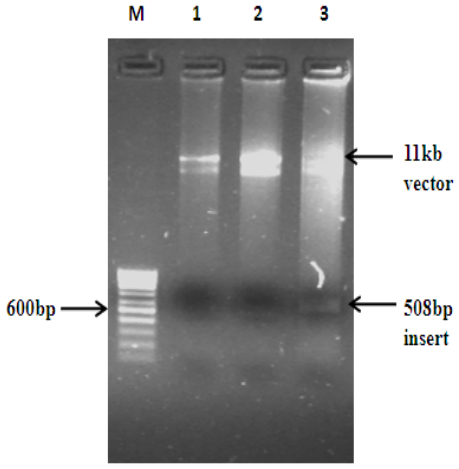


Fig. 3.13 (b) CCoAOMTpro1 insert in pCAMBIA1381Z vector. Lane M: 100 bp ladder. Lanes 1, 2 and 3: vector backbone and 508 bp insert

3.3.6 T-DNA construct of the recombinant vector obtained after cloning the CCoAOMT promoter sequence

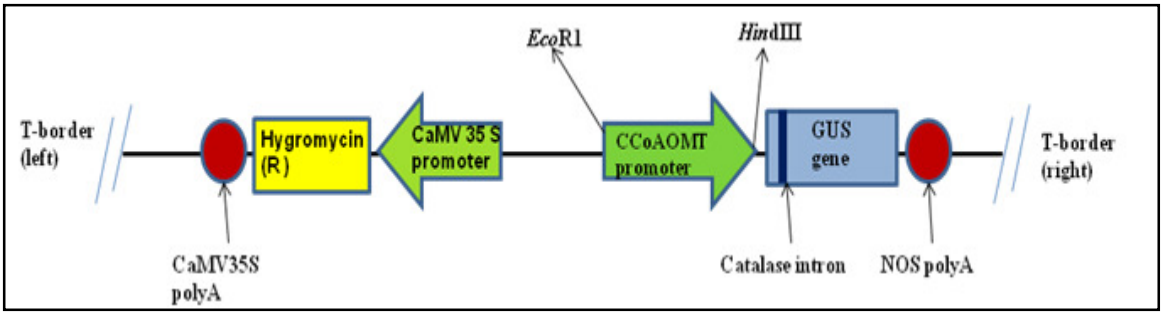


Fig. 3.14: The expression cassette between T-DNA border sequences is composed of a promoterless *gusA* reporter gene with the catalase intron (indigo blue) immediately downstream of the CCoAOMT promoter (green) cloned between the *Eco*R1 and *Hin*dIII sites lying in the MCS region of truncated *lacZα* (functional for blue/white screening;) and a constitutively expressed hygromycin resistance gene (HygR; plant selection maker gene; yellow) placed in a reverse orientation. CaMV35S poly-A and Nos poly-A: polyadenylation signals derived from CaMV or nopaline synthase gene (deep red).

3.3.7 Tissue-specific expression of CCoAOMT promoter driven *gus* gene in Tobacco

The recombinant binary vector having CCoAOMT promoter was used to transform competent *Agrobacterium* strain GV2260 and used for transforming tobacco leaf explants with the T-DNA construct. Simultaneously blank 1381Z vector was also used to transform the *Agrobacterium* strain, to be used as a control for promoter driven GUS expression studies. The genetic transformation experiments were carried out as per the protocol discussed in section 3.2.9. DNA was isolated from the regenerated putative transformants and PCR was performed to amplify *hpt*II gene from putative transformants using *hpt*II gene specific primers listed below:

Primer name	Primer sequence
HygBF	5-GTCGACCTATTTCTTTGCCCTCGGAC-3’
HygBR	5’-GGATCCCCTGACCTATTGCATCTCCC-3’

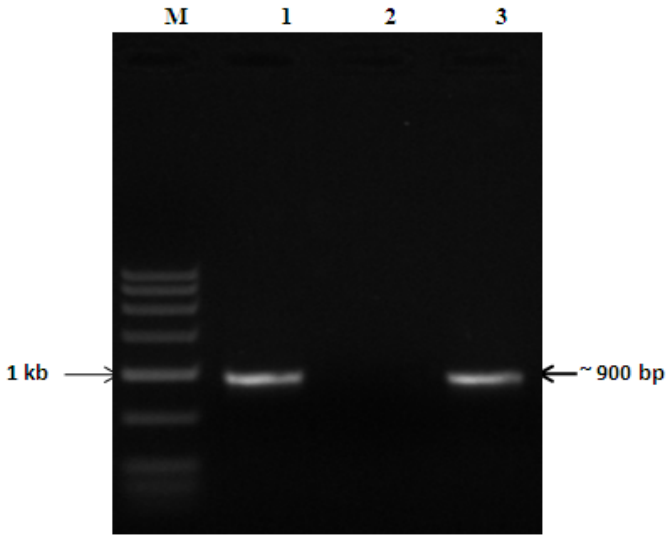


Fig 3.15: Putative transformants showing amplification with *hptII* primers. Lane M: MR DNA ladder; Lanes 1 &3: Putative transformants positive for hygromycin resistance gene and Lane 2: Control tobacco plant (without any construct) showing no amplification with the *hptII* primers.

An amplicon of ~900 bp confirmed the presence of *hptII* (hygromycin phosphotransferase gene) (Fig. 3.15) and the positive plants were used for GUS assay (Fig 3.16).

Moderately strong GUS expression as indicated by a dark blue precipitate, was found in the veins of the CCoAOMT pro::GUS construct transformed leaf (Fig 3.16, A1 and A2) as compared to the stronger expression derived from the 35S::GUS construct, which was not restricted to the veins but diffused over the whole leaf tissue (Fig 3.16, B1 and B2). The leaf tissue arising from the negative control i.e. promoter less::GUS construct presented negligible blue colour after bleaching (Fig 3.16, C1 and C2).

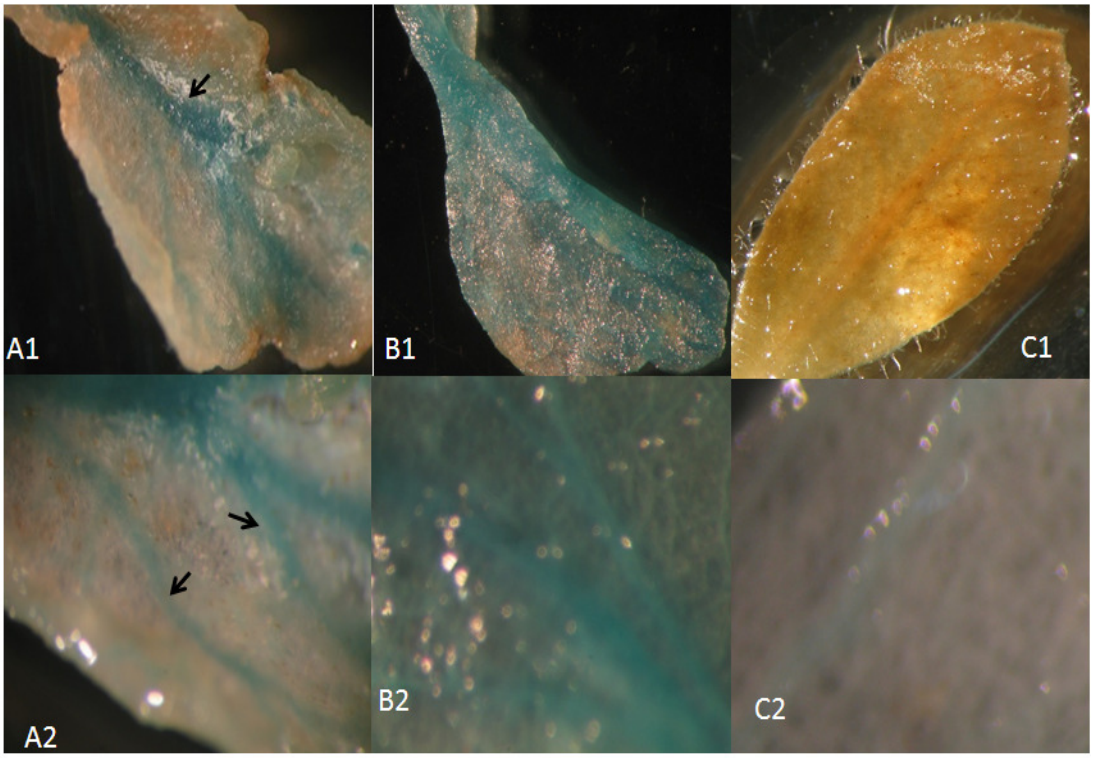


Fig 3.16: Histochemical detection of β-glucuronidase activity in transgenic tobacco leaves. Genetic transformation was performed using *Agrobacterium tumefaciens* strain GV2260 containing the individual constructs mentioned in the picture. Staining was performed after four months of regeneration.**A1, A2:** CCoAOMT pro::GUS construct; **B1, B2:** CaMV35S::GUS construct; **C1,C2**: promoter less::GUS construct.

3.4 Discussion

Manipulation of lignin biosynthetic pathway genes of a plant with minimum side-effects on other tissues requires a vascular-tissue specific promoter, as lignin is mainly deposited in vascular tissues. A targeted expression of transgene primarily in lignifying tissues through the use of a tissue-specific promoter would be more appropriate, as tissue specific promoters keep their specificity across species barriers (Fütterer, 1995a). In order to determine the decisive regulatory regions on a promoter required to confer full activity for driving the expression of a downstream gene, a detailed promoter analysis is required. Preliminary identification of the putative regulatory elements within the upstream region of the gene is performed using certain databases for searching plant promoter elements like PLACE (Higo *et al.* 1999), PlantCARE (Rombauts *et al.*, 1999) and TRANSFAC

(Hehl and Wingender, 2001). However, experimental characterization of the promoter regions, supported by EMSA and *in vivo* studies is required to determine the minimal promoter region required to drive the gene expression in a tissue specific manner. *In vivo* characterization of plant promoters is carried out by generation and analysis of transgenic plants expressing promoter/reporter gene fusion constructs (e. g. Pühringer *et al.* 2005; Puzio *et al.* 2000). Methods for transient expression analysis based on the biolistic system or the transformation of protoplasts have been used as convenient alternatives to stable transformation for the study of the regulation of gene expression (Carle Urioste *et al.* 1995; Russell *et al.*1995). Currently, *Agrobacterium*-based transient assay is being used efficiently and is reproducible for rapid analysis of plant promoters *in vivo* (Yang *et al.* 2000).

In this study four different partial 5' upstream untranslated genomic sequences were analysed for the lignin biosynthetic genes (C4H, CCoAOMT, CCR and CAD) from *L.leucocephala.* As all of these genes are involved in the same pathway, the study was aimed at determining the common cis-acting regulatory elements (CAREs) found on these putative promoter regions which might be co-regulated by certain transcription factors in controlling their tissue specific or temporal expression pattern. The minimum size of the promoter region was available for C4H (195 bp) and the maximum size (1961 bp) for CAD gene. Bioinformatic analysis was done for all the sequences using PLACE database search program and total 9 CAREs were found to be common among all of them. The most important common element was CAATBOX1 which is required for the basic transcriptional assembly. Although, TATA box was found on all the promoters, but different consensus sequences were identified for it on different promoter sequences. Other heterogenous groups of elements identified are unlikely to play a role in the studied promoters considering their function in the promoter where they were identified. This is the case for elements found in histone genes (nonamer motif), seed specific expressed gene (SEF4 binding site) and amylase gene (amylase-box; central element of gibberellin (GA) response complex. The homology of a number of putative *cis* elements in the 4 promoters to previously identified *cis* elements in angiosperm promoters of genes associated with the phenylpropanoid pathway supports the view that there is conservation in signal cascades of the phenylpropanoid pathway between distant species of the plant

kingdom. The most important element, MYBPLANT which is known to interact with MYB transcription factors and coordinating lignin biosynthetic genes was found in limited copies on all the promoter sequences. As the promoter sequences for most of the genes were partial, putative transcription start site could not be predicted for them. So based on the translation start site (TSS), position of the CAREs were determined. The location of occurrence of MYBPLANT element differed in all the gene promoters. In case of CAD it was located more than 1kb upstream of the TSS. However for CCR it was at -115 position and CCoAOMT it was at -201position. The position of the element was most close to the TSS in case of C4H i.e.-12 and -79, which could be present even after the transcription start site.

Based on the *in silico* studies of the promoter elements, determining the functional regions of the promoter were done, initially using EMSA studies, where the promoter regions carrying the MYBPLANT element were used as shown in section 3.3.4.3. Overall, the EMSA results indicate that there are differences between leaf and stem CNPs (Crude nuclear proteins) with respect to their interactions with the promoter regions. Stem CNPs appeared to form a stronger complex with the promoters as compared to the leaf CNPs, probably because of the presence of lignin regulatory proteins higher in stem tissue which is more lignified than the leaf tissues. More elaborate competition assays of the promoter sequences with nuclear protein extracts would clarify the differences in the interactions observed in the study.

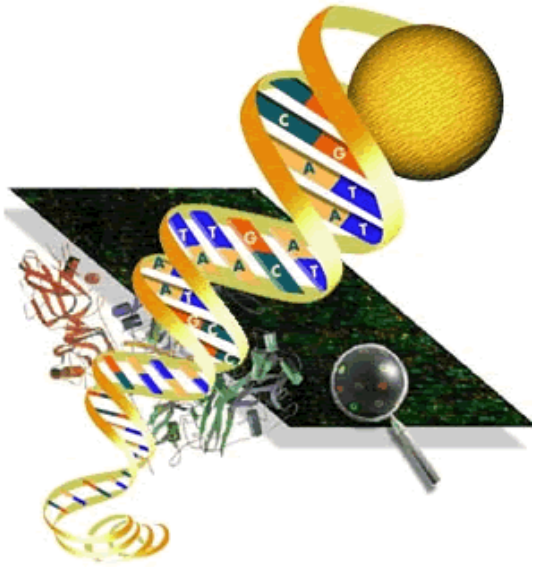
The partial promoter sequence isolated earliest was for the CCoAOMT gene and hence it was used for transient expression studies in tobacco. In order to characterize the partial CCoAOMT promoter, it was fused to the β-glucuronidase (GUS) reporter gene cloned in pCAMBIA-1381Z vector and subsequently analyzed by means of an *Agrobacterium*-mediated transient transformation assay. Histochemical analysis of the promoter driven GUS activity in transgenic tobacco showed that the partial promoter was sufficient to retain the vascular-tissue specific activity, as the GUS activity was restricted to the midrib and veins of the transgenic leaf. Thus, as reported earlier (Grimmig and Matern, 1997), even a minimal promoter region consisting of a TATA box and some enhancer region fused to a reporter gene is capable of driving its expression.

3.5 Conclusions

Promoter sequences for the lignin biosynthetic genes C4H, CCoAOMT, CCR and CAD were isolated using PCR-based approach and characterized using bioinformatic tools, to determine a putative cis-element, MYBPLANT, known to be involved in coordinating the regulation of phenylpropanoid pathway genes in plants. EMSA of the promoter regions carrying some of the important CAREs including MYBPLANT was performed using *L. leucocepohala* CNP extracts prepared from the leaf and stem tissues. The study showed that affinity of stem nuclear proteins was greater than leaf nuclear proteins for all the four putative promoters. Histochemical studies for CCoAOMT promoter driven GUS expression in tobacco showed the localization of GUS activity in the vascular tissues, indicating that even minimal promoter elements were capable of driving tissue-specific gene expression. Thus, CCoAOMT gene promoter may offer a more defined control over the constitutive promoters for future genetic engineering of traits related to lignification in plants, which are confined to the vascular tissues.

It can be concluded that, just like CCoAOMT promoter C4H, CCR and CAD partial promoters may also have a tissue-specific driving activity, although the developmental expression for the same need to be elaborated.

CHAPTER 4



Cloning and characterization of a novel R2R3 type MYB transcription factor gene (*LM*YB_SSM) from *Leucaena leucocephala*

4.1 Introduction

MYB transcription factors and their role in plant developmental processes have been already discussed in Chapter 1, section 1.4. This chapter deals with different strategies followed to fish out the full-length c-DNA clone of a phenylpropanoid pathway related R2R3type MYB gene and its characterization using bioinformatic and experimental tools from *Leucaena leucocephala*.

Isolation of a gene is the pre-requisite for its thorough study or characterization. To isolate any gene of interest generally there are two major approaches as follows:-

- Genomic/cDNA library screening- It is the classical method of gene isolation in which genomic/cDNA libraries are screened with homologous or heterologous probe. After 2-3 rounds of screenings, the plaques showing strong positive signals are excised and the DNA fragment sequenced.
- PCR- It is one of the most popular approaches of gene isolation because of its simplicity and rapidity. Forward and reverse primers are designed on the basis of available sequences in the database. PCR is performed using these primers and genomic/cDNA as template. The amplicon is sequenced to confirm its identity.

Genomic library screening was a lengthy process and resulted unfruitful in previous attempts of isolating genes for CCR and peroxidase in our lab (Shrivastava, 2009 and Abhilash, 2009). Also, its well known that R2R3 type MYB gene family is the largest in plants comprising over 100 members (Douglas *et al*., 2010), so such a technique would result in hundreds of clones and laborious screening for the gene of interest. Hence, the PCR-based approach was used to fish out a novel R2R3-type MYB gene from *Leucaena leucocephala* in the present study.

4.2 Materials and methods

4.2.1 RNA isolation and cDNA first strand synthesis

A good quality and DNA free, total RNA was isolated from *L. leucocephala* xylem tissue of stem (chapter 2, section 2.9.4.5); mRNA was purified from total RNA using oligotex dT resins (chapter 2, section 2.9.4.6) and cDNA Reverse Transcription (RT) 1st strand was synthesized (chapter 2, section 2.9.4.8).

4.2.2 Primer designing and Polymerase Chain Reaction

Degenerate forward and reverse primers were designed based on multiple sequence alignment of the available R2R3-type MYB gene sequences in the NCBI database which were used for PCR amplification using cDNA 1st strand or genomic DNA as template (chapter 2, section 2.9.4.9). Amplified PCR products were eluted from Agarose gel (chapter 2, section 2.9.4.4). The eluted PCR products were cloned in pGEM-T Easy cloning vector.

4.2.3 Rapid Amplification of cDNA Ends (RACE)

Based on the partial sequence of MYB gene obtained using previous step, forward and reverse gene specific primers (GSP) were designed for 3' and 5' RACE respectively (chapter 2, section 2.9.4.10) to fish out full length gene with its 5' and 3' UTRs. The RACE products were gel eluted and cloned in pGEM-T Easy cloning vector.

4.2.4 Sequencing

DNA sequencing was performed with the ABI Prism Big- Dye Terminator Cycle Sequencing Kit on the ABI Prism 3730 DNA analyzer (Applied Biosystems) at Genome Bio Biotech Pvt. Ltd., Pune, Maharashtra, India.

4.2.5 Bioinformatic analysis

Nucleotide and amino acid sequence analysis was done using software pDRAW 32, Clustal W 2.0.1 and online bioinformatics analysis facility available at www.justbio.com, www.expasy.org and www.ncbi.nlm.nih.gov. Multiple alignments of the amino acid sequences were carried out with the Clustal W 2.0.1 program (http://www.ebi.ac.uk/clustalw/). The evolutionary history was inferred using the

Neighbor-Joining method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed (Felsenstein, 1985). All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). Phylogenetic analyses were conducted in MEGA4 (Tamura *et al*., 2007). Conserved motifs were identified using MEME software (Bailey and Elkan, 1994) with required parameters settings.

4.2.6 Estimation of *Ll*MYB_SSM gene copy number

Genomic DNA was isolated from *L. leucocephala* using Lodhi *et al.* (1994) method. Gene copy number was estimated by Southern hybridization. DNA was digested with three restriction enzymes, *Eco* RI, *Hin*dIII and *Sac*I. Two of these enzymes, *Eco*RI and *Hin*dIII cut once in the carboxyl-domain of the gene while *Sac*I cuts thrice within the R2R3-MYB domain of the gene. Southern hybridization was done using two probes, one from the conserved domain and other from the most divergent C-terminal region of the gene. Hybridization was performed at 60 °C temperature (Sambrook *et al.*, 1989).

4.2.6.1 Genomic DNA extraction

Genomic DNA was extracted from *L. leucocephala* using the method given by Lodhi *et al*. (chapter 2, section 2.9.4.2).

4.2.6.2 Restriction digestion of DNA

Genomic DNA restriction digestions were set up as per manufacturer's (Promega, USA; NEB, UK; Amersham, USA) recommendations (chapter 2, section 2.9.4.3).

4.2.6.3 Southern hybridization

Southern hybridization was done (chapter 2, section 2.9.9.1) to estimate the copy number of *Ll*MYB_SSM gene in *L. leucocephala*.

4.3 Results

4.3.1 Multiple sequence alignment of nucleotide sequences of reported R2R3 MYB proteins

All known plant R2R3MYB proteins contain highly conserved stretches of amino acid residues within recognition helices of the R2 and R3 repeats from which R2R3-MYB specific mixtures of oligonucleotide primers can be derived (Avila *et al*., 1993). With the aim of isolating a lignin specific R2R3-MYB gene from *L. leucocephala*, the GenBank nucleotide database was searched for lignin-related MYB genes from Fabaceae family, to design the primer sequences. However, due to lack of functionally characterized R2R3-MYB sequences from the Fabaceae family, tree species from unrelated families, like *Poplar* and *Eucalyptus* were searched for MYB genes related to lignification. A protein BLAST against one of the retrieved sequences resulted in few other functionally related R2R3MYB genes with a significant homology in both the R2 and R3 domains. Few of the nucleotide sequences were retrieved and aligned using Clustal W 2.0.1 tool (Fig. 4.1). Although the sequences were highly divergent in the carboxy-terminal, they were significantly conserved in the Myb DNA-binding domains. So, a set of degenerate forward and reverse primers were designed from the aligned sequences as shown in the figure. The forward primer was designed from the ATG region of R2 domain and reverse primer was designed from the R3 domain (shown within boxes in Fig. 4.1).

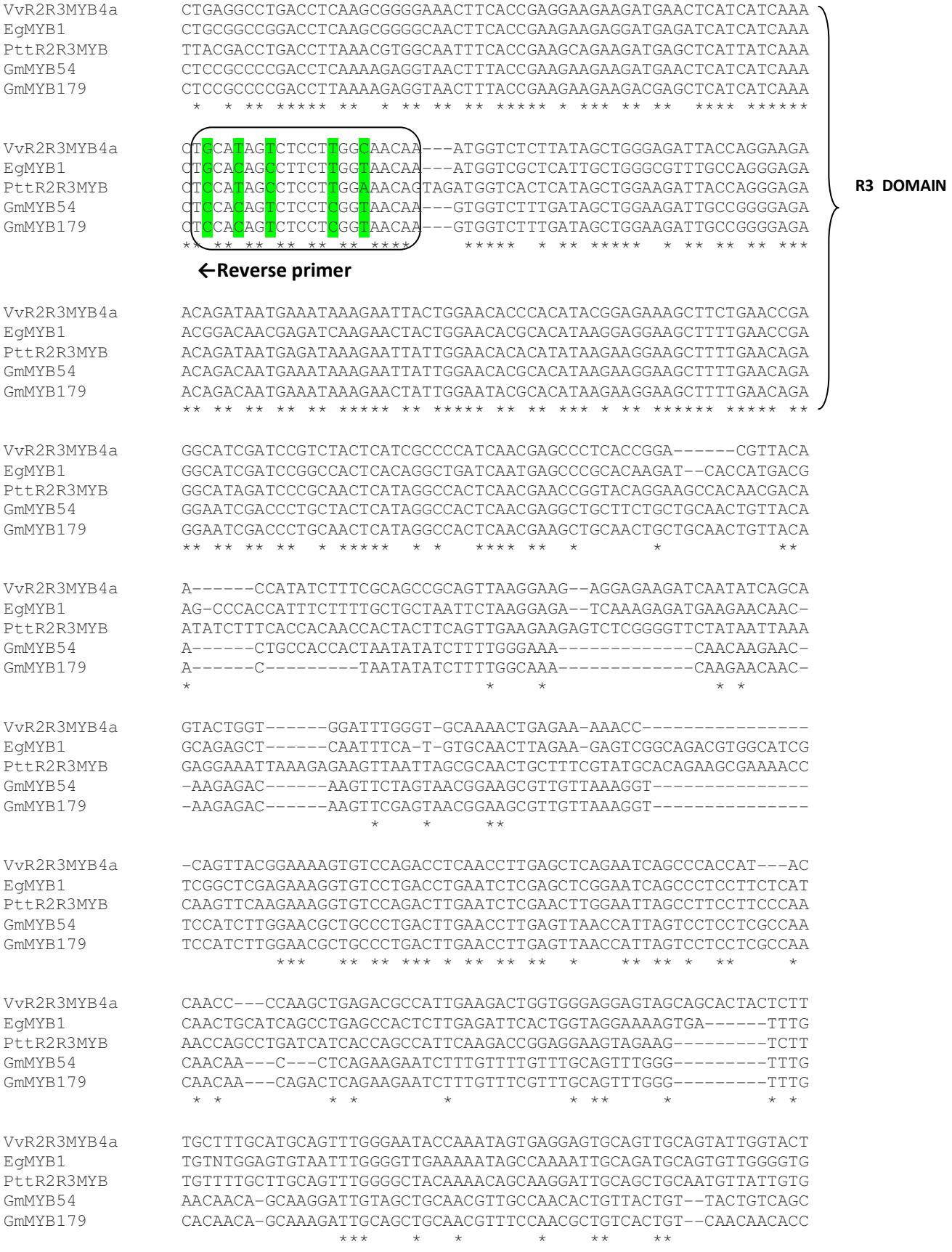
Forward primer→

```
VvR2R3MYB4a     ATGGGGAGGTCTCCTTGCTGTGAGAAAGCTCATACAAACAAAGGGGCATGGACCAAGGAG
EgMYB1          ATGGGAAGGTCTCCTTGCTGCGAGAAGGCTCACACAAACAAGGGCGCATGGACCAAGGAG
PttR2R3MYB      ATGGGAAGGTCTCCTTGCTGTGAAAAAGCCCATACAAACAAGGGTGCGTGGACCAAGGAG
GmMYB54         ATGGGAAGGTCCCCTTGCTGTGAGAAAGCTCACACAAACAAAGGTGCATGGACTAAAGAA
GmMYB179        ATGGGAAGGTCCCCTTGCTGTGAGAAAGCTCACACAAACAAAGGTGCATGGACTAAAGAA
                ***** ***** ******** ** ** ** ** ******** ** ** ***** ** **

VvR2R3MYB4a     GAAGATGATCGCCTCATCGCTTATATCCGGGCACACGGCGAGGGCTGCTGGAGGTCTCTC
EgMYB1          GAGGACGACAAGCTCATTGCCTACATAAGAGCGCACGGCGAGGGTTGCTGGCGGTCGCTC
PttR2R3MYB      GAAGACGATCGCCTTGTTGCTTACATTAGAGCTCATGGTGAAGGTTGCTGGCGTTCACTT
GmMYB54         GAAGATGACAGACTCATATCTTATATTCGAGCTCACGGAGAAGGCTGCTGGCGTTCACTC
GmMYB179        GAAGATGACAGACTCATATCTTATATTCGAGCTCACGGCGAAGGCTGCTGGCGTTCACTC
                ** ** **    **  *  * ** **  * ** ** ** ** ** ****** * ** **

VvR2R3MYB4a     CCCAAGGCCGCAGGCCTTCTCCGATGTGGGAAAAGTTGCCGCCTCCGATGGATAAACTAC
EgMYB1          CCGAAGGCCGCGGGCCTCCTCCGCTGTGGCAAGAGCTGCCGCCTCCGGTGGATCAATTAC
PttR2R3MYB      CCTAAAGCCGCTGGCCTTCTTAGATGTGGCAAGAGTTGCAGACTTCGCTGGATCAACTAC
GmMYB54         CCCAAAGCCGCCGGCCTTCTCCGGTGCGGCAAGAGCTGCCGTCTCCGGTGGATCAACTAC
GmMYB179        CCCAAAGCCGCCGGTCTTCTCCGGTGCGGCAAAAGCTGCCGTCTCCGGTGGATCAACTAC
                ** ** ***** ** ** **  * ** ** ** ** *** * ** ** ***** ** ***
```

R2 DOMAIN


VvR2R3MYB4a CTGAGGCCTGACCTCAAGCGGGGAAACTTCACCGAGGAAGAAGATGAACTCATCATCAAA
EgMYB1 CTGCGGCCGGACCTCAAGCGGGGCAACTTCACCGAAGAAGAGGATGAGATCATCATCAAA
PttR2R3MYB TTACGACCTGACCTTAAACGTGGCAATTTCACCGAAGCAGAAGATGAGCTCATTATCAAA
GmMYB54 CTCCGCCCCGACCTCAAAAGAGGTAACTTTACCGAAGAAGAAGATGAACTCATCATCAAA
GmMYB179 CTCCGCCCCGACCTTAAAAGAGGTAACTTTACCGAAGAAGAAGACGAGCTCATCATCAAA
VvR2R3MYB4a CTGCATAGTCTCCTTGGCAACAA---ATGGTCTCTTATAGCTGGGAGATTACCAGGAAGA
EgMYB1 CTGCACAGCCTTCTTGGTAACAA---ATGGTCGCTCATTGCTGGGCGTTTGCCAGGGAGA
PttR2R3MYB CTCCATAGCCTCCTTGGAAACAGTAGATGGTCACTCATAGCTGGAAGATTACCAGGGAGA
GmMYB54 CTCCACAGTCTCCTCGGTAACAA---GTGGTCTTTGATAGCTGGAAGATTGCCGGGGAGA
GmMYB179 CTCCACAGTCTCCTCGGTAACAA---GTGGTCTTTGATAGCTGGAAGATTGCCGGGGAGA
R3 DOMAIN
←Reverse primer
VvR2R3MYB4a ACAGATAATGAAATAAAGAATTACTGGAACACCCACATACGGAGAAAGCTTCTGAACCGA
EgMYB1 ACGGACAACGAGATCAAGAACTACTGGAACACGCACATAAGGAGGAAGCTTTTGAACCGA
PttR2R3MYB ACAGATAATGAGATAAAGAATTATTGGAACACACATATAAGAAGGAAGCTTTTGAACAGA
GmMYB54 ACAGACAATGAAATAAAGAATTATTGGAACACGCACATAAGAAGGAAGCTTTTGAACAGA
GmMYB179 ACAGACAATGAAATAAAGAACTATTGGAATACGCACATAAGAAGGAAGCTTTTGAACAGA
VvR2R3MYB4a GGCATCGATCCGTCTACTCATCGCCCCATCAACGAGCCCTCACCGGA------CGTTACA
EgMYB1 GGCATCGATCCGGCCACTCACAGGCTGATCAATGAGCCCGCACAAGAT--CACCATGACG
PttR2R3MYB GGCATAGATCCCGCAACTCATAGGCCACTCAACGAACCGGTACAGGAAGCCACAACGACA
GmMYB54 GGAATCGACCCTGCTACTCATAGGCCACTCAACGAGGCTGCTTCTGCTGCAACTGTTACA
GmMYB179 GGAATCGACCCTGCAACTCATAGGCCACTCAACGAAGCTGCAACTGCTGCAACTGTTACA
VvR2R3MYB4a A------CCATATCTTTCGCAGCCGCAGTTAAGGAAG--AGGAGAAGATCAATATCAGCA
EgMYB1 AG-CCCACCATTTCTTTTGCTGCTAATTCTAAGGAGA--TCAAAGAGATGAAGAACAAC-
PttR2R3MYB ATATCTTTCACCACAACCACTACTTCAGTTGAAGAAGAGTCTCGGGGTTCTATAATTAAA
GmMYB54 A------CTGCCACCACTAATATATCTTTTGGGAAA-------------CAACAAGAAC-
GmMYB179 A------C----------TAATATATCTTTTGGCAAA-------------CAAGAACAAC-
VvR2R3MYB4a GTACTGGT------GGATTTGGGT-GCAAAACTGAGAA-AAACC----------------
EgMYB1 GCAGAGCT------CAATTTCA-T-GTGCAACTTAGAA-GAGTCGGCAGACGTGGCATCG
PttR2R3MYB GAGGAAATTAAAGAGAAGTTAATTAGCGCAACTGCTTTCGTATGCACAGAAGCGAAAACC
GmMYB54 -AAGAGAC------AAGTTCTAGTAACGGAAGCGTTGTTAAAGGT---------------
GmMYB179 -AAGAGAC------AAGTTCGAGTAACGGAAGCGTTGTTAAAGGT---------------
VvR2R3MYB4a -CAGTTACGGAAAAGTGTCCAGACCTCAACCTTGAGCTCAGAATCAGCCCACCAT---AC
EgMYB1 TCGGCTCGAGAAAGGTGTCCTGACCTGAATCTCGAGCTCGGAATCAGCCCTCCTTCTCAT
PttR2R3MYB CAAGTTCAAGAAAGGTGTCCAGACTTGAATCTCGAACTTGGAATTAGCCTTCCTTCCCAA
GmMYB54 TCCATCTTGGAACGCTGCCCTGACTTGAACCTTGAGTTAACCATTAGTCCTCCTCGCCAA
GmMYB179 TCCATCTTGGAACGCTGCCCTGACTTGAACCTTGAGTTAACCATTAGTCCTCCTCGCCAA
VvR2R3MYB4a CAACC---CCAAGCTGAGACGCCATTGAAGACTGGTGGGAGGAGTAGCAGCACTACTCTT
EgMYB1 CAACTGCATCAGCCTGAGCCACTCTTGAGATTCACTGGTAGGAAAAGTGA------TTTG
PttR2R3MYB AACCAGCCTGATCATCACCAGCCATTCAAGACCGGAGGAAGTAGAAG---------TCTT
GmMYB54 CAACAA---C---CTCAGAAGAATCTTTGTTTTGTTTGCAGTTTGGG---------TTTG
GmMYB179 CAACAA---CAGACTCAGAAGAATCTTTGTTTCGTTTGCAGTTTGGG---------TTTG
VvR2R3MYB4a TGCTTTGCATGCAGTTTGGGAATACCAAATAGTGAGGAGTGCAGTTGCAGTATTGGTACT
EgMYB1 TGTNTGGAGTGTAATTTGGGGTTGAAAAATAGCCAAAATTGCAGATGCAGTGTTGGGGTG
PttR2R3MYB TGTTTTGCTTGCAGTTTGGGGCTACAAAACAGCAAGGATTGCAGCTGCAATGTTATTGTG
GmMYB54 AACAACA-GCAAGGATTGTAGCTGCAACGTTGCCAACACTGTTACTGT--TACTGTCAGC
GmMYB179 CACAACA-GCAAAGATTGCAGCTGCAACGTTTCCAACGCTGTCACTGT--CAACAACACC

```
VvR2R3MYB4a      AGTA------------GTGGAAGCAGCAGCTC------TGGGTATGACTTCTTAGGGTTG
EgMYB1           ATCG------------AGAGTGAAACTAGTGT------TGGGTATGACTTCTTGGGCTTG
PttR2R3MYB       AGCACTGTTGGGAGCAGTGGCAGCACTAGCACAAAGAATGGCTATGACTTCTTGGGCATG
GmMYB54          AACACTACTCCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTTATGATTTCTTGGGCATG
GmMYB179         A------------------------------------------------------------
                 *

VvR2R3MYB4a      ACATCT---GGGGTTTTGGATTACAGAGGTTTGGAGATGAAATAA
EgMYB1           AAGGCA---AGTGTTTTGGATTATAGGAGCTGA------------
PttR2R3MYB       AAAAGT---GGTGTTTTGGATTATAGAAGTTTAGAGATGAAATAA
GmMYB54          AAAACCAACGGTGTTTGGGATTGCACCCGCTTGGAAATGAAATGA
GmMYB179         -------------------------------------------
```

Fig. 4.1: Multiple sequence alignment of nucleotide sequences of phenylpropanoid pathway related R2R3 MYB genes. GenBank Accession numbers of the R2R3MYB genes from different plant species: *Vitis* (*Vv*R2R3MYB4a: XM_002278186), *Eucalyptus* (EgMYB1:CAE09058), *Poplar* (PttR2R3MYB: AJ567346.1), *Glycine* (GmMYB179: DQ822961.1) and (GmMYB54:DQ822881) The regions of primer designing are shown within boxes and the regions where degeneracy was incorporated are shaded.

4.3.2. Forward and Reverse Primers

As evident from the multiple sequence alignment of different R2R3 MYB genes, the sequences are highly divergent in the trans-activation region outside the MYB domain. However, significant homology is present in the R2 and R3 domains of the gene sequences. Hence, a few sets of degenerate primers were designed from those domains. The primer combination which worked for *L. leucocephala*, have been indicated in Fig. 4.1 and the primer sequences are as follows:

Primer name	**Primer sequence**
Forward primer (Sum Myb Star)	5'-ATGGG(G/A)AGGTC(T/C)CCTTGCT-3'
Reverse primer (MybGR)	5'-TTGTT(A/G/T)CC(G/A)AGGAG(A/G)CT(A/G)TG(G/C) A-3'

4.3.3 Isolation of total RNA from *L. leucocephala*, its cDNA synthesis and PCR

RNase free environment was created and maintained as described by Blumberg (1987). The glassware and plasticware were treated overnight with DEPC (0.1% in water) and autoclaved. The pestle and mortar were also DEPC treated and then baked at 100 ºC for 6-8 h.

Total RNA was isolated from xylem tissue of field grown plants or from *in vitro* grown seedlings of *L. leucocephala* (as mentioned in chapter 2, section 2.9.4.5). The quantity and integrity of the isolates was checked by Agilent 2100 Bioanalyser system and also by agarose gel electrophoresis (Fig. 4.2). Concentration of the RNA was estimated using Nanodrop. 1 µg of total RNA was used for cDNA 1st strand synthesis (as mentioned in Chapter 2, Section 2.9.4.8) using MMLV-RT based SuperscriptIII reverse transcriptase system (Invitrogen).

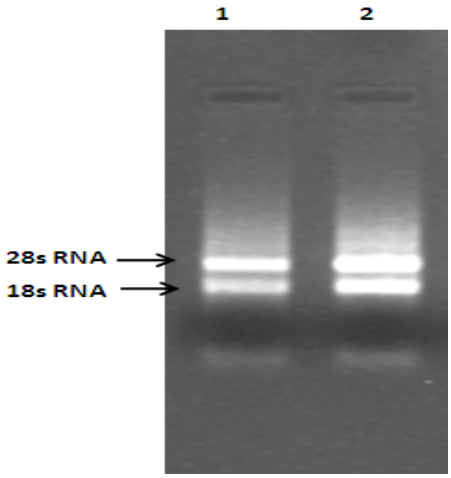


Fig. 4.2 Electrophorogram of RNA isolation from *L. leucocephala* run on 1% agarose gel.
Lanes 1 and 2: Total RNA from xylem tissue of matured plant and 30 day seedling stem respectively.

4.3.4 PCR amplification of partial cDNA fragments of R2R3 type MYB gene from *L. leucocephala*

Various primer sets were used for amplification of a partial MYB gene sequence from the cDNA pool of the xylem tissue of *L.leucocephala*. However, only one set of primers designed from the highly conserved N-terminal domain of the gene resulted in an amplification of ~250bp fragment (Fig. 4.3a) which corroborated to the band size expected from the primer combination used.

The amplified fragment was eluted from the agarose gel using gel extraction kit (Axyprep, USA) and cloned into pGEM-T easy vector (Promega, USA). Recombinant clones were confirmed by colony PCR and *Eco*RI digestions which released a ~ 250 bp

insert (Fig. 4.3b). Sequencing of the positive clones resulted in a single type of sequence of 263 bp, which was assigned **MYB*Leu*1** for reference.

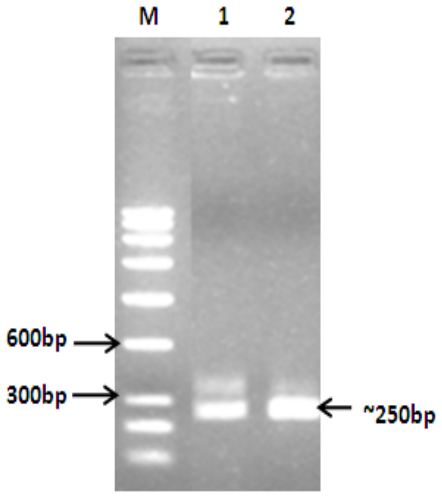


Fig. 4.3(a) PCR amplicon of MYB*Leu*1 on 1% agarose gel. Lane M: LR ladder; Lanes1&2: ~250 bp cDNA fragment.

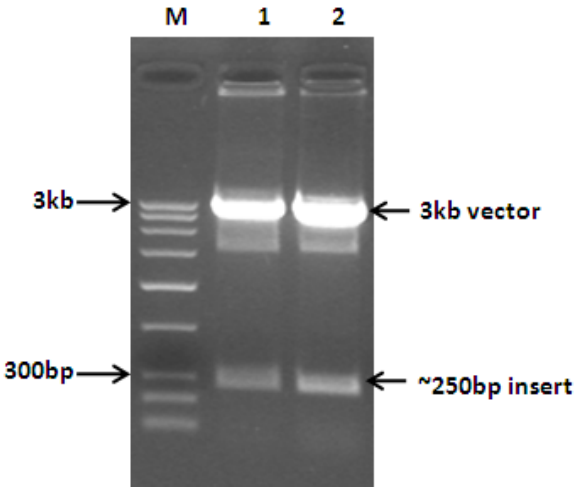


Fig. 4.3(b) MYB*Leu*1 insert in pGEM-T easy vector. Lane M: LR DNA ladder, Lanes 1&2: 3kb vector backbone and ~250 bp insert in clones 1&2 respectively

4.3.4.1 MYB*Leu*1 sequence analysis

MYB*Leu*1 nucleotide sequence was subjected to NCBI nucleotide blast, which shows 84-86% identities and about 85% sequence similarity with reported R2R3-type MYB gene sequences among various plants. The deduced amino acid sequence shows maximum sequence identity of 86% with MYB54 gene from *Glycine max*. *At*MYB4 (*Arabidopsis*), *Vv*MYB4a (*Vitis*), ZmMYB31 (*Zea mays*) and EgMYB1 (*Eucalyptus*) which are functionally known repressors of phenylpropanoid pathway showed 83- 84% identity at nucleotide level.

Forward primer→

```
  1 ATGGGGAGAG CTCCGTGCTG CGAGAAAGCT CACACTAACA AAGGAGCATG
 51 GACCAAGGAA GAAGACGATC GTCTCGTTGC TTATATTCGA GCTCACGGCG
101 AGGGCTGCTG GCGCTCTCTC CCCAAAGCCG CCGGCCTCCT CCGCTGTGGC
151 AAGAGCTGCC GTCTCCGCTG GATCAATTAC CTCCGACCAG ACCTCAAACG
201 CGGCAATTTT ACCGAAGAAG AAGATGAGCT CATCATCAAG CTCCATAGTC
251 TCCTCGGCAA CAA
```

←Reverse primer

Fig. 4.4: Nucleotide sequence (263 bp) of partial MYB *Leu*1, showing the forward and reverse primers

4.3.4.2 Predicted protein sequence for MYB*Leu*1:

Based on translation results obtained using Expasy proteomics tool protein sequence of 87 amino acids was retrieved and putative conserved domains (complete and a partial SANT domain, representing the R2 and R3 repeats of an R2R3-type MYB gene) were identified (Fig. 4.6). The three regularly spaced tryptophan residues (positions 17, 37 and 57) present in R2 repeat of animal MYB proteins and known to be important in maintaining the hydrophobic core of the DNA-binding domain were also located. A phenylalanine residue (70th position) representing the partial R3 domain was also located.

M G R A P C C E K A H T N K G A **W** T K E E D D R L V A Y I
R A H G E G C **W** R S L P K A A G L L R C G K S C R L R **W** I
N Y L R P D L K R G N **F** T E E E D E L I I K L H S L L G N

Fig 4.5 Translated amino acid sequence of MYB *Leu*1, showing the conserved tryptophan (W) residues and a phenylalanine (F) residue in bold.

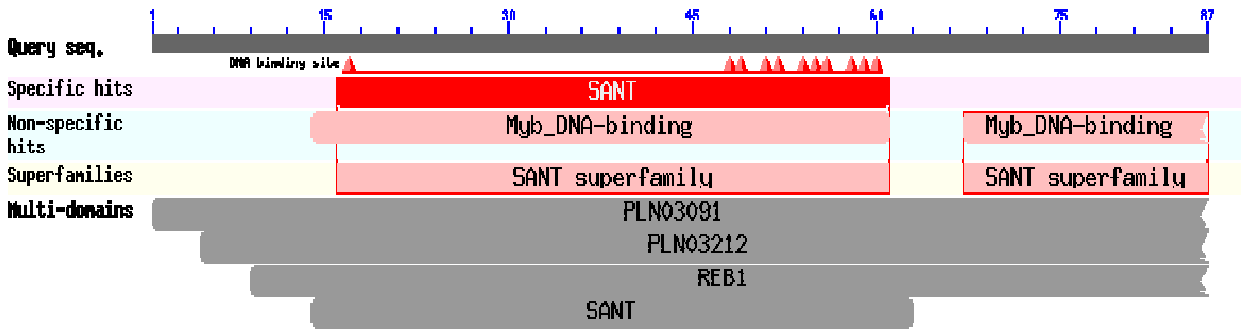


Fig. 4.6 Putative conserved domains identified on the translated MYB *Leu*1

SANT domain: 'SWI3, ADA2, N-CoR and TFIIIB' DNA-binding domains. Tandem copies of the domain bind telomeric DNA tandem repeats as part of the capping complex. Binding is sequence dependent for repeats which contain the G/C rich motif. The domain is also found in regulatory transcriptional repressor complexes where it also binds DNA.

Myb-DNA-binding domain: pfam00249; This family contains the DNA binding domains from Myb proteins, as well as the SANT domain family.

PLN03091 and PLN03212: Classified as a model that may span more than one domain and not assigned to any domain superfamily.

REB1: COG5147, Myb superfamily proteins, including transcription factors and mRNA splicing factors [Transcription / RNA processing and modification / Cell division and chromosome partitioning]

4.3.4.3 Alignment of amino acid sequence of MYB*Leu*1 with known R2R3 type-MYB sequences

Protein BLAST of the amino acid sequence showed 95-99 % identity with other reported R2R3-type MYBs (Fig. 4.7).This shows that the DNA-binding domain sequence is highly conserved in functionally related MYBs. MYB*Leu*1 showed the highest identity of 99% with *Am*MYB308 (*Antirrhinum majus*), which is reported to be a repressor of lignin biosynthetic pathway.

```
MYBLeu1      MGRAPCCEKAHTNKGAWTKEEDDRLVAYIRAHGEGCWRSLPKAAGLLRCGKSCRLRWINY 60
AmMYB308     MGRSPCCEKAHTNKGAWTKEEDDRLVAYIRAHGEGCWRSLPKAAGLLRCGKSCRLRWINY 60
VvMYB4a      MGRSPCCEKAHTNKGAWTKEEDDRLIAYIRAHGEGCWRSLPKAAGLLRCGKSCRLRWINY 60
GmMYB54      MGRSPCCEKAHTNKGAWTKEEDDRLISYIRAHGEGCWRSLPKAAGLLRCGKSCRLRWINY 60
GhMYB9       MGRSPCCEKAHTNKGAWTKEEDDRLIAYIRAHGEGCWRSLPKAAGLLRCGKSCRLRWINY 60
EgMYB1       MGRSPCCEKAHTNKGAWTKEEDDKLIAYIRAHGEGCWRSLPKAAGLLRCGKSCRLRWINY 60
             ***:*******************:*::*********************************
                                    R2 DOMAIN

MYBLeu1      LRPDLKRGNFTEEEDELIIKLHSLLGN 87
AmMYB308     LRPDLKRGNFTEEEDELIIKLHSLLGN 87
VvMYB4a      LRPDLKRGNFTEEEDELIIKLHSLLGN 87
GmMYB54      LRPDLKRGNFTEEEDELIIKLHSLLGN 87
GhMYB9       LRPDLKRGNFTEEEDELIIKLHSLLGN 87
EgMYB1       LRPDLKRGNFTEEEDEIIIKLHSLLGN 87
             ****************:**********
                  Partial R3 DOMAIN
```

Fig. 4.7 Amino acid sequence alignment of MYB*Leu*1 with other functionally characterized MYBs. GenBank Accession Numbers: *Am*MYB308 (P81393), *Vv*MYB4a (EF113078), *Gm*MYB54 (DQ82288), *Gh*MYB9 (AAK19619), *Eg*MYB1 (AJ576024).

4.3.5. PCR for 3’ RACE reaction

The protocol was followed as per the Invitrogen’s Gene Racer kit instructions (chapter 2, section 2.9.4.10) to fish out 3’ end of the partial MYB gene. The partial MYB cDNA sequence (MYB*Leu*1) begins with the putative translation start site of an R2R3-type MYB gene, as determined in the nucleotide as well as Protein BLAST hits. So, it was first necessary to isolate the 3’ end of the gene in order to get a complete R2R3-type MYB gene sequence. As, the partial sequence was highly conserved, two sets of forward primers were designed from the sequence portion representing the R3 repeat to be used in combination with 3’ Gene racer primers for fishing out 3’UTR of the MYB gene :

Primer name	Primer sequence
MYBGF1	5'- ACCTCAAACGCGGCAATTTTACCGAA -3'
MYB GF2	5'- CTCCATAGTCTCCTCGGCAACAA-3'

The 3'RACE reaction was performed as discussed in Chapter 2, section 2.9.4.10, using 3'RACE ready cDNA as template. Primary reaction using MYBGF1 and Gene Racer primer resulted in no specific amplified product, but a smear; so it was diluted and further re-amplified using 3' nested universal primer (3'NUP- is provided with kit) and both MYBGF1 and MYBGF2. However, an amplification of desired size was obtained with MYBGF1 itself. The resultant PCR product of approximately 700 bp (Fig 4.8a) was cloned into pGEM-T easy vector (Promega, USA) and the recombinant construct was confirmed by *Eco*RI restriction analysis (Fig 4.8 b) and by sequencing. Sequencing of the 10 positive clones resulted in atleast two different types of clones, which were 95% similar in the R3 domain (comprising 60 amino acids) at protein level but highly divergent in the remaining C-terminal region. Nucleotide and protein BLAST for the two types of sequences resulted in one of the clones being functionally related to other database sequences which are reported to be phenylpropanoid pathway repressors. This clone was of 738 bp and assigned **MYB*Leu*3'RACE** for reference and a reverse STOP primer was designed to fish out a full length MYB cDNA clone, from the 3'RACE cDNA pool.

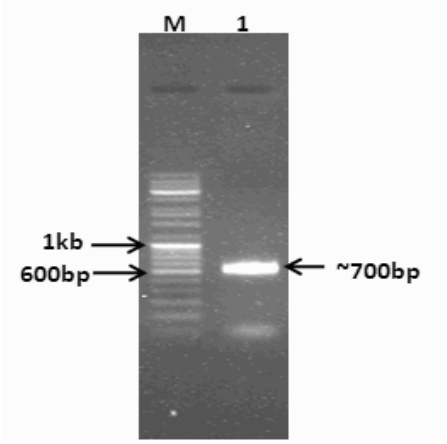


Fig. 4.8a: PCR amplicon of MYB*Leu*3'RACEon 1% agarose gel. Lane M: MR ladder; Lane1: ~700 bp 3'RACE cDNA fragment

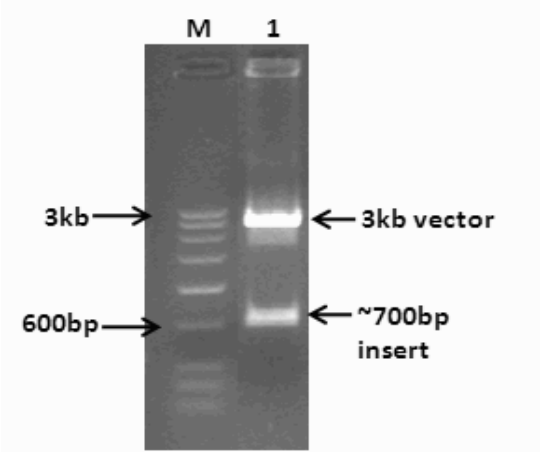


Fig. 4.8b: MYB*Leu*3'RACE insert in pGEM-T easy vector. Lane M: LR DNA ladder 3kb vector backbone and ~700 bp insert

4.3.5.1 MYB*Leu*3'RACE sequence analysis

The analysis of the sequenced 3'RACE product revealed a partial coding sequence of 509 bases, followed by a stop codon (TGA) identified using expasy translation tool and 203 bp of 3'UTR region (Fig 4.9). A putative polyadenylation signal sequence was also detected on the UTR sequence, which is shown shaded. A reverse Stop primer was designed from the underlined region in the sequence to fish out a full length MYB gene from the 3'RACE cDNA pool.

Forward nested GSP→

```
  1 ACCTCAAACG CGGCAATTTT ACCGAAGAAG AAGATGAGCT CATCATCAAG
 51 CTCCATAGTT TCCTCGGCAA CAAGTGGTCG TTGATAGCCG GGAGATTACC
101 GGGAAGAACG GACAACGAGA TAAAGAATTA CTGGAATACT CATATAAGAA
151 GGAAGCTTCT GAGCAGAGGA ATTGACCTTA CGACTCATAG GCCTTTCAAC
201 GATGACTTTC AGAACCAATC CCATGCCTCA TTTGCTTCTG CTGCTAAACA
251 GCAGGAACCG GAGATTCTAA TTACAGATGA GAATTCCAAA CGACCCGGCG
301 TCTTAGAAGT ATGCCCCGAC TTGAACCTCG AGCTAACCAT CAGTTTTCCC
351 CACCACCATC ATCAGAAGCA ACCAGATCTG CCATCGCTCA AGAGCATTAA
401 TAGTGGAAGC AGTACGGCTG GCGATGGTTA TGACTCCTTG GGATTGAAAA
451 GCGGTGTTTT GGATTGCAGG AGCAGCGTAA CCCAAATAGA TTGGAGAGCA
501 CTCACAGAAT GAtttgtgtg tacccatcat tttaaagttc tcttctttct
551 ttcttcgatt gtataatttg ttaattacga tgcatatagg atatatatag
601 tatcaggaca tttggatcct cttgttatat ttaaaccatg cgttttttt
651 aacaaagaaa tcaacctgat attcattggt tgttttgttc aaaaaaaaaa
701 aaaaaaaaaa aaaaaCACTG TCATGCCGTT ACGTAGCG
```

←Reverse 3'NUP

Fig. 4.9: Nucleotide sequence (738 bp) of MYB *Leu* 3' RACE product. Forward nested gene specific primer (MYBGF1) and reverse 3'NUP primer are shown in bold; 3' UTR of 203bases is shown in italics lowercase and the STOP codon (TGA) is highlighted in red colour. The primary cleavage site selection in pre- mRNA (CP Joshi, 1987) is highlighted yellow. The reverse STOP primer designed for fishing out the full length gene has been shown underlined.

4.3.5.2 Predicted protein sequence for MYB*Leu* 3' RACE:

Based on translation results obtained using expasy proteomics tool, protein sequence of 164 amino acids (Fig. 4.10 a) was retrieved and putative conserved domains (a second complete SANT domain, representing the R3 repeat of an R2R3-type MYB gene) were identified (Fig 4.10 b). Another non-specific hit, **PLN03091** also identified in MYB*Leu*1

was detected. The three tryptophan residues (W) forming the hydrophobic core of the R2 repeat are replaced by a Phenylalanine (F) residue in the 1st position of the R3 repeat, a characteristic of all the plant R2R3 MYB genes.

F T E E E D E L I I K L H S L L G N K **W** S L I A G R L P G R
T D N E I K N Y **W** N T H I R R K L L S R G I D P T T H R P L
N D D S Q N Q S H A S S A S A A K Q Q E P E I L I T D E D S
K R P G V L E V C P D L N L E L T I S L P H H H H Q K Q P D
L P S L K G I N S G S S T A G D G Y D F L G L K S G V L D C
R S S V T Q I D W R A L T E

Fig 4.10(a): Translated amino acid sequence of MYB *Leu*3'RACE, showing the conserved Phenylalanine (F) and 2 tryptophan (W) residues in bold.

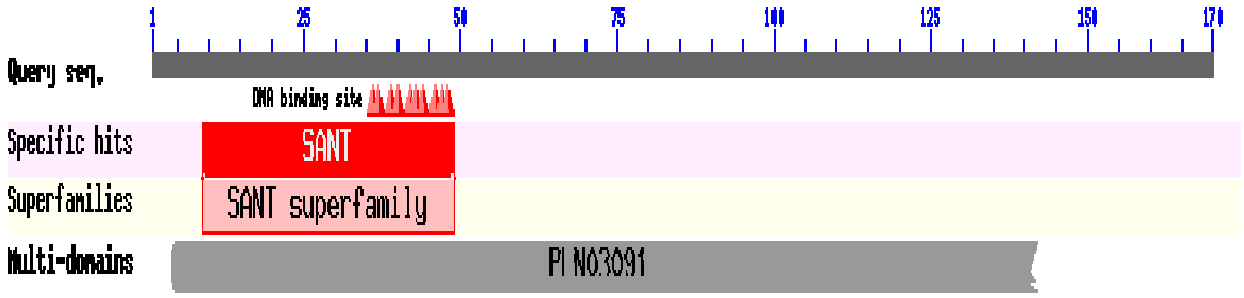


Fig. 4.10(b): Putative conserved domains identified on the translated MYB *Leu* 3' RACE sequence

4.3.5.3 Alignment of amino acid sequence of MYB*Leu* 3' RACE with known R2R3 type-MYB sequences

Protein BLAST of the amino acid sequence showed 85-90 % identity in the R3 domain (1st 60 amino acids) and only 30-40% identity outside the MYB domain with other functionally reported R2R3-type MYBs (Fig 4.11) when aligned using ClustalW2 tool. This further shows that the DNA-binding domain sequence is highly conserved in functionally related R2R3-type MYBs. However, the region outside the MYB-domain also known as the trans-activation domain is quite divergent in different members of the MYB gene family. MYB*Leu* 3' RACE showed the highest similarity of 57% with *Gh*MYB9 (*Gossypium hirsutum*), which is reported to play a role in initiation and elongation of cotton fiber cells.

```
GhMYB9        FTEEEDELIIKLHSLLGNKWSLIAGRLPGRTDNEIKNYWNTHIRRKLLSRGIDPATHRPL 60
VvMYB4a       FTEEEDELIIKLHSLLGNKWSLIAGRLPGRTDNEIKNYWNTHIRRKLLNRGIDPSTHRPI 60
AmMYB308      FTEEEDELIIKLHSLLGNKWSLIAGRLPGRTDNEIKNYWNTHIRRKLLSRGIDPTTHRSI 60
EgMYB1        FTEEEDEIIIKLHSLLGNKWSLIAGRLPGRTDNEIKNYWNTHIRRKLLNRGIDPATHRLI 60
MybLeu3R      FTEEEDELIIKLHSLLGNKWSLIAGRLPGRTDNEIKNYWNTHIRRKLLSRGIDPTTHRPL 60
AtMYB4        FTEEEDELIIKLHSLLGNKWSLIAGRLPGRTDNEIKNYWNTHIRRKLINRGIDPTSHRPI 60
              *******:***************************************:.*****::** :
                <--------------------------------------------->

                            R3 DOMAIN

GhMYB9        NE--ASQD------------VTTISFSGAKEEK---EKINTNSNNNPIGFITKDEKK--- 100
VvMYB4a       NE--PSPD------------VTTISFAAAVKEE---EKINISSTG---GFGCKTEK---- 96
AmMYB308      NDGTASQDQ-----------VTTISFSNANSKE---EDTKHKVAV---DIMIKEEN---- 99
EgMYB1        NEPAQDHHD-----------EPTISFAANSKEI---KEMKNNAEL---NFMCNLEESADV 103
MybLeu3R      ND--DSQNQ-----------SHASSASAAKQQE---PEILITDED------SKRPG---- 94
AtMYB4        QESSASQDSKPTQLEPVTSNTINISFTSAPKVETFHESISFPGKSEKISMLTFKEEKDE- 119
                ::   . .                * :   .      .

GhMYB9        -IPVQERCPDLNLDLRISPPYYQQ--TQPESFKTGGR----TLCFICS--LGVKN-SKDC 150
VvMYB4a       -NPVTEKCPDLNLELRISPPYQPQ--AETP-LKTGGRSSSTTLCFACS--LGIPN-SEEC 149
AmMYB308      -SPVQERCPDLNLDLKISPPCQQQINYHQENLKTGGRNGSSTLCFVCR--LGIQN-SKDC 155
EgMYB1        ASSARERCPDLNLELGISPPSHQLH-QPEPLLRFTGRK---SDLCECN--LGLKN-SQNC 156
MybLeu3R      ---VLEVCPDLNLELTISLPHHHHQ-KQPDLPSLKGINSGSSTAGDGYDFLGLKSGVLDC 150
AtMYB4        -CPVQEKFPDLNLELRISLPDDVDR------LQGHGKS-TTPRCFKCS--LGMIN-GMEC 168
                   . *  *****:* ** *               *     .        **: .   :*

GhMYB9        TCSTITTAAGSSSSSSSHSNSNNSSGYDFLGLK----SGILEYRSLEMK 195
VvMYB4a       SCS-IGTSSGSSSS-----------GYDFLGLT----SGVLDYRGLEMK 182
AmMYB308      SCS---DGVGN-------------------------------------- 163
EgMYB1        RCS--VGVIESETS----------VGYDFLGLK----ASVLDYRS---- 185
MybLeu3R      RSS-----------------------------------VTQIDWRALTE- 164
AtMYB4        RCGRMRCDVVGGSSKGS----DMSNGFDFLGLAKKETTSLLGFRSLEMK 213
          ..
```

Fig. 4.11 Amino acid sequence alignment of MYB*Leu* 3' RACE with other functionally characterized MYBs. GenBank Accession Numbers: *Am*MYB308 (P81393), *Vv*MYB4a (EF113078), *Gh*MYB9 (AAK19619), *Eg*MYB1 (AJ576024), AtMYB4 (AF062860).

4.3.6 Isolation of full length coding region of R2R3 type MYB gene from *L. leucocephala.*

Based on the forward START primer designed using the partial MYB*Leu*1sequence and reverse STOP primer designed based on MYB*Leu* 3' RACE sequence, a full length R2R3 type-MYB gene was fished out from the 3'RACE cDNA pool.

Primer name	**Primer sequence**
Sum Myb Star	5'ATGGGGAGAGCTCCGTGCT 3'
MYBStop	5'-TCATTCTGTGAGTGCTCTCCAATCT-3'

The resultant PCR product of approximately 700 bp (Fig. 4.12 a) was cloned into pGEM-T Easy vector (Promega, USA) and the recombinant construct was confirmed by *Eco*RI restriction analysis (Fig 4.12b), which resulted in two bands of ~200 and ~500bp, indicating the presence of the site in the gene sequence. Sequencing of 8 positive clones resulted in two-three types of clones, which differed only by 2-3 % towards the C-terminal end at amino acid level, as already reported about the structural and functional redundancy of MYB genes. One of the clones in the correct reading frame was selected for further studies, which contained an insert of 705 bp. It was assigned *Ll*MYB_SSM and submitted to the ncbi database with an Accession Number GU901209.

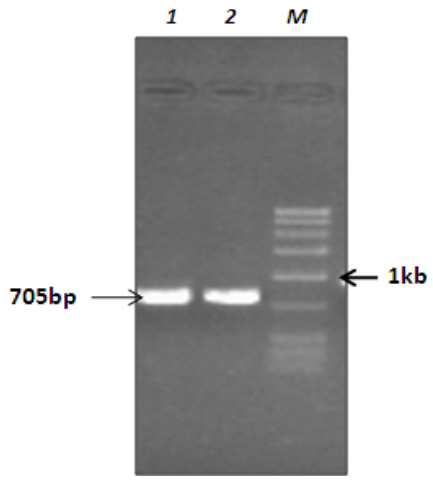


Fig. 4.12(a) PCR amplicon of *Ll*MYB_SSM on 1% agarose. Lanes1&2: ~700 bp M: LR DNA Ladder

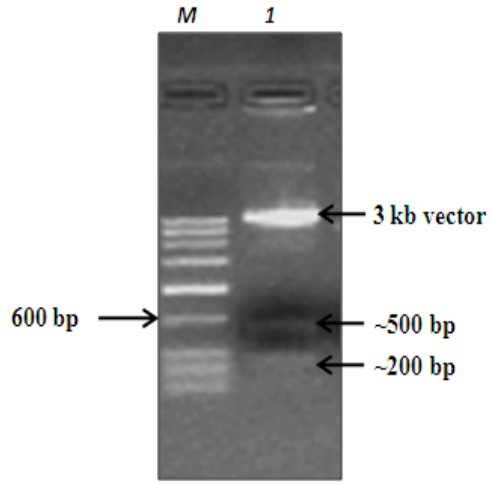


Fig. 4.12(b) *Ll*MYB_SSM insert in pGEM-T easy vector. LaneM: LR DNA ladder; Lane 1: 3kb vector backbone and ~200 and ~500 bp fragments of the 700 bp insert.

4.3.6.1 *Ll*MYB_SSM sequence analysis

An ORF of 705 bp beginning with the putative start codon (ATG) and terminating with the stop codon (TGA) was identified.

```
1   ATGGGGAGAG CTCCGTGCTG CGAGAAAGCT CACACTAACA AGGAGCATG
51  GACCAAGGAA GAAGACGATC GTCTTGTTGC TTATATTCGA GCTCACGGCG
101 AGGGCTGCTG GCGCTCTCTC CCCAAAGCCG CCGGCCTCCT CCGTTGTGGC
151 AAGAGCTGCC GTCTCCGCTG GATCAACTAC CTCCGTCCCG ACCTCAAACG
201 CGGCAATTTT ACTGAAGAAG AAGATGAGCT CATCATCAAG CTCCACAGTC
251 TCCTCGGCAA TAAGTGGTCG CTGATAGCCG GGAGATTACC GGGAAGAACG
301 GACAACGAGA TGAAGAATTA CTGGAATACT CATATAAGAA GGAAGCTTCT
351 GAGCAGAGGA ATTGACCCTA CAACTCACAG GCCTCTCAAC GATGCTTCTC
                                  ← SSM 5R2(nested)
401 AGAACCAATC TCATGCCTCA TCTGCTTCTG CTGCTAAACA GCACGACTCG
                              ←SSM 5R1
451 GAGATTCTGA TTACGGACAC AGATGCGAATTCCAAACGAC CCGGCGTCTT
501 TGAAGTGTGC CCCGACTTGA ACCTTGAGCT AACCATCAGT CTTCCCCATC
551 ATCATCAGCA TCAGCAACCA GATCAGATAT CGCTCAAGAG CATTAATAGT
601 GAGAGCAGTA CGGCTGGCGA TGGTTTTGAT TTCTTGGAAT CGAAAAGCAG
651 TGTTTTGGAT TGCAGGAGCA GCGTAACCCA AATAGATTGG AGAGCACTCA
701 CAGAATGA
```

Fig. 4.13: Nucleotide sequence (705 bp) of *Ll* MYB_SSM, showing the putative initiation codon (ATG) and termination codon (TGA). An *Eco*RI site lying in the C-terminal domain has been underlined. The reverse primers designed from the region outside the MYB domain have been highlighted.

Protein BLAST of the retrieved 235 amino acid sequence (Fig. 4.14), showed the presence of two highly conserved R2 and R3 MYB-DNA binding domains along with a non-conserved carboxyl-terminal domain (Fig 4.15).The three tryptophan residues (W) forming the hydrophobic core of the R2 repeat and 1 Phenylalanine and 2 tryptophan residues in the R3 repeat, a characteristic of all the plant R2R3 MYB genes are also located on the sequence.

4.3.6.2 Predicted protein sequence for *Ll* MYB_SSM:

The nucleotide sequence was translated into a 235 amino acid sequence using justbio as shown in Fig. 4.14.

```
  1-ATGGGGAGAGCTCCGTGCTGCGAGAAAGCTCACACTAACAAAGGAGCATGGACCAAGGAA - 60
  1- M  G  R  A  P  C  C  E  K  A  H  T  N  K  G  A  W  T  K  E   - 20

 61-GAAGACGATCGTCTTGTTGCTTATATTCGAGCTCACGGCGAGGGCTGCTGGCGCTCTCTC - 120
 21- E  D  D  R  L  V  A  Y  I  R  A  H  G  E  G  C  W  R  S  L   - 40

121- CCCAAAGCCGCCGGCCTCCTCCGTTGTGGCAAGAGCTGCCGTCTCCGCTGGATCAACTAC - 180
 41 - P  K  A  A  G  L  L  R  C  G  K  S  C  R  L  R  W  I  N  Y   - 60

181- CTCCGTCCCGACCTCAAACGCGGCAATTTTACTGAAGAAGAAGATGAGCTCATCATCAAG - 240
 61 - L  R  P  D  L  K  R  G  N  F  T  E  E  E  D  E  L  I  I  K   - 80

241- CTCCACAGTCTCCTCGGCAATAAGTGGTCGCTGATAGCCGGGAGATTACCGGGAAGAACG - 300
 81- L  H  S  L  L  G  N  K  W  S  L  I  A  G  R  L  P  G  R  T   - 100

301- GACAACGAGATGAAGAATTACTGGAATACTCATATAAGAAGGAAGCTTCTGAGCAGAGGA - 360
101- D  N  E  M  K  N  Y  W  N  T  H  I  R  R  K  L  L  S  R  G   - 120

361- ATTGACCCTACAACTCACAGGCCTCTCAACGATGCTTCTCAGAACCAATCTCATGCCTCA - 420
121- I  D  P  T  T  H  R  P  L  N  D  A  S  Q  N  Q  S  H  A  S   - 140

421- TCTGCTTCTGCTGCTAAACAGCACGACTCGGAGATTCTGATTACGGACACAGATGCGAAT - 480
141- S  A  S  A  A  K  Q  H  D  S  E  I  L  I  T  D  T  D  A  N   - 160

481- TCCAAACGACCCGGCGTCTTTGAAGTGTGCCCCGACTTGAACCTTGAGCTAACCATCAGT - 540
161- S  K  R  P  G  V  F  E  V  C  P  D  L  N  L  E  L  T  I  S   - 180

541- CTTCCCCATCATCATCAGCATCAGCAACCAGATCAGATATCGCTCAAGAGCATTAATAGT - 600
181- L  P  H  H  H  Q  H  Q  Q  P  D  Q  I  S  L  K  S  I  N  S   - 200

601- GAGAGCAGTACGGCTGGCGATGGTTTTGATTTCTTGGAATCGAAAAGCAGTGTTTTGGAT - 660
201- E  S  S  T  A  G  D  G  F  D  F  L  E  S  K  S  S  V  L  D   - 220

661- TGCAGGAGCAGCGTAACCCAAATAGATTGGAGAGCACTCACAGAATGA - 708
221- C  R  S  S  V  T  Q  I  D  W  R  A  L  T  E STOP
```

Fig. 4.14: Translated amino acid sequence of *Ll* MYB_SSM, showing the conserved Phenylalanine (F) and tryptophan (W) residues in red colour, in the R2 and R3 DNA-binding domains. The initiation codon(ATG) and termination codon (TGA) are shaded in green and blue colours respectively.

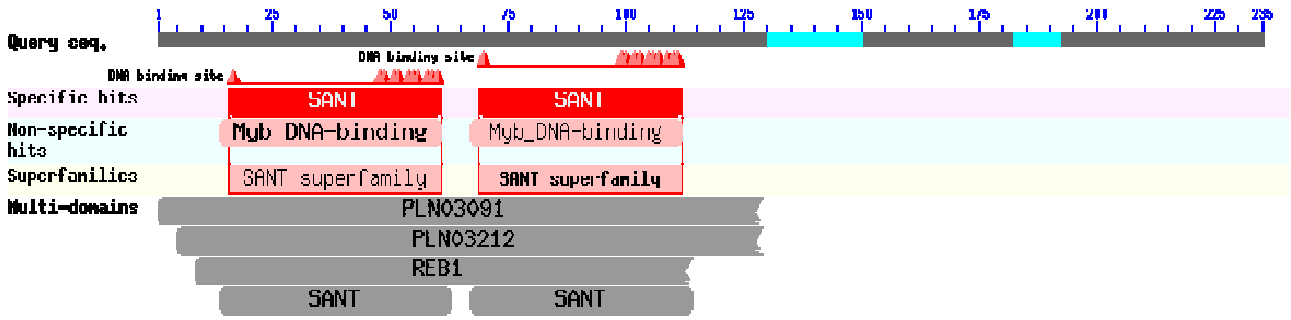


Fig. 4.15: Putative conserved domains identified on the *Ll* MYB_SSM translated sequence

4.3.6.3 Codon usage of the predicted protein sequence

GC% of the gene was found to be 50.56% and its codon usage was predicted based on the standard genetic code using justbio tool.

codon	mean	codon	mean	codon	mean	codon	mean
UUU ( )	1.27	UCU ( )	2.12	UAU ( )	0.42	UGU ( )	0.42
UUC ( )	0.42	UCC ( )	0.42	UAC ( )	0.85	UGC ( )	2.54
UUA ( )	0.42	UCA ( )	0.42	UAA ( )	0.00	UGA ( )	0.42
UUG ( )	1.27	UCG ( )	1.69	UAG ( )	0.00	UGG ( )	2.54
CUU ( )	1.69	CCU ( )	0.85	CAU ( )	2.54	CGU ( )	1.69
CUC ( )	5.51	CCC ( )	2.12	CAC ( )	2.12	CGC ( )	1.27
CUA ( )	0.42	CCA ( )	0.42	CAA ( )	1.27	CGA ( )	0.85
CUG ( )	1.27	CCG ( )	0.85	CAG ( )	2.12	CGG ( )	0.00
AUU ( )	2.12	ACU ( )	1.69	AAU ( )	2.54	AGU ( )	2.12
AUC ( )	1.69	ACC ( )	1.27	AAC ( )	2.54	AGC ( )	2.97
AUA ( )	1.69	ACA ( )	1.27	AAA ( )	2.97	AGA ( )	2.54
AUG ( )	0.85	ACG ( )	1.27	AAG ( )	2.97	AGG ( )	1.27
GUU ( )	0.85	GCU ( )	3.81	GAU ( )	3.81	GGU ( )	0.42
GUC ( )	0.42	GCC ( )	1.69	GAC ( )	2.97	GGC ( )	3.39
GUA ( )	0.42	GCA ( )	0.85	GAA ( )	3.39	GGA ( )	1.27
GUG ( )	0.42	GCG ( )	0.42	GAG ( )	2.97	GGG ( )	0.85

Fig. 4.16: Codon usage of *Ll*MYB_SSM gene expressed as percentage of all codons

4.3.7 PCR for 5' RACE reaction

Based on the putative full length *Ll*MYB_SSM cDNA sequence, appropriate reverse primers were designed from the non-conserved domain outside the Myb DNA-binding domain as shown in Fig. 4.13, to fish out the actual 5'UTR for the isolated gene. The 5' RACE cDNA was prepared as described in Chapter 2, Section 2.9.4.10. A primary 5' RACE reaction with AAP and SSM5R1 primers was followed by secondary reaction with AUAP and nested gene specific primers designed from *Ll*MYB_SSM gene sequence shown in Fig.4.13. The reaction conditions and the primer sequences for AAP and AUAP are mentioned in Chapter 2, Section 2.9.4.10.

Primer name	Primer sequence
SSM 5R1	5'- GCAGATGAGGCATGAGATTGGTTCTGA-3'
SSM 5R2(nested)	5'- AGGCCTGTGAGTTGTAGGGTCAAT-3'

A secondary 5'RACE amplicon of ~500 bp was obtained by running on 1% agarose gel, which was excised, gel eluted and ligated with pGEM-T easy vector for cloning. The positive clones obtained after blue-white screening and colony PCR were confirmed by sequencing the plasmid DNA. A single 5'UTR of 25 bases including a coding sequence of 384 bases was obtained which aligned well with *Ll*MYB_SSM gene sequence and assigned **5'*Ll*MYB_SSM**.

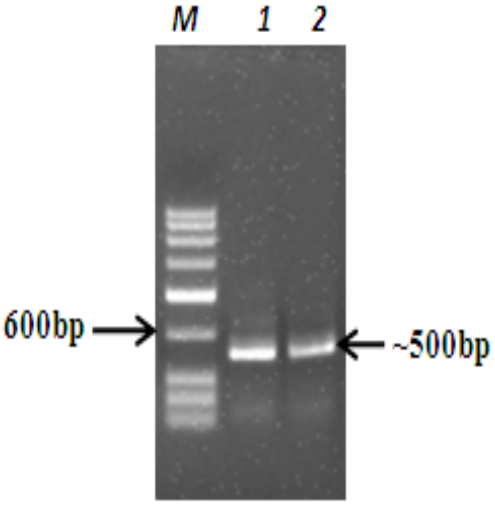


Fig. 4.17(a): PCR amplicon of *Ll*MYB_SSM on 1% agarose gel. Lane M: ʁ DNA ladder; Lanes 1 & 2: ~500bp condary5' RACE product

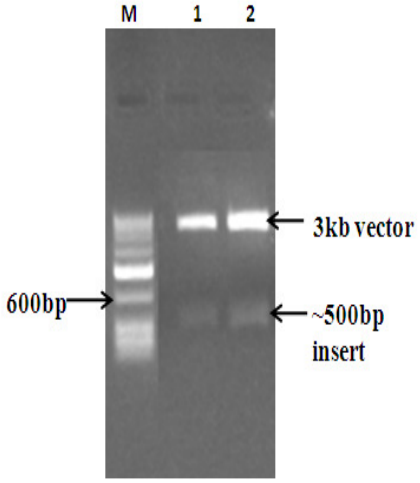


Fig. 4.17(b): 5'*Ll*MYB_SSM insert in pGEM-T easy vector. Lane M: LR DNA ladder ; Lanes 1 & 2: 3kb vector backbone and ~500bp insert of 5'*Ll*MYB_SSM.

4.3.7.1 Sequence analysis of 5'*Ll*MYB_SSM

The 5'UTR sequence overlapped with 99% identity at nucleotide level with the full length *Ll*MYB_SSM gene sequence. The 1% difference was in the forward primer region as it was designed based on other MYB gene sequences. The putative initiation codon

ATG is flanked by a G at + 4, in keeping with the nucleotide commonly found flanking the initiator methionine in plants (Lutcke *et al.,* 1987).

Forward nested AUAP→

```
  1 GGCCACGCGT CGACTAGTAC GGGGGGGGGG GGGGagatgt tgcagaagcc
 51 atccttaacA TGGGAAGGTC TCCTTGCTGC GAGAAAGCTC ACACTAACAA
101 AGGAGCATGG ACCAAGGAAG AAGACGATCG TCTTGTTGCT TATATTCGAG
151 CTCACGGCGA GGGCTGCTGG CGCTCTCCCC CCAAAGCCGC CGGCCTCCTC
201 CGTTGTGGCA AGAGCTGCCG TCTCCGCTGG ATCAACTACC TCCGTCCCGA
251 CCTCAAACGC GGCAATTTTA CTGAAGAAGA AGATGAGCTC ATCATCAAGC
301 TCCACAGTCT CCTCGGCAAT AAGTGGTCGC TGATAGCCGG GAGATTACCG
351 GGAAGAACGG ACAACGAGAT AAAGAATTAC TGGAATACTC ATATAAGAAG
401 GAAGCTTCTG AGCAGAGGAA TTGACCCTAC AACTCACAGG CCT
```

← SSM 5R2 (nested)

Fig. 4.18: Nucleotide sequence (443bp) of 5'*Ll*MYB_SSM RACE product. Forward nested 5'RACE primer (AUAP) and reverse SSM5R2 (nested) primer are shown in bold; 5' UTR of 25 bases is shown in italics lowercase. The initiation codon ATG is shown underlined.

4.3.8 Bioinformatic characterization of deduced *Ll*MYB_SSM protein

The *L .leucocephala Ll*MYB_SSM cDNA was conceptually translated and protein sequence was analyzed using various bioinformatic tools.

4.3.8.1 Amino acid composition

The deduced amino acid sequence of *Ll*MYB_SSM gene shows that the coding region consists of 235 amino acids. The theoretical molecular weight for the protein was estimated to be 26.52 kD and pI was estimated to be 7.68 using the tool ProtParam (Gasteiger *et al*., 2005), accessed from expasy proteomics server. The amino acid composition of the translated protein sequence for the gene is given in table 4.1. The total number of negatively charged residues (Asp + Glu) equals 31and positively charged residues (Arg + Lys) equal 32 and the empirical atomic formula for the protein was found to be: $C_{1145}H_{1821}N_{351}O_{358}S_{9.}$

Table 4.1 Amino acid composition of *Li*MYB_SSM protein

Amino acid	Total No.	Percentage	Amino acid	Total No.	Percentage
Ala (A)	15	6.4	Leu (L)	25	10.6
Arg (R)	18	7.7	Lys (K)	14	6.0
Asn (N)	12	5.1	Met (M)	2	0.9
Asp (D)	16	6.8	Phe (F)	4	1.7
Cys (C)	7	3.0	Pro (P)	10	4.3
Gln (Q)	8	3.4	Ser (S)	24	10.2
Glu (E)	15	6.4	Thr (T)	13	5.5
Gly (G)	14	6.0	Trp (W)	6	2.6
His (H)	11	4.7	Tyr (Y)	3	1.3
Ile (I)	13	5.5	Val (V)	5	2.1

4.3.8.2 Prediction of sub-cellular localization of *Li*MYB_SSM

In silico sub-cellular localisation of the protein was predicted to be **nuclear** using the program ProtComp Version 9.0, accessed from the softberry server. ProtComp combines several methods of protein localization prediction - neural networks-based prediction; direct comparison with updated base of homologous proteins of known localization; comparisons of pentamer distributions calculated for query and DB sequences; prediction of certain functional peptide sequences, such as signal peptides, transit peptides of mitochondria and chloroplasts, and transmembrane segments. It means that the program treats correctly only complete sequences, containing signal sequences and other functional peptides, if any. The output result for the prediction is summarized in Fig. 4.19.

Predicted by Neural Nets - Nuclear with score 2.9; Integral Prediction of protein location: Nuclear with score 9.8

Location weights:	LocDB /	PotLocDB /	Neural Nets /	Pentamers /	Integral
Nuclear	5.0 /	3.0 /	2.94 /	5.03 /	9.83
Plasma membrane	0.0 /	0.0 /	0.00 /	0.01 /	0.00
Extracellular	0.0 /	0.0 /	0.00 /	0.13 /	0.08
Cytoplasmic	0.0 /	0.0 /	0.00 /	0.00 /	0.00
Mitochondrial	0.0 /	0.0 /	0.00 /	0.00 /	0.00
Endoplasm. retic.	0.0 /	0.0 /	0.00 /	0.00 /	0.02
Peroxisomal	0.0 /	0.0 /	0.00 /	0.00 /	0.00
Golgi	0.0 /	0.0 /	0.00 /	0.06 /	0.07
Chloroplast	0.0 /	0.0 /	0.00 /	0.00 /	0.00
Vacuolar	0.0 /	0.0 /	0.06 /	0.00 /	0.00

Fig. 4.19 *Li*MYB_SSM ProtComp result for identification of sub-cellular localization.
LocDB: scores based on query protein's homologies with proteins of known localization. **PotLocDB:** scores based on homologies with proteins which locations are not experimentally

known but are assumed from strong theoretical evidence; **Neural nets:** Scores assigned by neural networks.
Pentamers are scores based on comparisons of pentamer distributions calculated for QUERY and DB sequences; **Integral** are final scores that combine all above previous scores

4.3.8.3 Hydropathy Plot

The hydropathy index of an amino acid is a number representing the hydrophobic or hydrophilic properties of its side-chain (Jack Kyte and Russell Doolittle, 1982). The larger the number is, the more hydrophobic the amino acid. The most hydrophobic amino acids are isoleucine (4.5) and valine (4.2). The most hydrophilic ones are arginine (- 4.5) and lysine (-3.9). This is very important in protein structure; hydrophobic amino acids tend to be internal (with regard to the protein's native 3- dimensional structure) while hydrophilic amino acids are more commonly found towards the protein surface. The amino acid sequence of *Ll*MYB_SSM protein was analyzed using Kyte-Doolittle Hydropathy plot at http://gcat.davidson.edu/rakarnik/kd.cgi (Fig 3.16A) window size 9. When the window size is **9**, strong negative peaks indicate possible surface regions of globular proteins, which may be putative antigenic epitope while positive values show the hydrophobic regions. The grand average of hydropathicity (GRAVY) i.e calculated as the sum of hydropathy values (Kyte and Doolittle, 1982) of all the amino acids, divided by the number of residues in the sequence, was calculated to be -0.743 as predicted by ProtParam tool of expasy server (Gasteiger *et al*., 2005).

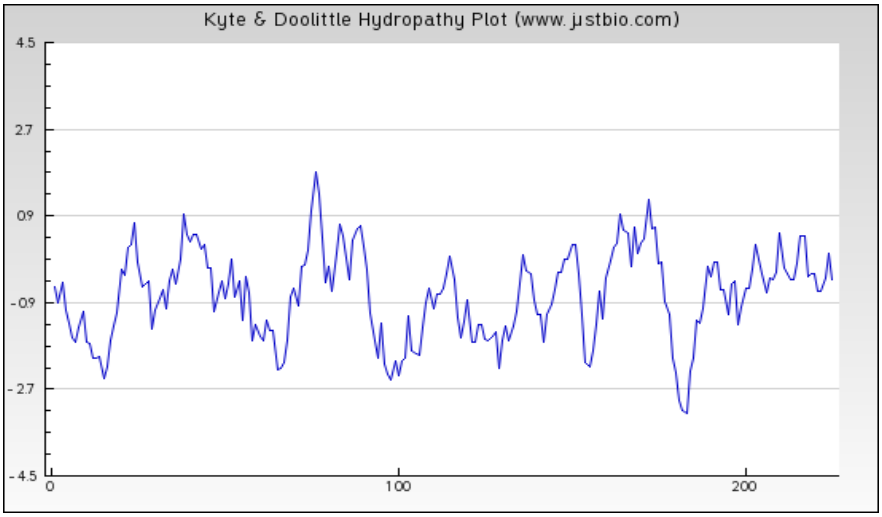


Fig. 4.21: Kyte-Doolittle Hydropathy for *Ll*MYB_SSM. Window size of 9 suggested a good value for finding putative surface-exposed regions.

4.3.8.4 Phylogenetic analysis of *Ll*MYB_SSM

Phylogenetic analysis of *Ll*MYB_SSM was done at amino acid level to investigate its molecular relationship with the functionally reported R2R3 MYB genes in the protein database of ncbi. Amino acid sequences of R2R3-type MYB genes from various plant species were obtained from GenBank by homologous blastp, non-redundant database search using full length *Ll*MYB_SSM amino acid sequence as query sequence. The BLAST search identified several homologous R2R3 MYB proteins all belonging to subgroup 4 (Kranz *et al*., 1998). The sequences were aligned using Clustal W 2.0.1 with the default settings. No manual adjustment was found to be necessary (Fig. 4.22).

The alignment resulted in significant similarity in the N-terminal R2R3 domains (1st 120 amino acids) of all the protein sequences, but the C-terminal region showed a very high diversity with various gaps due to different sizes However, a few conserved motifs were identified in the C-terminal region which is characteristic of R2R3 MYB genes belonging to subgroup 4 (Kranz *et al*., 1998), which have been discussed in Section 4.3.8.5. Overall, similarity score was the highest for *Ptt*MYB4a (66%), followed by *Eg*MYB1 (64%), *Am*MYB308 (62%), *At*MYB4 (61%) and ZmMYB31 and *Zm*MYB42 (59%).

```
EgMYB1          MGRSPCCEKAHTNKGAWTKEEDDKLIAYIRAHGEGCWRSLPKAAGLLRCGKSCRLRWINY 60
PttMYB4a        MGRSPCCEKAHTNKGAWTKEEDDRLVAYIRAHGEGCWRSLPKAAGLLRCGKSCRLRWINY 60
LlMYB_SSM       MGRSPCCEKAHTNKGAWTKEEDDRLVAYIRAHGEGCWRSLPKAAGLLRCGKSCRLRWINY 60
ATMYB4          MGRSPCCEKAHTNKGAWTKEEDERLVAYIKAHGEGCWRSLPKAAGLLRCGKSCRLRWINY 60
AmMYB308        MGRSPCCEKAHTNKGAWTKEEDDRLVAYIRAHGEGCWRSLPKAAGLLRCGKSCRLRWINY 60
ZmMYB31         MGRSPCCEKAHTNKGAWTKEEDERLVAHIRAHGEGCWRSLPKAAGLLRCGKSCRLRWINY 60
ZmMYB42         MGRSPCCEKAHTNRGAWTKEEDERLVAYVRAHGEGCWRSLPRAAGLLRCGKSCRLRWINY 60
                ************:********::*:*:::***********:*****************

EgMYB1          LRPDLKRGNFTEEEDEIIIKLHSLLGN-KWSLIAGRLPGRTDNEIKNYWNTHIRRKLLNR 119
PttMYB4a        LRPDLKRGNFTEAEDELIIKLHSLLGNSRWSLIAGRLPGRTDNEIKNYWNTHIRRKLLNR 120
LlMYB_SSM       LRPDLKRGNFTEEEDELIIKLHSLLGN-KWSLIAGRLPGRTDNEMKNYWNTHIRRKLLSR 119
ATMYB4          LRPDLKRGNFTEEEDELIIKLHSLLGN-KWSLIAGRLPGRTDNEIKNYWNTHIRRKLINR 119
AmMYB308        LRPDLKRGNFTEEEDELIIKLHSLLGN-KWSLIAGRLPGRTDNEIKNYWNTHIRRKLLSR 119
ZmMYB31         LRPDLKRGNFTEEEDELIVKLHSVLGN-KWSLIAGRLPGRTDNEIKNYWNTHIRRKLLSR 119
ZmMYB42         LRPDLKRGNFTADEDDLIVKLHSLLGN-KWSLIAARLPGRTDNEIKNYWNTHIRRKLLGS 119
                ***********  **.:*:****:*** :***** *********:************:.

EgMYB1          GIDPATHRLINEPAQDHHDE---------PTISFAANSKEIKEMKNN-------AELNFM 163
PttMYB4a        GIDPATHRPLNEPVQEAT-----------TTISFTTTTTSVEEESRGSIIKEEIKEKLIS 169
LlMYB_SSM       GIDPTTHRPLNDASQNQS----------------HASSASAAKQHDS----------EIL 153
ATMYB4          GIDPTSHRPIQESSASQDSKPTQLEPVTSNTINISFTSAPKVETFHESISFPGKSEKISM 179
AmMYB308        GIDPTTHRSINDGTASQD-----------QVTTISFSNANSKEEDTK----------HKV 158
ZmMYB31         GIDPVTHRPVTEHHASNIT------------ISFETEVAAAARDDKK--------GAVFR 159
ZmMYB42         GIDPVTHRRVAGGAATTIS------------FQPSPNSAAAAAAAET--------AAQAP 159
                ****.:** :
```

R2 Domain

R3 Domain

```
EgMYB1          CNLEESA-------------------------DVASSARERCPDLNLELGISPPSHQL- 196
PttMYB4a        ATAFVCT-------------------------EAKTQVQERCPDLNLELGISLPSQNQ- 202
LlMYB_SSM       ITDTDAN-------------------------SKRPGVFEVCPDLNLELTISLPHHHQ- 186
ATMYB4          LTFKEEK-------------------------DECP-VQEKFPDLNLELRISLPDDVD- 211
AmMYB308        AVDIMIK-------------------------EENSPVQERCPDLNLDLKISPPCQQQI 192
ZmMYB31         LEDEEEEERNKATMVVGRDRQSQSHSHSHPAGEWGQGKRPLKCPDLNLDLCISPPCQEEE 219
ZmMYB42         IKAEET---------------------------AAVKAPRCPDLNLDLCISPPCQHED 190
                                                    *****:* ** * .

EgMYB1          ----HQPEPLLRFTGRK------------SDLCECNLGLKN-SQNCRCSVG--VIESETS 237
PttMYB4a        ----PDHHQPFKTGGSR------------SLCFACSLGLQN-SKDCSCNVI--VSTVGSS 243
LlMYB_SSM       ----HQQPDQISLKSIN------------SESSTAGDGFDF-LESKSS-----VLDCRSS 224
ATMYB4          ----RLQG---HGKSTT------------PRCFKCSLGMIN-GMECRCGRMRCDVVGGSS 251
AmMYB308        ----NYHQENLKTGGRNGSS---------TLCFVCRLGIQN-SKDCSCSDG-----VGN- 232
ZmMYB31         --EMEEAAMRVRPA-VKREA---------GLCFGCSLGLPR-TADCKCSS---------- 256
ZmMYB42         DGEEEDEELDLKPAFVKREALQAGHGHGHGLCLGCGLGGQKGAAGCSCSNG--------- 241
                                             .  *        .

EgMYB1           ------VGYDFLGLK----ASVLDYRS---- 254
PttMYB4a         GSTSTKNGYDFLGMK----SGVLDYRSLEMK 270
LlMYB_SSM        -------------------VTQIDWRALTE- 235
ATMYB4           KGSDMSNGFDFLGLAKKETTSLLGFRSLEMK 282
AmMYB308         -------------------------------
ZmMYB31          --------SSFLGLR----TAMLDFRSLEMK 275
ZmMYB42          --------HHFLGLR----TSVLDFRGLEMK 260
```

Fig. 4.22: Multiple sequence alignment of phenylpropanoid pathway related R2R3MYB amino acid sequences retrieved from the ncbi database. The boxed region in the R3 domain represents the bHLH protein interaction motif with consensus sequence [DE]LX2[RK]X3LX6LX3R. GenBank accession numbers: *Ll*MYB_SSM (GU901209), EgMYB1 (CAE09058), AtMYB4 (AF062860), ZmMYB31 (CAJ42202), ZmMYB42 (CAJ42204), AmMYB308 (P81393), PttMYB4a (AJ567346).

Phylogenetic analysis was conducted using MEGA version 4.0.2 (Tamura *et al*., 2007, http://www.megasoftware.net/) based on the full length amino acid sequences of the aligned sequences as well as few other reported subgroup 4 members from different taxa and 2 members from subgroup 3 were used as landmark. Human *v*MYB sequence was used as outgroup. Bootstrapping (1000 replicates) was performed to evaluate the statistical reliability of the inferred topology (Felsenstein, 1985). The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). A total of 196 positions were there in the final data set.

Mybs from angiosperm and conifer species are not found in distinct groups, but instead are interspersed. This suggests that significant expansion of plant R2R3 Myb genes occurred before the divergence of the two growth forms. This conclusion is in agreement with the previous studies (Lipsick, 1996; Rabinowicz *et al*., 1999). Even members of the same taxa form different branches, due to the divergence in the C-terminal domains. *Ll*MYB_SSM was grouped with *Am*MYB308 with a bootstrap value of 21 which was quite low, along with all the known repressors of the lignin pathway being associated in closer branches (Fig. 4.23). The bootstrap values are low for several branches, because of the large number of taxa included for the study.

Therefore, to support the reliability of the classification, only the R2R3 domains of the taxa used in Fig.4.23 were used to generate a new neighbor joining tree to see the difference in topology of the phylogeny. The analysis was conducted with p-distances, which has been suggested for the analysis of large numbers of genes with relatively short sequences (Nei, 1996; Rabinowicz *et al*., 1999), as the R2R3 domain alone comprises of only 120 amino acids (Fig. 4.24).

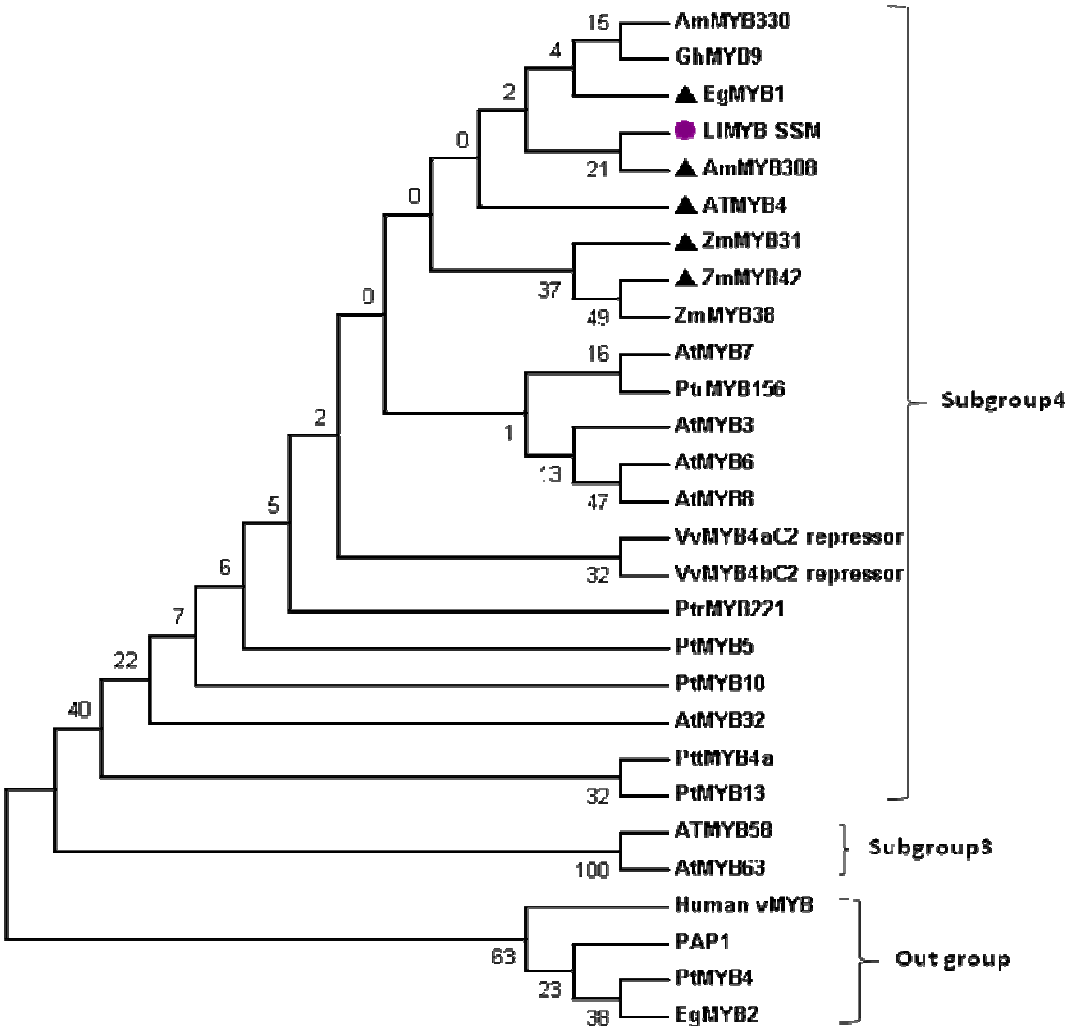


Fig. 4.23: Rooted Neighbour-Joining tree of *Ll*MYB_SSM with the angiosperm and conifer full length R2R3MYB amino acid sequences, belonging to subroup4 obtained using Mega 4 software. *Ll*MYB_SSM is flanked by a filled circle ● ; known repressors of lignin pathway are flanked by filled triangle ▲. Subgroup 4 members form a separate clade and members of subgroup 3 (known activators of lignin pathway) have been used as a landmark. Human *v*MYB has been included for out-grouping and along with it PAP1, PtMYB4 and EgMYB2 also known as activators of pathway have been grouped.

GenBank Accession numbers: *Ll*MYB_SSM (GU901209); *Pinus taeda*:PtMYB5 (FJ469924), PtMYB10 (FJ469925), PtMYB13 (FJ469926), PtMYB4 (AY356371); *Populus trremuloides X Populus trycocarpa* : PttMYB4a (AJ567346); *Populus trichocarpa*:PtrMYB221(estExt_fgenesh4_pg.C-L_G_IV1453), *tr*MYB156(eugene3.00090345); *Vitis vinifera*:VvMYB4a(ABL61515),VvMYB4b(ACN94269); *Arabidopsis thaliana*: AtMYB3 (AF062859), AtMYB4 (AF062860), AtMYB6 (U26936), AtMYB8(AF207991), AtMYB32(AF062874),AtMYB85(AF175993),AtMYB58(AF062893),AtMYB63(AF062898),PAP1(At1g56650); *Zea mays*: ZmMYB31 (CAJ42202), ZmMYB42 (CAJ42204),ZmMYB38();

Antirrhinum majus: AmMYB308 (P81393),AmMYB330 (P81395); *Eucalyptus gunnii*: EgMYB1(CAE09058), EgMYB2(AJ576023);*Gossypium hirsutum*:GhMYB9 (AAK19619); *Homo sapiens*: Hu*v*MYB (NP_001155131)

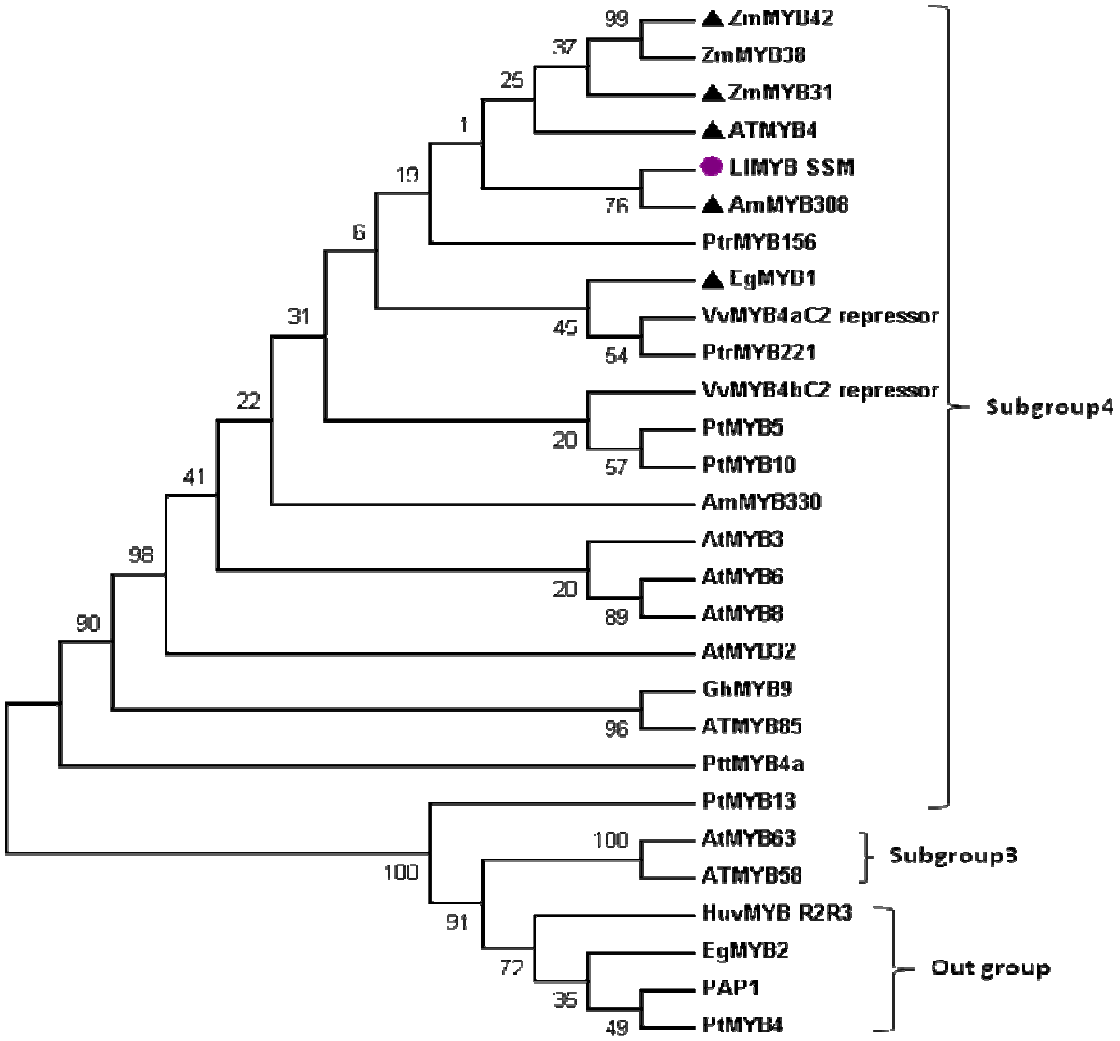


Fig. 4.24: Rooted Neighbour-Joining tree generated by alignment of R2R3 domain of *Ll*MYB_SSM protein with the angiosperm and conifer sequences, belonging to subgroup 4 obtained using Mega 4 software. *Ll*MYB_SSM is flanked by a filled circle ●; known repressors of lignin pathway are flanked by filled triangle▲. Subgroup4 members form a separate clade and members of subgroup 3 (known activators of lignin pathway) have been used as a landmark. Human *v*MYB has been included for out-grouping and along with it PAP1, PtMYB4 and EgMYB2 alsot known activators of pathway have been grouped.

The inferred topology of the tree generated in Fig. 4.24, shows that *Ll*MYB_SSM is associated with *Am*MYB308, with a significantly higher bootstrap value of 76 and still associated with the known repressors of the pathway, with improved bootstrap values. A

slight change is seen in sub-grouping of few other members; which can be attributed to the p-distance parameter used for generating the tree, as compared to the gamma parameter used for previous tree (Fig.4.23).

4.3.8.5 Identification of conserved motifs on the C-terminal region of protein

Plant R2R3Myb genes possess highly conserved N-terminal Myb domains, but have dramatically divergent C-terminal regions. However, conserved motifs within the C-terminal regions of paralogous and orthologous Myb genes have been identified (Stracke *et al*., 2001), which may reflect constraints upon the functions of related Myb proteins. Preliminary results support the idea that genes with similar functions are clustered together as subgroups based on their phylogeny and similarity. Additionally, many of the individual subgroups contain C-terminal motif sequences, which can be used as distinct markers for Myb gene subgroups and on the basis of this fact, MYB proteins have been classified into 22 subgroups. As, inferred from the phylogenetic study of section 4.3.8.4, *Ll*MYB_SSM is grouped with subgroup 4 members with 90-98% similarity in the MYB domains. Hence, motif identification was performed with the carboxy-terminal domains of a few closely related members of the gene using MEME software (Bailey and Elkan, 1994).

Motifs were detected using MEME software with the following parameter settings: the distribution of motifs: zero or one per sequence; maximum number of motifs to find: 6; minimum width of motif: 4; maximum width of motif: 6; minimum number of sites for each motif: 10, i.e., the number of sequences so that the motif must be present in all members within the same subgroup. Other options used the default values. Only motifs with e-value <= 5e-10 were kept for further analysis, to ensure its statistical significance. To test whether the carboxy-terminal motifs are specific to Myb genes, motif sequences were used to perform homology search in Swiss-Prot database and EST data set from GenBank. The latter can also provide information on the expression pattern of Myb genes. Low complexity was turned off for optimal short sequence search in both homology searches. Out of the total 6 motifs identified by the software, 3 were most conserved and found in the C-terminal region of all the aligned subgroup 4 amino acid sequences.

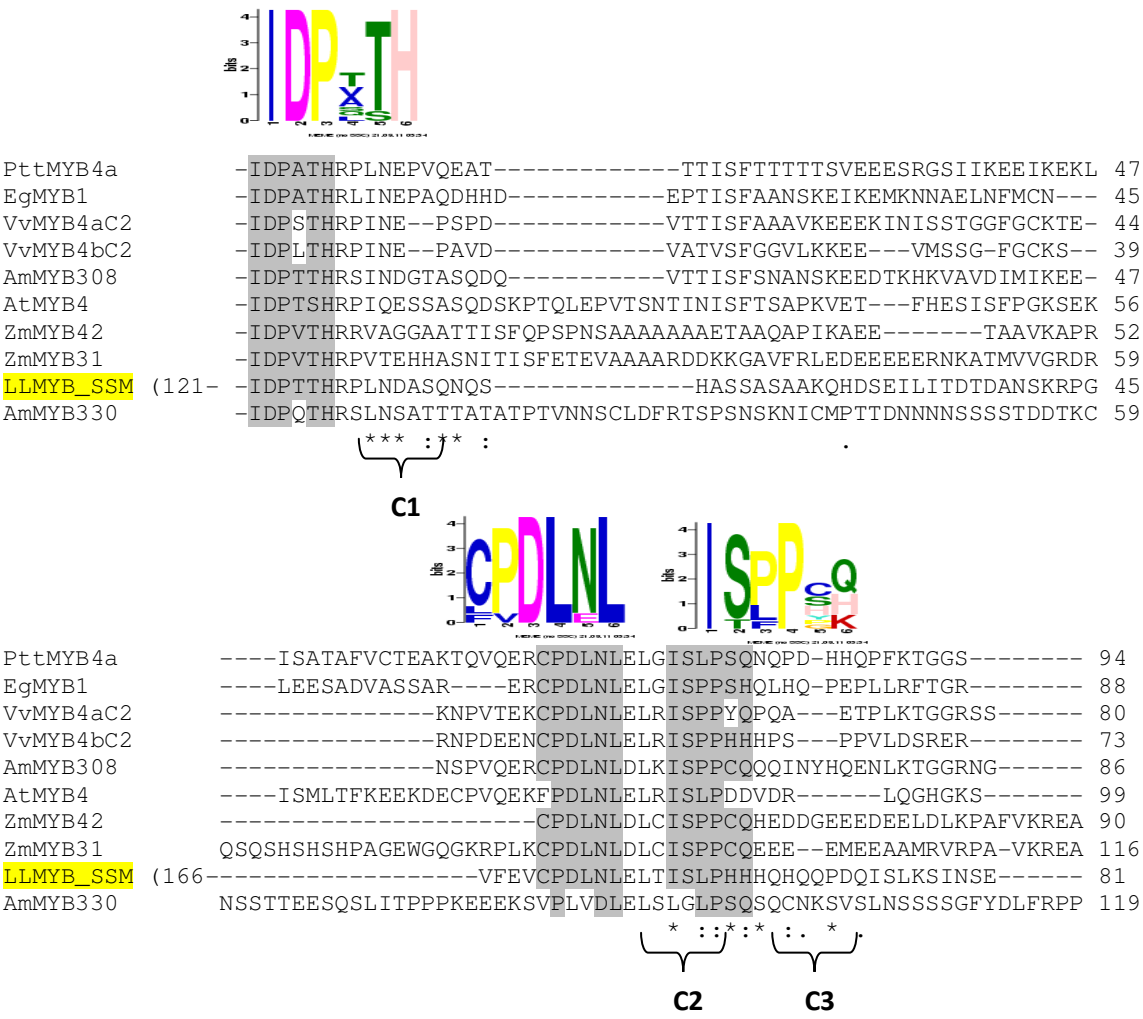


Fig. 4.25: Amino acid motifs identified in the C-termini of *Li*MYB_SSM and shared with members of subgroup 4, which are known repressors of phenylpropanoid pathway using MEME software, shown shaded on the alignment of predicted amino acid sequences generated using ClustalW. MEME motif logos on top with bit scores indicate the information content for each position. The three most conserved motifs identified have been labeled as C1, C2 and C3 respectively.

The degree of C-terminal conservation of *Li*MYB_SSM amino acid sequence with other subgroup 4 members is shown in Fig. 4.25 as interpreted using MEME software. Despite the highly variable C-terminal in MYBs, three conserved amino acid motifs were identified in the sequence. The C1 motif of unknown function and another C2 motif also known as the repressor domain containing the core from the EAR motif (Kazan, 2006), typical signature motif of subgroup 4 (Kranz *et al.,* 1998) was harboured by all the sequences. A third C3 motif was identified just adjacent to C2 but its function is not known.

4.3.9 Restriction map analysis of *Ll*MYB_SSM gene

Restriction analysis of *Ll*MYB_SSM was done using bioinformatic software pDRAW32 (Fig. 4.26). Analysis was limited to the parameters: including 6 bp or longer cutters; excluding non-palindromic recognition sites and enzyme site cutting not more than four times in the sequence. The parameters were restricted to simplify the sequence analysis for cloning experiments and Southern hybridization.

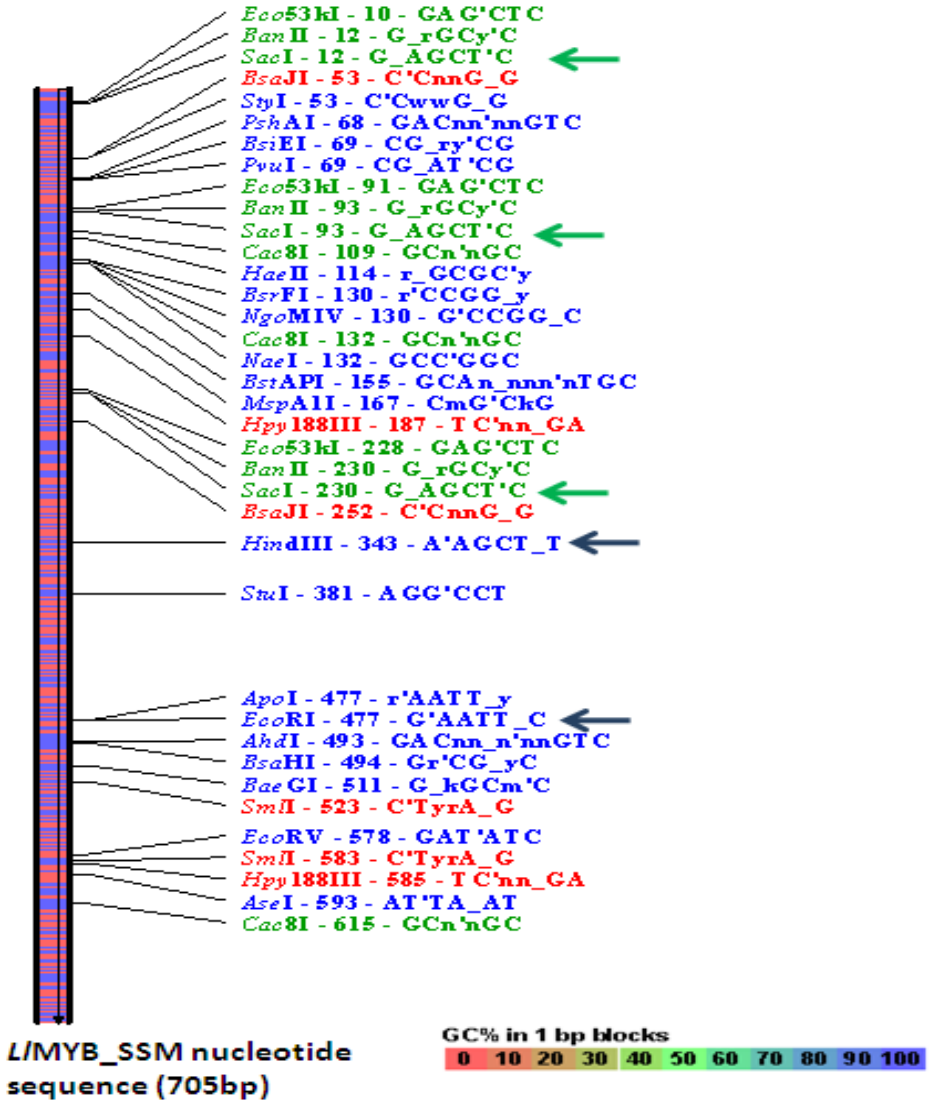


Fig. 4.26: Restriction map of *Ll*MYB_SSM (GU901209) gene using pDRAW32. The green arrows indicate *Sac*I sites (3 sites at positions: 12, 93, 230) and deep blue arrows *Hin*dIII (single site, position: 343) and *Eco*RI (single site, position: 477).

4.3.10 Genomic DNA extraction

Genomic DNA was isolated (Fig. 4.27) by using the protocol of Lodhi *et al.,* (1994) as mentioned in Chapter 2, Section 2.9.4.2. The concentration of the gDNA was measured by spectrophotometer as mentioned in Chapter 2, Section 2.9.4.7. The DNA was loaded on 0.8% agarose gel along with known concentration of λ-DNA to confirm its concentration and quality.

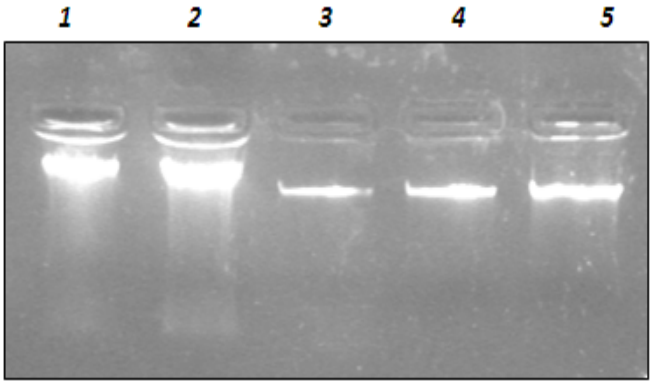


Fig. 4.27: *Leucaena leucocephala* genomic DNA (2µL each in lanes 1 & 2) **resolved on 0.8% agarose gel.** Lanes 3, 4 & 5: 100, 200 and 400ng of λ-DNA

PCR was performed using 100 ng of good quality gDNA as the template and gene specific primers for *Ll*MYB_SSM. The PCR reaction resulted in an amplicon of nearly the same size as the cDNA amplicon for *Ll*MYB_SSM. The amplicon was cloned and sequenced to be almost 99% identical to its cDNA clone. Thus, the genomic clone for *Ll*MYB_SSM was inferred to be intron-less.

4.3.11 Southern screening for *Ll*MYB_SSM copy number in *L. leucocephala*

Southern hybridization was performed to unveil the representation of *Ll*MYB_SSM gene in the *L. leucocephala* genome and thus determining its copy number. Being a member of multigene family of R2R3MYB, a probe specific for the gene sequence and hence designed from the most divergent region of the gene was required. Since, the N-terminal DNA-binding domain comprising initial 360 bases are highly conserved in MYB genes; a probe was designed from the carboxyl-terminal region of the gene as identified by multiple sequence alignment.

Good quality of genomic DNA was extracted from tender leaves of *Leucaena* plant according to the protocol described by Lodhi *et al.* (1994). Digestibility of the DNA was checked with a few restriction enzymes which cut at different positions in the *Ll*MYB_SSM gDNA sequence. After checking the digestion on 0.8% agarose gel, few enzymes (*Sac*I, *Eco*RI and *Hin*dIII) which produced complete digestion in 16-18h were selected for final reaction. According to the standard protocols, 10- 15 µg of genomic DNA were digested with the three enzymes individually. The completely digested DNA was electrophoretically separated on 0.8% agarose gel at 70 V for 6h (Fig. 4.28 A), and transferred to positively charged Hybond N^+ nylon membrane (Amersham). Hybridization was performed with two types of probes, which were designed from different regions of the gene and radiolabeled with P^{32} dCTP incorporated in it. Post hybridization membrane was exposed on intensifying screen. The screen was scanned for signal development using a very robust and much sensitive instrument named Typhoon Trio + scanning system.

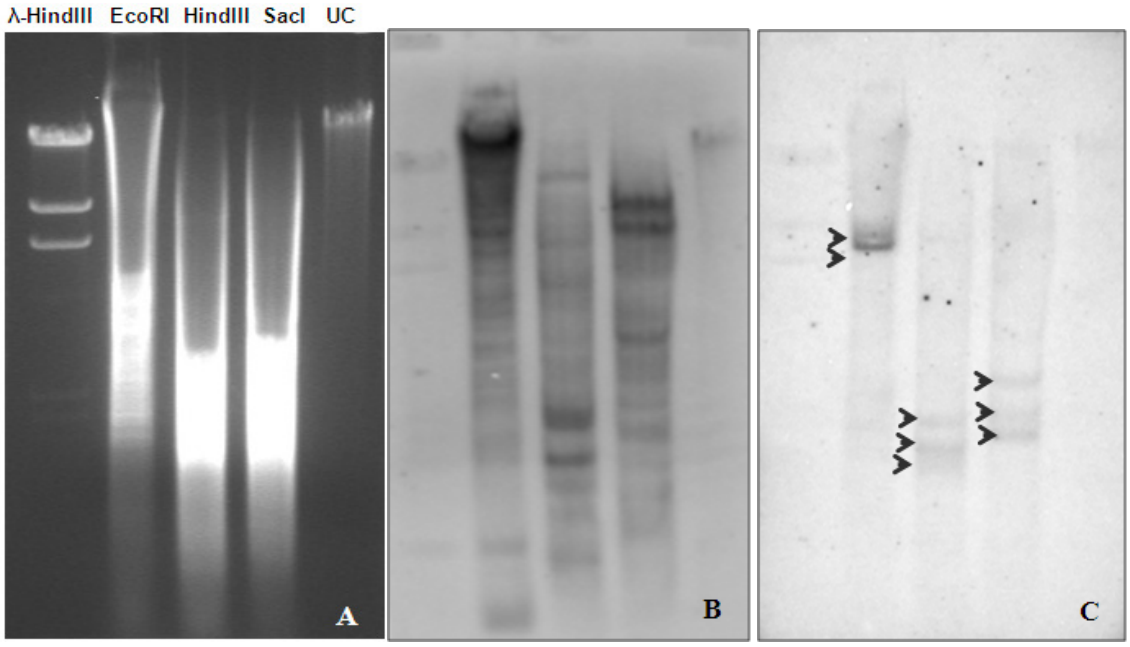


Fig. 4.28: (A) 0.7% agarose gel showing *L. leucocephala* genomic DNA digestion with *Eco*RI, *Hin*dIII and *Sac*I enzymes; λ-*Hin*dIII: Lambda DNA digested with *Hin*dIII enzyme; UC: Uncut genomic DNA. (B) Southern hybridisation signals obtained using R2R3 domain of *Ll*MYB_SSM as probe. (C) Southern hybridisation signals obtained using C-terminal domain of *Ll*MYB_SSM as probe.

Initially hybridization was performed with a probe synthesized from the R2R3 domain of the gene sequence. After performing the washes in stringent conditions, multiple bands were observed in each of the digestion lanes (Fig. 4.26 B). Due to the presence of 3 *Sac*I sites within the R2R3 domain region of the gene, a large number of hybridization signals were detected in the *Sac*I digest lane. However, *Eco*RI and *Hin*dIII sites were present outside the conserved domain, but still a number of signals were detectable. The results showed the presence of multiple copies of R2R3 type MYB genes in *L. leucocephala*, just as reported for other plant species with characterized genomes (*Arabidopsis, Vitis* & *Poplus*).

To find the representation of specifically *Ll*MYB_SSM in the genome, another probe was designed from the most divergent C-terminal region of the gene. The hybridization was repeated after stripping the same membrane, in similar conditions. The hybridization signals were much fewer this time (Fig. 4.26 C). Careful observation revealed the presence of 4-5 signals in *Sac*I lane and 3-4 signals were visible in *Eco*RI and *Hin*dIII lanes. As the C-terminal region of the gene has one site each for the latter two enzymes, number of signals would have been more; but as the lanes were over-run, probably the smaller fragments were not detectable in the screen. Based on the results it can be concluded that atleast 3-4 copies or isoforms of the gene are present in the *L. leucocephala* genome.

4.4 Discussions

The transcriptional activation of lignin biosynthetic genes is likely mediated through multiple cis-elements and a combinatorial interaction of multiple transcription factors. Earlier studies of the promoter activities of lignin biosynthetic genes have indicated the presence of cis elements besides the AC elements and of multiple proteins binding to the promoter sequences (Hauffe *et al*., 1993; Lacombe *et al*., 2000; Hatton *et al*., 1995). In addition, some MYBs have been suggested to function as repressors in fine-tuning the expression level of phenylpropanoid biosynthetic genes. Identification and characterization of all the cis elements and transcription factors involved in regulation of lignin biosynthesis is essential for gaining an understanding of the complexity of

transcriptional control of lignin biosynthesis, especially in woody plant species like *Leucaena*.

Promoter-related studies presented in chapter 3, showed the presence of these highly conserved AC elements, which are known to be crucial in regulating the pathway genes by interacting with R2R3MYB transcription factors. To understand the role of MYB transcription factors in the pathway regulation, a lignin-related MYB transcription factor was required. R2R3MYB is known to be a multigene family with more than 100 members having highly conserved N-terminal DNA-binding domain and highly divergent C-terminal domain in plants, with various functions in different plant tissues. So, a set of degenerate primers was designed from the most conserved N-terminal R2R3 domain of few functionally known R2R3-type MYB genes from tree species, to fish out a partial MYB gene fragment from the *L. leucocephala* cDNA pool. As, no EST library was available for isolation of xylem specific genes, cDNA prepared from stem tissues of the plant, which are supposed to have actively lignifying cells was used as template for the PCR-based approach to isolate the gene of interest from the plant. Finally, *Ll*MYB_SSM gene was isolated to its full length by using 3' RACE and finally 5' RACE and characterized by various bioinformatic tools as an R2R3 Myb gene belonging to subgroup 4 of the R2R3 Myb gene family based on the classification criteria of Stracke *et al.*, 2001 and Kranz *et al.*, 1998. *In silico* translated amino acid sequence of the gene showed the presence of the characteristic conserved R2 and R3 domains in its N-terminal, showing upto 95% similarity with the R2R3 Myb genes mainly involved in repressing the phenylpropanoid pathway, as shown in the section 4.3.8.4. An interesting fact, was the presence of a b-HLH interaction domain in the R3 domain of the sequence (Fig. 4.22), which together with well documented partnership between MYB and b-HLH proteins (Ramsay and Glover, 2005) suggests that *Ll*MYB_SSM might interact with a specific b-HLH to control transcription process. However, like previous reports, very poor homology was seen in the C-terminal region of the sequence with the same aligned sequences. Phylogenetic analysis using the full length amino acid sequence of *Ll*MYB_SSM was performed to identify its most closely related member belonging to subgroup 4. The analysis showed that *Am*MYB308, known to be a repressor of lignin pathway genes (Tamagnone *et al.*, 1998), was the member closest to *Ll*MYB_SSM. To

determine any other conserved regulatory motifs in the C-terminal region of the gene, MEME software analysis was done along with the most closely related subgroup4 members. Amongst the three conserved C-terminal motifs proposed by Kranz *et al*., (1998), two were identified in the gene, namely the C1 motif (LlsrGIDPX[T/S]HRX[I/L]) and C2 motif (pdLNL[D/E]LXi[G/S]). The C2 motif containing the core EAR-motif was previously shown to play an essential role in active repression (Kazan, 2006; Hiratsu *et al*., 2003). However, a third C-3 motif also called as the zinc finger motif (CX1–2CX7–12CX1–2C) is absent in *Ll*MYB_SSM, as compared to the other members, to which no definite function has been assigned yet. As, the C-terminal characteristics of *Ll*MYB_SSM are consistent with its classification in R2R3MYB subgroup 4, it suggests functional conservation of the gene with other subgroup members. In contrast to the carboxy-terminal coding regions, no conserved DNA sequence motifs were identified in the Myb gene noncoding regions (5' or 3' UTR sequences) of the same subgroup. This could be due to the fact that the Myb genes clustered in each subgroup are probably not orthologs or paralogs.

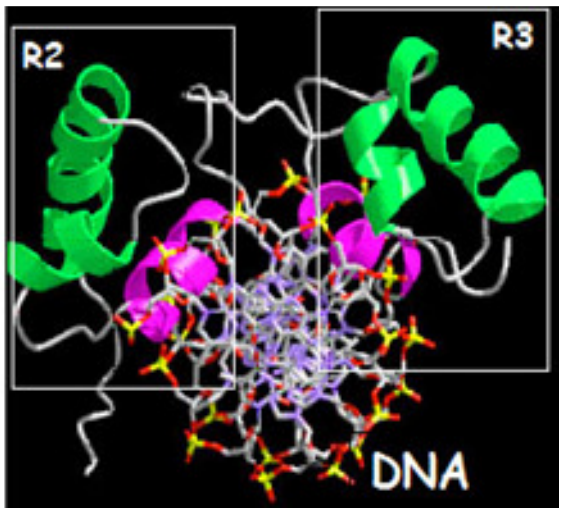
Southern analysis for the gene was done with *L. leucocephala* gDNA digested with different enzymes to get an idea about the copy number of the gene. Initially, the membrane was probed with the region of the gene encoding N-terminal, DNA-binding domain, which resulted in several signals in each lane, indicating that the plant genome consists of several R2R3-type MYB genes, which is consistent with the previous reports in plants. Probing with the C-terminal non-conserved domain encoding fragment resulted in very few signals of 3-4, indicating that atleast 3-4 isoforms of *Ll*MYB_SSM are present in the *L.leucocephala* genome.

4.5 Conclusions

PCR based RACE approach was used to fish out the first R2R3-type MYB gene, *Ll*MYB_SSM (GenBank: GU901209) from *Leucaena leucocepahala*. The full length gene of 705 bp, encoding for 235 amino acid long polypeptide was characterized using various bioinformatic tools. Putative polyadenylation site and the poly A tail was identified in the 3' UTR of the gene. The translated amino acid sequence of the gene using BLAST analysis revealed about 59-66% similarity with the functionally reported

R2R3Myb proteins in the database. Analysis of deduced amino acid sequence of *Ll*MYB_SSM revealed highly conserved R2 and R3, DNA-binding domains, including the highly divergent C-terminal region. Phylogenetic analysis of deduced amino acid sequence was done using 28 protein sequences; 25 of them belonging to subgroup 4 of the R2R3MYB gene family from different taxa using the neighbor joining method. The results showed that isolated *Ll*MYB_SSM is evolutionarily most similar to *Am*MYB308, especially in the R2 and R3 domains (98% identity) than other R2R3MYB genes. Apart from this a highly conserved C2- signature motif for subgroup4 genes was also detected in the C-terminal region of the gene, having a repressor function. Banding pattern in Southern hybridization suggested that at least 3-4 copies or isoforms of *Ll*MYB_SSM gene are present in *L. leucocephala* genome. Considering the information retrieved from its bioinformatic characterization, functional characterization of *Ll*MYB_SSM has been taken up for establishing its putative role in lignification in *L. leucocephala* in chapters 5 and 6.

CHAPTER 5



Heterologous expression, purification and characterization of the MYB transcription factor

5.1 Introduction

The R2R3 two-repeat MYB family occurs specifically in plant lineages and its members have been postulated to participate in the regulation of a wide range of developmental and metabolic processes, notably the phenylpropanoid biosynthetic pathway (Du *et al.,* 2009). Although the involvement of MYB proteins in the monolignol-specific branch pathway has been suggested (Demura and Fukuda, 2007), the functional role of the conserved MYB *cis*-elements (also called AC elements) and the molecular mechanism, by which they participate in promoter activities has not yet been investigated.

Several MYBs from pine, *Eucalyptus* and *Arabidopsis* have been shown to be strong candidates as transcriptional activators of lignin biosynthetic genes (Patzaff *et al*., 2003 a, b; Zhou *et al*., 2009), which have been shown to bind to the AC elements and directly activate the expression of lignin biosynthetic genes as supported by the experimental findings. However, the regulatory cascade of lignification shows that several R2R3 MYB genes act as upstream transcription factors that control the formation of secondary walls by activating a range of other transcription factors (Zhong *et al*., 2010), which explains why several of the currently described MYB transcription factors lead to enhanced or reduced lignification when mis-expressed in plants while do not directly regulate the lignin biosynthetic genes by binding to their promoter regions (Zhong *et al*., 2006, 2008). Therefore, yeast one-hybrid assays, protoplast transient expression assays; and electrophoretic mobility shift assays have been essential to prove the direct binding of a given transcription factor.

Functional analysis of any gene requires it to be expressed in biologically active form in a suitable expression system. Expression of genes in heterologous organisms has allowed the isolation of many important genes (e.g. for nutrient uptake and transport) and has contributed a lot to the functional analysis of the gene products. Heterologous expression systems are powerful tools for isolating new genes and for characterizing proteins from all organisms. Heterologous expression of plant genes provides a new technique for determining gene-product function. *E. coli* remains a valuable organism for the high-level production of recombinant proteins, especially those from eukaryotes in functional form (Das *et al*., 1990; Hewitt and McDonnell, 2004). *E. coli* expression systems have been genetically modified to meet several criteria including:

(I) Minimal basal expression of the gene to be expressed under repressed conditions,

(II) Fast and uncomplicated induction of a wide variety of genes to a high level of expression,

(III) Easy cloning and DNA manipulation features.

The most common expression system is the T7 expression system derived from bacteriophage T7. The T7 expression system is based on the use of the T7 bacteriophage promoter and RNA polymerase. The T7 RNA polymerase is useful for synthesizing selectively large amounts of RNA because the T7 RNA polymerase recognizes only the T7 promoter and not the *E. coli* promoters. Conversely, the *E. coli* RNA polymerase does not recognize the T7 promoter. The T7 RNA polymerase is able to transcribe genes five times faster than the *E. coli* RNA polymerase. The gene encoding the T7 RNA polymerase was inserted into the chromosome of the bacteria used for over expression. Expression of the target gene is induced by addition of IPTG (isopropyl-β-D-thiogalactopyranoside) to a growing culture.

However, production of optimum protein depends on many factors like, cell growth characteristics, expression levels, intracellular and extracellular expression, and biological activity of the protein of interest. pET expression vectors, which are medium-copy plasmids, offer several useful features for expressing the target gene of interest in a compatible host strain of *E. coli* (BL21), for achieving some of the mentioned criteria. The heterologously expressed protein can be thus utilized for various *in vitro* studies like raising antibodies or EMSA by binding to specific DNA sequences.

However, binding to DNA is not always synonymous with transcriptional regulation function of transcription factor, which can be addressed only by *in vivo* studies in a plant system. Moreover, post-translational modifications (phosphorylation, nitrosylation or ubiquitination) and protein–protein interactions (with bHLH proteins) are also known to have a significant impact on the regulatory activity of MYB transcription factors (Daniel *et al*., 2004, Zimmermann *et al*., 2004). Hence, *in planta* overexpression in a heterogenous plant system would help in establishing the role of MYB transcription factors in regulating the biological functions of a plant.

Based on these findings, in an attempt to get an insight of the central objective of the thesis, a cDNA encoding an R2R3MYB transcription factor, *Ll*MYB_SSM from *Leucaena leucocephala* was cloned and characterized as described in the previous chapter. To study its functional expression, the cDNA has been cloned into the expression vectors pET28b+ and pET41a+. Heterologous expression of the recombinant protein in BL21 (DE3) and its purification using Ni-NTA affinity column from inclusion bodies (pET28b+) and using GST-bind resin in soluble form (pET41a+) has been shown. Ni-column purified protein was used for raising polyclonal antibodies while GST-bind-column purified protein in active form was used for gel-retardation studies. Molecular modeling of the protein has been described based on the primary structure of the predicted protein in the subsequent section, to gain further understanding of the three dimensional structure of the protein. Finally, *in planta* studies by heterologous expression of *Ll*MYB_SSM was carried out in *Nicotiana tobacum* to address the issue of post-translational modifications and partner proteins required for regulatory activity of the protein.

5.2 Materials and methods

5.2.1 Materials

Glassware and plasticware: As discussed in chapter 2, sections 2.4 and 2.5

Chemicals: As discussed in chapter 2, section 2.6

Bacterial strains and plasmids used in the study

Escherichia coli XL-1 Blue (Stratagene, USA)

E. coli BL 21(DE3) (Invitrogen, USA)

Agrobacterium tumefaciens GV2260

pGEM-T Easy Cloning vector (Promega, USA)

pET28b (+) Expression vector (Novagen, USA)

pET41a (+) Expression vector (Novagen, USA)

pCAMBIA 1301 binary cloning plant transformation vector (CAMBIA)

5.2.2 Methods

5.2.2.1 Bacterial culture conditions: As discussed in chapter 2, section 2.9.1

5.2.2.2 Bacterial cells transformation: *E. coli* transformation and selection was done as discussed in chapter 2, section 2.9.2.

5.2.2.3 Isolation of plasmid DNA from *E. coli*: As described in chapter 2, section 2.9.4.1.

5.2.2.4 Restriction digestion of DNA: As described in chapter 2, section 2.9.4.3

5.2.2.5 Extraction and purification of DNA from agarose gels: As described in chapter 2, section 2.9.4.4.

5.2.2.6 Polymerase Chain Reaction (PCR): As described in chapter 2, section 2.9.4.9.

5.2.2.7 Colony PCR method: As described in chapter 2, section 2.9.2.5.

5.2.2.8 Cloning of *Ll*MYB_SSM cDNA in pET28b + expression vector

The pET-28b (+) vector (Novagen, USA) was used for expression of *Ll*MYB_SSM gene in *E. coli* BL21 (DE3). This vector carries an N-terminal His•Tag®/thrombin/ T7•Tag® configuration plus an optional C-terminal His•Tag sequence. This vector has unique restriction sites for cloning (Fig. 5.1). The cloned gene is expressed under the T7 RNA

polymerase promoter. The sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circular map.

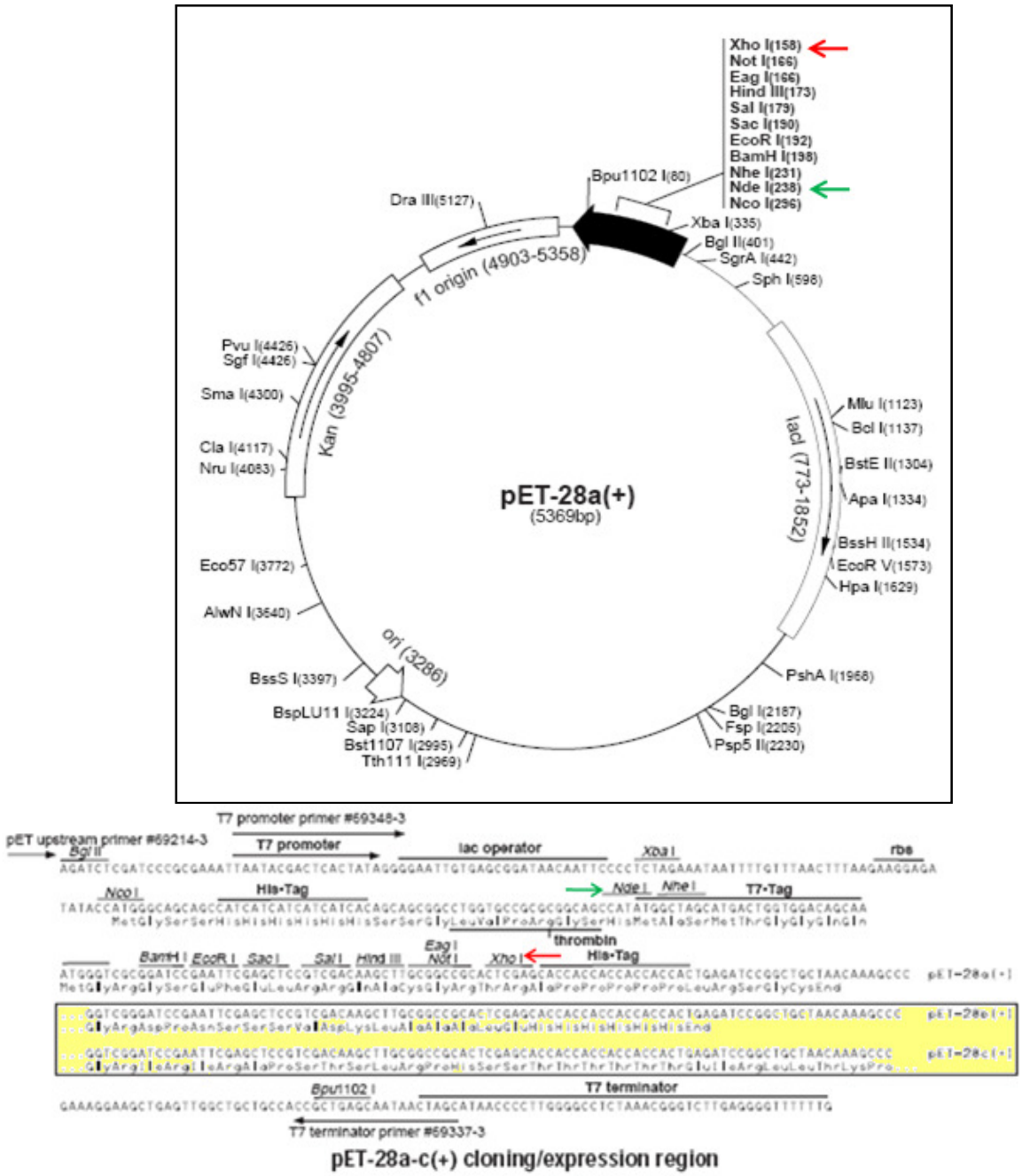


Fig. 5.1: Vector diagram of expression vector pET 28 a, b and c (+).The maps for pET 28b (+) and pET-28c (+) are the same as pET 28a (+) (shown) with the following exceptions: pET-28b (+) is a 5421 bp plasmid; subtract 1 bp from each site beyond *Bam*H I at 198. pET-28c (+) is a 5423 bp plasmid; add 1 bp to each site beyond *Bam*H I at 198. The sequence of pET 28b (+) is highlighted in yellow. Arrows indicate the sites where *Li*MYB_SSM gene is cloned in frame with the promoter.

5.2.2.9 Expression and purification of recombinant protein from inclusion bodies

E.coli BL21 (DE3) cells transformed with recombinant pET 28b (+) plasmid were screened for over-expression as mentioned in chapter 2, section 2.9.11.1. The clones showing optimum expression were preserved by preparing glycerol stocks for future studies (chapter 2, section 2.9.3).

5.2.2.9.1 Recombinant protein expression in *E. coli* (BL21)

A single bacterial (BL21) clone carrying recombinant pET-28b (+) with *Ll*MYB_SSM gene was inoculated from the glycerol stock culture in 5 mL liquid LB medium containing 50 µg / mL kanamycin and grown overnight with shaking at 200 rpm at 37 °C. 500 µL of the culture was transferred to 50 mL LB-kan media ($1/100^{th}$ of cultured volume) and grown at 37 °C for 1-2 h until the cultures reached A_{600} of 0.4 to 0.5. The culture was then induced with 1 mM IPTG and grown upto 6 h. After 6 h, the culture was harvested for soluble (from lysate) and insoluble (from inclusion bodies dissolved in 8M Urea buffer) protein as described in chapter 2, section 2.9.11.2 and the expression profile was checked by running a 12% SDS PAGE gel, as the expected size of the recombinant protein was ~30kD.

5.2.2.9.2 Affinity purification of *Ll*MYB_SSM Protein from inclusion bodies

His-tagged recombinant protein was purified by Metal Chelate Affinity Chromatography. The initial stage of His-tagged protein purification is based on the remarkable selectivity and high affinity of patented Ni-NTA (nickel-nitrilotriacetic acid) resin for proteins containing an affinity tag of six consecutive histidine residues, the 6xHis-tag. NTA, which has four chelating sites for nickel ions, binds nickel more tightly than metal chelating purification systems that have only three sites available for interaction with metal ions. The extra chelation site prevents nickel ion leaching, providing a greater binding capacity and high-purity protein preparations. Purification of recombinant *Ll*MYB_SSM protein carrying two 6xHis-tags was conducted using Ni NTA Agarose beads (Qiagen) (chapter 2, section 2.9.11.3). Purity of protein was checked on 12% SDS-PAGE (chapter 2, section 2.9.11.5).

5.2.2.10 Protein estimation

Protein estimation was done using Bradford assay (1976). A standard graph was made for BSA and concentration of unknown sample was determined by plotting standard graph (chapter 2, section 2.9.17).

5.2.2.11 MALDI MS-MS analysis of purified protein

For MALDI analysis, 10% SDS –PAGE was run for the *Ll*MYB_SSM recombinant protein. Gel was then stained with Coomassie Blue stain. Expected stained protein bands were excised from the gel. The gel pieces were then destained by destaining solution (50% acetonitrile (ACN) / 50% 50 mM NH_4HCO_3) till colour was gone. Gel was then dehydrated by treating with 100% ACN. After dehydration, ACN was completely removed by evaporating briefly in speedvac till noticeably shrunken and white. Then gel piece was dissolved in 10 mM DTT in 100 mM NH_4HCO_3 and protein was reduced for 45-60 min at 56 °C. After cooling to room temperature, DTT solution removed and 55 mM iodoacetate in 100 mM NH_4HCO_3 was added. This was followed by iodoacetamide removal and washing of gel pieces with 100 mM NH_4HCO_3 for 5 min. Again the gel pieces were washed twice with 50% ACN / 50% 50 mM NH_4HCO_3 and dehydrated with 100% ACN as mentioned above. Then enough trypsin solution was added to cover the gel pieces (usually around 20 µL) and the gel pieces were rehydrated at 4 °C for 30 min in buffer containing 50 mM NH_4HCO_3 and trypsin. Spun briefly and more NH_4HCO_3 was added to cover gel pieces (typically another 25 µL). This was followed by overnight digestion at 37 °C. The digested solution (supernatant) was transferred into clean 1.5 mL microfuge tube. 50% ACN /2% formic acid solution was added to the gel pieces which was incubated and vortexed for 20 min. This was spun and sonicated for 5 min in a water bath with no heat. Supernatant was then removed and combined with initial digestion solution (supernatant). Digested extract was then vortexed and evaporated to reduce to 5-10 µL. The remaining 5-10 µL was spun at 14k rpm for at least 10 min to remove any micro particulates. The supernatant was carefully transferred to a fresh 1.5 mL microfuge tube. The sample was then ready for loading on to MALDI MS/MS. Simultaneously MALDI plate was also washed in order to remove the particulate matter deposited, if any. Then sample and MALDI matrix (α-cyano 4-hydroxy cinnamic acid) in the proportion of

2:1 ratio were loaded on to the MALDI plate. This was followed by reading of MALDI plate with laser energy of about 280 V using MALDSNYPT equipment (Waters).

5.2.2.12 Raising polyclonal antibody against purified *Ll*MYB_SSM protein in rabbit

The purified *Ll*MYB_SSM protein was used for raising polyclonal antibodies in New Zealand White rabbit.

5.2.2.12.1 Pre-treatment of serum:

Rabbit immune serum (stored at -70 ºC), containing polyclonal antibodies against the antigen, was thawed overnight at 4 ºC. The serum was kept at 55 °C for one hour to deactivate the complement system and the immune serum was centrifuged at 12,000 g for 15 min at 4 ºC. Aliquots according to the requirements were made and stored at -70 ºC. Prior to use again antiserum was thawed overnight at 4 ºC.

5.2.2.12.2 Determination of titre of antibodies

ELISA was performed to determine the titre of first bleed of rabbit serum (chapter 2, section 2.9.14.1). Once the antibody titre was determined then, a fixed dilution of antibody was used for rest of the experiments.

5.2.2.13 Cloning of *Ll*MYB_SSM cDNA in pET41a + expression vector

It is desirable to express proteins in their soluble, active form for functional studies. The choice of vector and expression host can significantly increase the activity and amount of target protein present in the soluble fraction. One of the ways in which a vector can enhance solubility and/or folding, is by providing fusion to a polypeptide that itself is highly soluble [e.g. glutathione-S-transferase (GST), thioredoxin (Trx)]. These tags can also facilitate detection and purification of the target protein, or may increase the probability of biological activity by affecting solubility in the cytoplasm.

The pET-41 series is designed for cloning and high-level expression of peptide sequences fused with the 220 aa GST•Tag™ protein. Vector encoded sequence can be completely removed when cloning into the *Psh*AI site (as shown below) and then cleaving the GST fusion protein with Enterokinase. pET 41a(+) vector was chosen to express GST-fused *Ll*MYB_SSM recombinant protein for increasing its solubility.

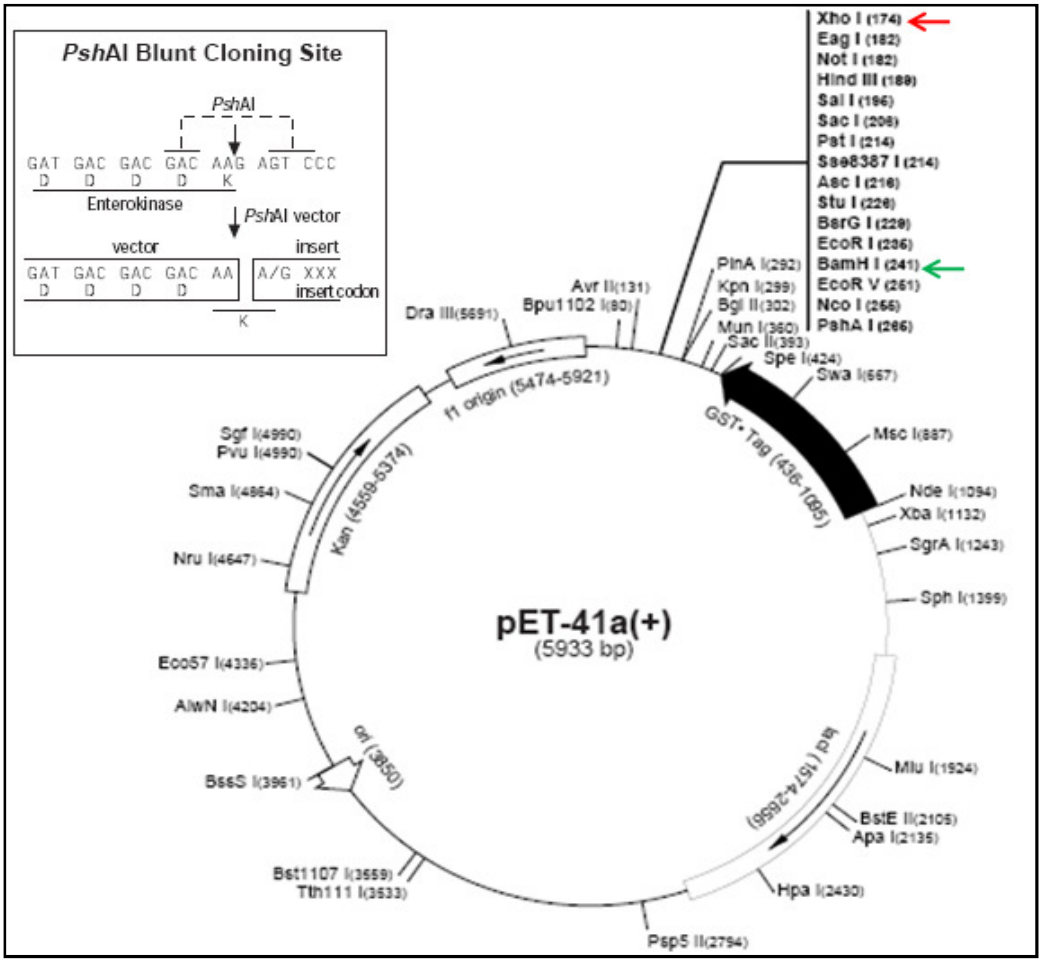


Fig 5.2: Vector diagram of expression vector pET 41 a(+) 5933 bp plasmid. Arrows indicate the sites where *Ll*MYB_SSM gene is cloned in frame with the promoter.

5.2.2.14 Expression and purification of *Ll*MYB_SSM in soluble form

E. coli BL21 (DE3) cells transformed with recombinant pET 41a (+) plasmid were screened for over-expression as mentioned in chapter 2, section 2.9.11.1. The clones showing optimum expression were preserved by preparing glycerol stocks for future studies (chapter 2, section 2.9.3).

5.2.2.14.1 Recombinant protein expression in *E. coli* (BL21)

A single bacterial (BL21) clone carrying recombinant pET-41a(+) with *Ll*MYB_SSM gene was inoculated from the glycerol stock culture in 5 mL liquid LB medium containing 50 µg / mL kanamycin and grown overnight with shaking at 200 rpm at 37 °C. 500 µL of the culture was transferred to 50 mL LB-kan media (1/100th of cultured

volume) and grown at 37 °C for 1-2 h until the cultures reached A_{600} of 0.4 to 0.5. The culture was then induced with 0.5 mM IPTG and grown upto 16 h at 18 °C, as the GST-tag shows the optimum yield at this temperature. After 16 h, the culture was harvested for soluble protein (from lysate) as described in chapter 2, section 2.9.11.4 and the expression profile was checked by running a 10% SDS PAGE gel.

5.2.2.14.2 Affinity Purification of pET41a (+) cloned *Ll*MYB_SSM protein using GST•Bind ™ Resin

The recombinant protein having a GST-tag was purified as per the protocol mentioned in chapter 2, section 2.9.11.4. The purified protein was checked for its purity on 10% SDS gel, followed by silver-staining.

5.2.2.15 Gel-retardation assay of purified protein with AC-elements

As shown in previous studies, the GST-tag was retained in the purified recombinant protein to carry out gel retardation experiments. The purified protein was quantified using Bradford Assay and used for DNA-binding experiments with the partial promoter regions harboring the MYB-binding elements (AC-elements) shown in chapter 3, section 3.3.4.3. The EMSA protocol was followed as per chapter 2, section 2.9.6.

5.2.2.16 Molecular modeling of *Ll*MYB_SSM protein using I-TASSER tool

The three-dimensional structure of *Ll*MYB_SSM was generated using the multiple threading alignments and iterative structural assembly simulations using iterative threading assembly refinement (I-TASSER) server (Roy *et al*., 2010). Further, the predicted structural model was evaluated by using structure quality assessment programs like PROCHECK, ProSA and Verify 3D.

5.2.2.17 Construction of *Ll*MYB_SSM-sense construct in pCAMBIA 1301 vector

To express the complete gene and monitor its effect on the lignin biosynthetic pathway genes in *Nicotiana tobacum,* the complete open reading frame of the *Ll*MYB_SSM was cloned in sense orientation in the modified pCAMBIA1301 vector available in the lab. In pCAMBIA1301- *Ll*MYB_SSM construct the insert was cloned in the sense orientation.

The map of this construct between the left and right border is shown in fig. 5.3. The *Li*MYB_SSM cDNA cloned in pCAMBIA 1301 plant transformation vector is under the transcriptional control of CaMV 35S promoter and poly-A sequences from nopaline synthase gene. The recombinant plant transformation vector further comprises a selection cassette containing hygromycin marker gene for selection of transformed plants and plant cells, also under the control of CaMV 35S promoter and poly-A sequences derived from nopaline synthase. The vector also includes a reporter gene construct comprising of GUS gene operably linked to CaMV 35S promoter and poly-A sequences from nopaline synthase gene.

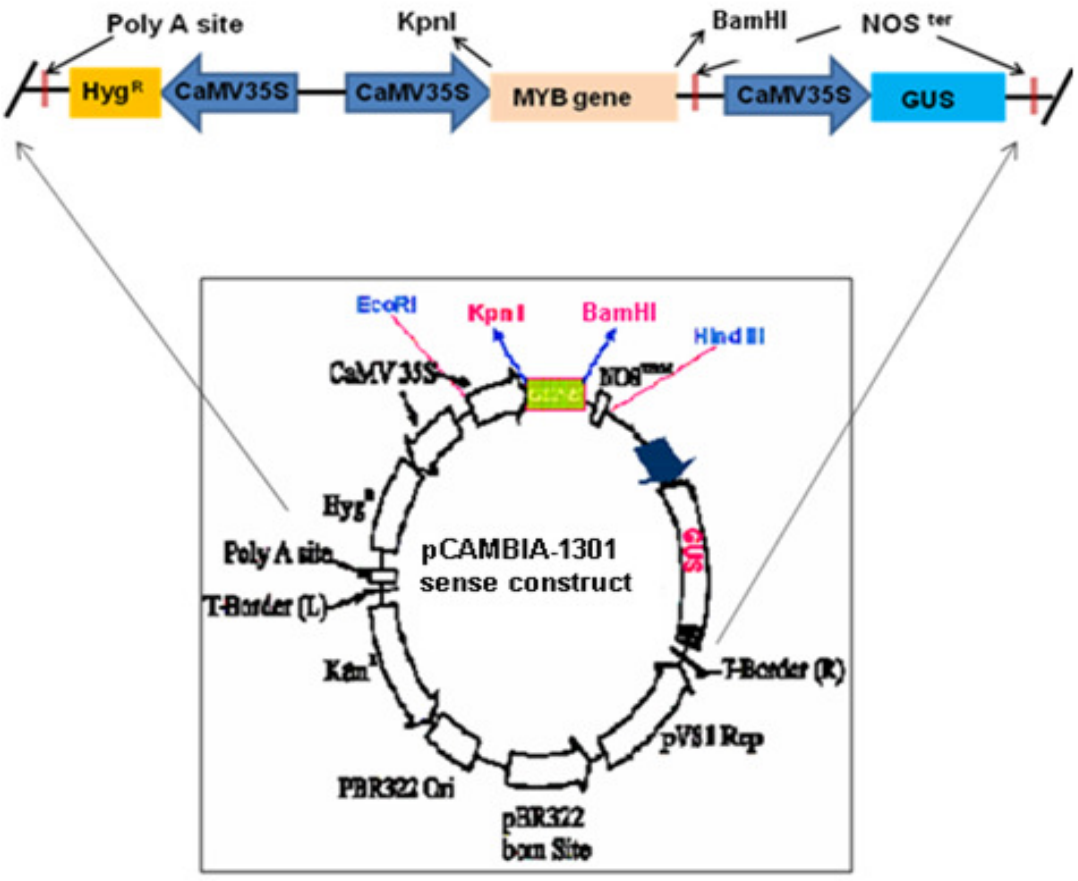


Fig. 5.3 Detailed Diagram of the T-DNA construct of *Li*MYB_SSM gene cloned in sense orientation in pCAMBIA-1301 vector

5.2.2.18 *Agrobacterium*-mediated transformation of tobacco

Tobacco (*N. tabaccum* var. Anand 119) leaf discs were transformed using *A.tumefaciens* culture harboring the pCAMBIA1301-*Ll*MYB_SSM construct as per the protocol described in chapter 2, section 2.9.15.

5.2.2.19 GUS Assay and PCR for analysis of transformants

GUS assay for the putative transformants at early stages was performed as per the protocol described in chapter 2, section 2.9.16. DNA isolation for putative transformants was done with a single leaf for each plant based on the modified protocol described by Lodhi *et al* (1994) (chapter 2, section 2.9.4.2) and used for carrying out PCR with hygromycin phosphotransferase gene (*hptII)* specific primers as described in section 5.3.8.3.2 of this chapter for confirmation of the transformants.

5.2.2.20 Quantitative Real-Time PCR for analysis of putative transgenic tobacco plants

RNA extraction and cDNA synthesis from tobacco leaves of the putative transgenics was performed as for *L. leucocephala* (chapter 2, sections 2.9.4.5 and 2.9.4.8 respectively) and real-time RT-qPCR analyses were performed as described in chapter 2, section 2.9.8.2 with minor modifications. For QRT-PCR quantification of each target RNA, a 5-fold dilution of cDNA mixture was used as template. In the present study MESA GREEN qPCR Master Mix Plus, low ROX (Eurogentec) was used for SYBR Assay, with Meteor *Taq* DNA polymerase, which is a hotstart enzyme.

5.3 Results

5.3.1 Construction of *Ll*MYB_SSM protein expression vector in pET28b (+) and pET41a (+) vectors

pET28b(+) vector was chosen for cloning *Ll*MYB_SSM for expression studies as it is one of the simplest vectors providing only the 6-histidine-tag to facilitate protein purification with an inducible T7 *lac* promoter.

pET41a (+) vector was chosen for cloning *Ll*MYB_SSM for expression studies to obtain the protein in soluble form in fusion with the GST-tag. In addition to enhancing the solubility of the partner protein, the tag is highly specific for affinity purification of the protein in active form as it can be performed in mild conditions.

5.3.1.1 Incorporation of restriction sites

A high fidelity *Taq pol* (LA-Taq DNA polymerase, Sigma-Aldrich, USA) was used to amplify *Ll*MYB_SSM gene to avoid any mis-incorporation of nucleotide and the consequent frame-shift in the open reading frame, using restriction site incorporated forward and reverse primers (shown below) from the cDNA clone, sequenced and maintained in pGEM-T Easy vector as the template. *Nde*I and *Bam*HI sites were incorporated in the forward primers for cloning in pET28b (+) and pET41a (+) respectively, while *Xho* I site incorporated reverse primer was common for both the vectors.

Primer name	**Primer sequence**
SSM*Nde*I-F	5’- *CATATG*GGGAGAGCTCCGTGCT -3’
SSM*Bam*HI-F	5’- *GGATCC*GGGAGAGCTCCGTGCT -3’
SSM*Xho* I-R	5’-*CTCGAG*TTCTGTGAGTGCTCTCCA-3’

PCR cycling conditions:

Temperature	Time	Cycles
94 °C	5 min	1
94 °C	30 s	35
60 °C	30 s	
72 °C	1 min	
72 °C	5 min	1
4 °C	hold	1

~700bp fragment was amplified (exactly 705 bp coding region of *Ll*MYB_SSM gene + *Nde*I/ *Bam*HI and *Xho* I sites added into primer sequences). The fragment was excised from gel, purified, ligated in pGEM-T Easy vector and transformed in *E. coli* XL10 cells. Clones with *Ll*MYB_SSM gene with flanking restriction sites were screened by colony PCR and a few colonies were inoculated in 5 mL LB (Ampicilin 100 µg/mL) tubes. Isolated individual plasmids were restriction digested with *Nde*I/*Bam*HI and *Xho*I enzymes to confirm the integration of *Ll*MYB_SSM gene as well as obtaining the insert to be directionally cloned in the corresponding sites in pET28b (+) and pET41a (+) vectors and named **pGEM-28-*Ll*MYB_SSM** and **pGEM-41-*Ll*MYB_SSM** respectively.

5.3.1.2 Directional cloning of *Ll*MYB_SSM gene in pET28b (+)/ pET41a (+)

Double digestion reaction

Both the blank vector plasmids pET28b(+)/ pET41a(+) and pGEM-28-*Ll*MYB_SSM/ pGEM-41-*Ll*MYB_SSM plasmids (to obtain insert) were separately kept for double digestion using restriction enzymes (*Nde*I/ *Bam*HI and *Xho*I) in the presence of Buffer D (Promega).

Plasmid DNA	4 µL
Buffer D(10X)	3 µL
*Nde*I / *Bam*HI	1 µL
*Xho*I	1 µL
BSA	1 µL
SMQ	20 µL
Total reaction	**30 µL**

The mixture was kept at 37 $^{\circ}$C for 3-4 h and run on 1% agarose gel with medium range marker. The digested insert and vector (Fig. 5.4 A and B) were excised from the gel, eluted and quantified for deciding the vector: insert ratio to set up ligation reaction.

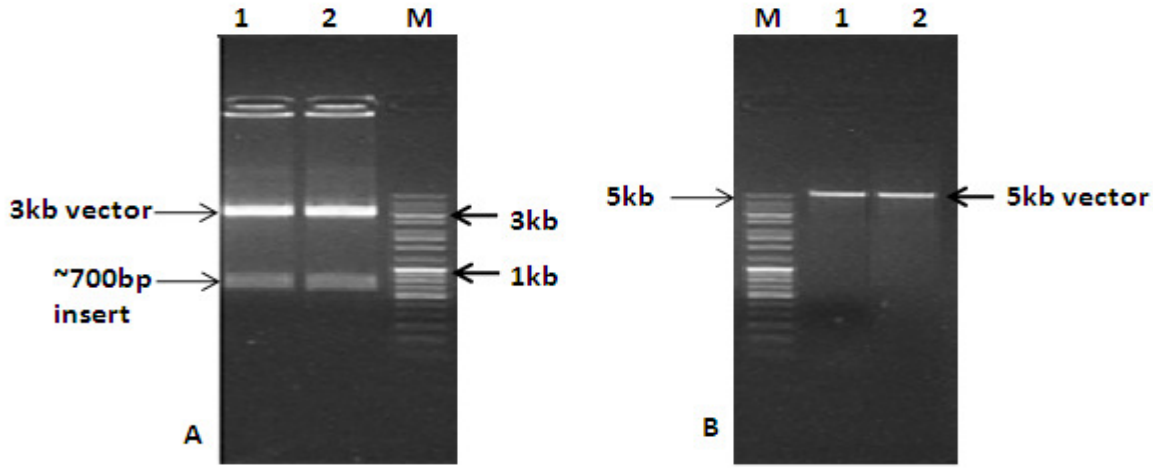


Fig. 5.4 (A): pGEM-28-*Ll*MYB_SSM/pGEM-41-*Ll*MYB_SSM plasmids double digested with *Nde*I- *Xho*I and *Bam*HI-*Xho*I respectively. Lanes 1 & 2: 3kb vector backbone and ~700bp insert, Lane M: MR DNA ladder. **(B): pET28b (+)/ pET41a (+) vectors double digested with *Nde*I- *Xho*I and *Bam*HI-*Xho*I respectively.** Lane M: MR DNA ladder, Lanes 1 & 2: Linearised vector backbone.

Ligation reaction

The digested pET28b (+)/ pET41a (+) vectors and insert were subjected to ligation reaction. The ligation mixture was kept at 16 °C overnight and used for transformation of XL10 cells. Colony PCR (chapter 2, section 2.9.2.5) was done to screen the recombinant pET-28b (+)/ pET41a (+) clones in XL10 cells (Fig. 5.5A)

Vector	2 µL
Insert	8 µL
Ligase buffer (10X)	1.5 µL
T_4 DNA ligase	2 µL
ATP (100mM)	1.5 µL
Total reaction	**15 µL**

Integration of *Ll*MYB_SSM gene in pET-28b (+)/pET-41a (+) was confirmed by digestion with *Nde /Bam*HI and *Xho*I (Fig. 5.5 B). Plasmids isolated from the positive clones were further confirmed by sequencing for correct reading frame and the correctly cloned recombinant plasmids were mobilized in BL21 (DE3) cells for expression studies.

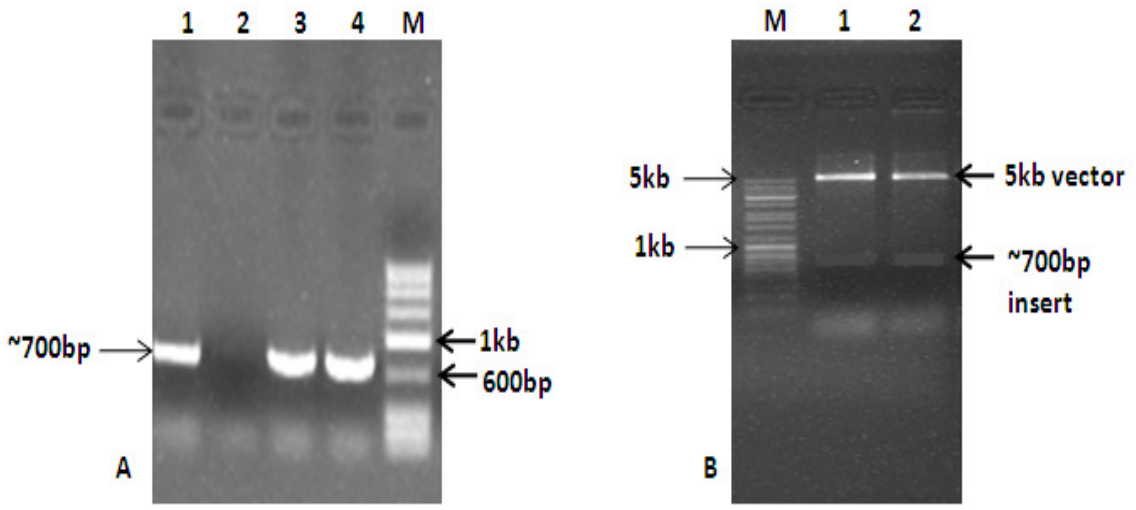


Fig. 5.5 (A): Colony PCR for recombinant clones with *Ll*MYB_SSM insert. Lanes 1, 3 & 4: ~700 bp PCR amplicons obtained in positive clones, Lane M: LR DNA ladder. **(B): Restriction digestion of recombinant pET28b (+)/pET41a (+) vector.** Lane M: MR DNA ladder, Lanes 1 & 2: Recombinant plasmids releasing ~700 bp insert from 5kb vector on digestion with *Nde*I/ *Bam*HI and *Xho*I.

5.3.2 Heterologous expression of pET28b (+)-*Ll*MYB_SSM protein in BL21 cells

E. coli BL21 (DE3) was used as a host for the expression of mature *Ll*MYB_SSM protein. A pilot experiment with 5 ml culture of cells induced with 1mM IPTG was used to screen the clones showing optimum expression of the recombinant protein. For protein visualization cells were harvested at 5 h after IPTG induction and isolated by centrifugation. The pelleted cells were resuspended in sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) loading buffer and heated at 95 °C. Proteins were separated by SDS-PAGE on 12% gel and stained with Coomassie blue. Molecular mass of the expressed protein was ~30 kDa as confirmed by SDS- PAGE and consistent with the predicted sequence including the His-tags. The expression level of the *Ll*MYB_SSM protein was clearly detectable in induced cells as compared to uninduced bacterial cells, where the expression level was merely detectable (Fig. 5.6).

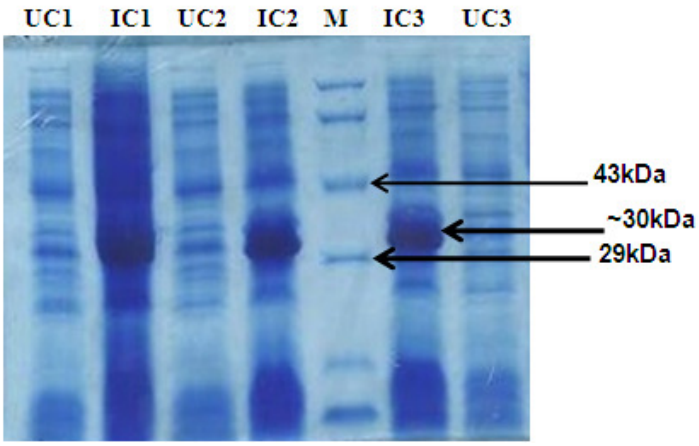


Fig. 5.6: Crude protein lysate from recombinant clones run on 12% SDS gel; Lanes: UC1, UC2 & UC3: Uninduced Crude lysate from different clones, IC1, IC2 & IC3: Induced Crude lysate for the corresponding uninduced cultures, showing expression of ~30kDa protein, M: Medium Range Protein Molecular Weight Marker

However, this study doesn't show whether the protein is getting expressed in the soluble form or it is embodied in as inclusion body. To clarify the obvious doubt, cytosolic as well as insoluble fractions were prepared and analyzed for the *Ll*MYB_SSM protein expression. The study revealed the accumulation of *Ll*MYB_SSM protein mainly in inclusion bodies (insoluble aggregates of denatured or partly denatured protein) that lack biological activity but often give a high yield of the recombinant protein.

5.3.2.1Standardisation of IPTG concentration for protein expression in soluble form:

Temperature and IPTG concentration are two important factors which influence protein expression in soluble fraction of the cell lysate. Preliminary screening of the best expressing clone at 37 °C with 1mM IPTG induction showed that most of the protein was accumulated in the inclusion bodies. So, optimization of temperature was tried with 1mM IPTG concentration at 28 °C, 25 °C, 20 °C, 18 °C and 15 °C. Results (not shown here) showed that growth of culture at 15 °C resulted in comparatively better localization of the protein in soluble form. However, the yield was not very effective for purification, so different IPTG concentrations (0.5 mM, 0.1mM and 0.05 mM) were used for induction at 15 °C for 15h. Analysis of result showed that 0.5 mM IPTG concentration was most effective for protein expression in soluble form, as equal amount of protein was seen in

both inclusion bodies and lysate (Fig. 5.7 A, lanes 1&2) while, 0.05 mM was the least effective. Therefore, 0.5 mM IPTG concentration was utilized for obtaining the protein in soluble form and purification. However, purification from the lysate resulted in a very low yield of purified protein along with several contaminating bands. So, purification of recombinant protein was tried from the inclusion bodies as shown in following section.

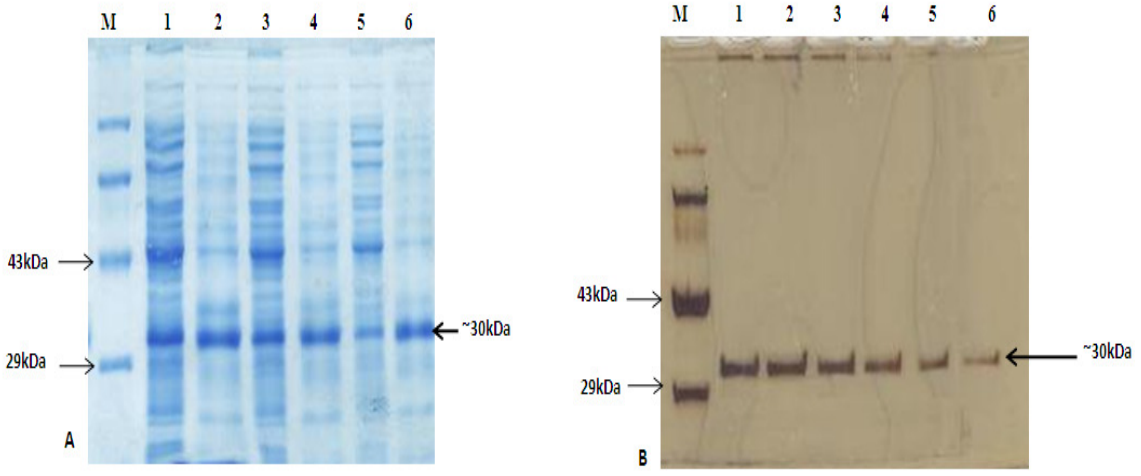


Fig. 5.7 (A): Expression profile for one of the pET28b-*Ll*MYB_SSM BL21 clones with different IPTG concentration grown at 15 °C for 15h: Equal quantity of total protein was loaded in each well. Lane M: MR protein marker, Lanes 1 & 2: lysate and IB with 0.5 mM IPTG, Lanes 3&4: lysate and IB with 0.1 mM IPTG and Lanes 5&6: lysate and IB with 0.05 mM IPTG, respectively; **IB**: Inclusion bodies

Fig. 5.7 (B): Purification of over-expressed *Ll*MYB_SSM protein from Inclusion bodies using Ni-NTA affinity chromatography: Lane M: MR protein marker, Lanes 1-6: Elution fractions of the purified protein in chronological order.

5.3.2.2Purification of recombinant protein from inclusion bodies

Since the purification of recombinant protein in soluble form was not achieved in good amount, large-scale production and purification to raise antibodies was done from inclusion bodies as described in section 5.2.2.9.2 of this chapter. Once confirmed the protein expression in pilot experiments, the induction of protein expression was done at large scale. It was found that six hours of growth post-induction at 37 °C was enough to achieve the purification of protein from inclusion bodies. The protein isolation from inclusion bodies was done in denaturing conditions as the protocol prescribed in chapter 2; section 2.9.11.2. The protein was purified to homogeneity with Ni-NTA affinity chromatography as described in chapter 2; section 2.9.11.3. Different eluted fractions of

*Ll*MYB_SSM protein of ~30 kDa were separated under denaturing condition by 12% SDS-PAGE and visualized with silver staining (Fig. 5.7 B)

5.3.3 MALDI MS-MS analysis of recombinant *Ll*MYB_SSM protein

MALDI MS/MS (Matrix assisted laser desorption ionization, with tandem TOF/TOF analyser) was done for the purified gel digested protein as described in section 5.2.2.11of this chapter. MS/MS analysis (Tandem mass spectrometry) was done for confirmation of recombinant *Ll*MYB_SSM protein by comparing the peptide mass fingerprinting data with the data for sequences already submitted to the Uniprot database. The MALDI MS/MS ionization spectra and coverage map of the protein and the first best hit are shown in Fig. 5.8 and 5.9 respectively. Ionisation spectra shows mass of different peptides of the query protein and coverage map shows number of peptides of query protein which shows exact match with the template proteins present in the Uniprot database.

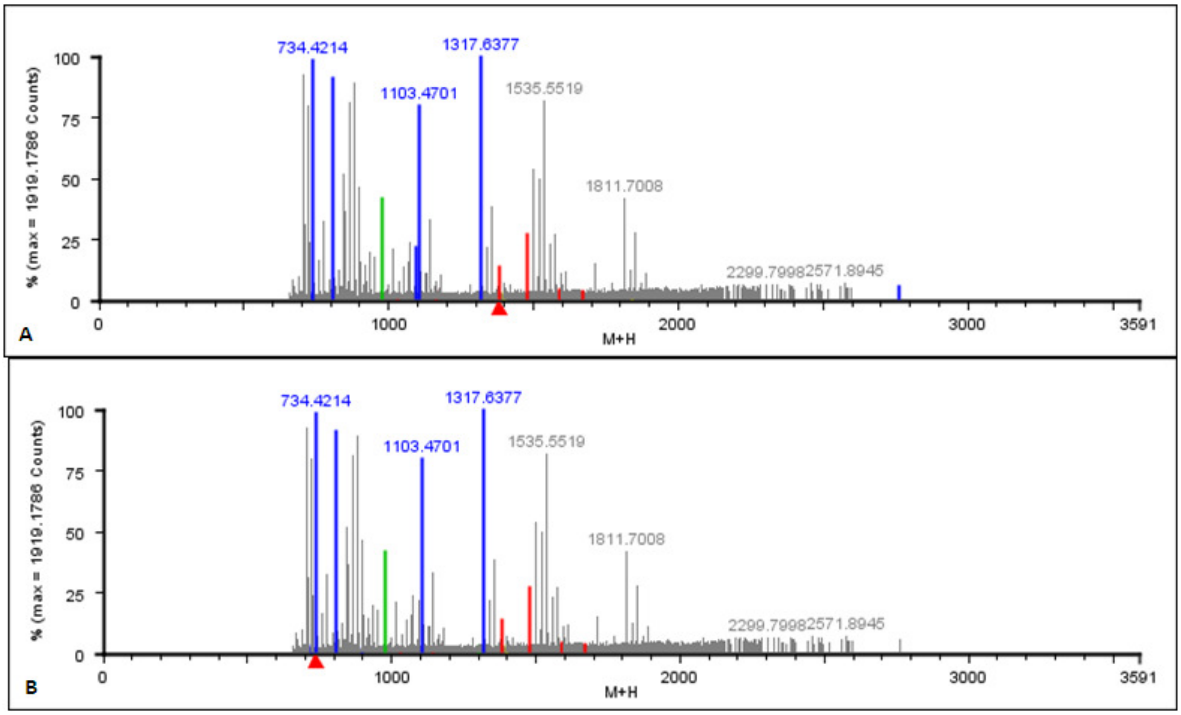


Fig. 5.8: Ionisation spectrum of peptidal cleavage identified by PLGS Global server for MALDI MS/MS of (A) R2R3MYB from *Leucaena leucocephala* with 17 peptides showing highest probability of 88.8% and **(B) *Am*MYB308 from *Antirrhinum majus*** with 15 peptides showing probability of 6.71%. The heights of the peaks correspond to the % intensity i.e. the probability of occurrence of the generated peptide mass.

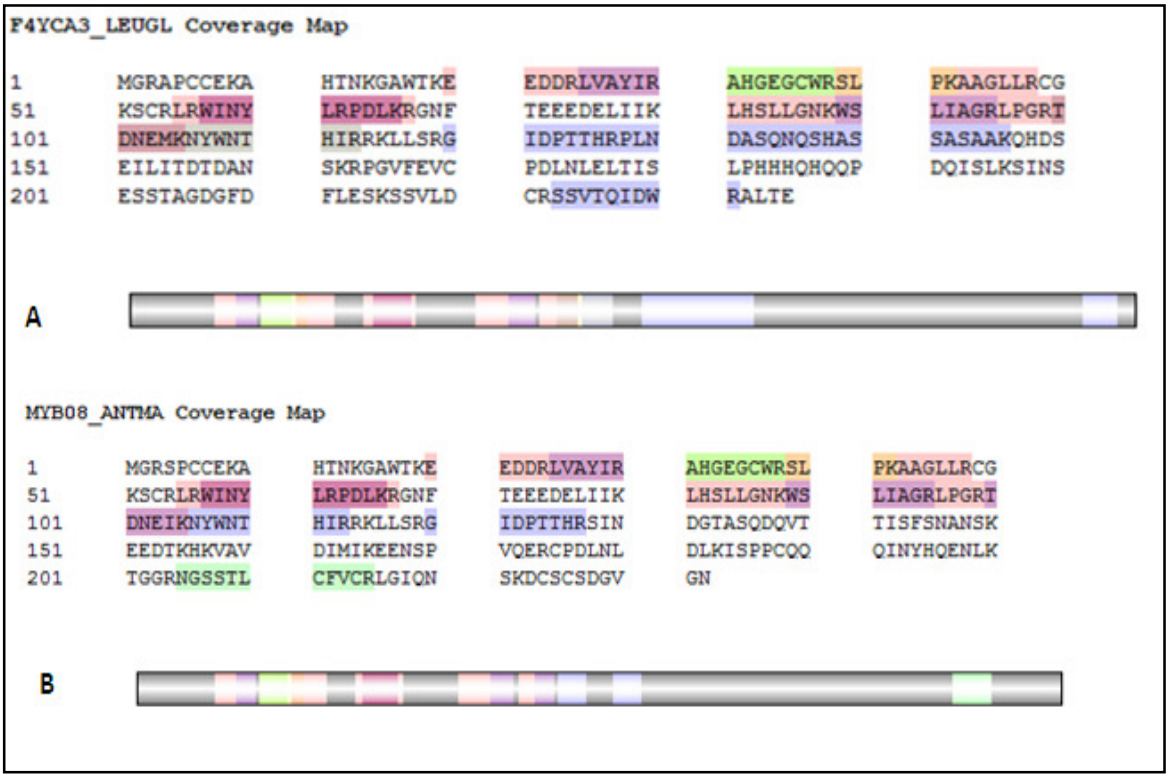


Fig. 5.9: Coverage map of peptidal cleavage identified by MALDI MS/MS (A): R2R3-type MYB transcription factor from *Leucaena leucocephala* with 47.65% coverage (B) MYB transcription factor AmMYB308 from *Antirrhinum majus* with 40.1% coverage. Regions of the protein sequence that match peptides are highlighted in colour, according to the key below:
Matched to a peptide: Matched to a partial peptide:
Matched to a modified peptide: Matched to a partial modified peptide: .
The highlights are transparent, so that regions where peptides overlap are visible on the coverage map. E.g. an overlap between a standard peptide and a modified peptide would appear as shown:

The data retrieved from the two figures show that *Ll*MYB_SSM shows the best similarity of its PMF data with the theoretical proteolytic peptidal data for *Am*MYB308. Most of the peptides with highest probabilty are seen mapping within the 120 amino acid sequence region which corresponds to the MYB DNA-binding domain of the two proteins (Fig. 5.9).

5.3.4 Raising polyclonal antibodies against purified *Ll*MYB_SSM protein in rabbit

Purified *Ll*MYB_SSM protein as shown in Fig.5.7 (B) was dialysed against 1X PBS buffer supplemented with 0.1 mM PMSF overnight in cold room with two changes of fresh buffer. Dialysed *Ll*MYB_SSM protein was concentrated and quantified by Bradford assay (Bradford reagent, 1976). This purified, desalted *Ll*MYB_SSM protein was used for raising polyclonal antibodies in rabbit.

5.3.4.1 Determination of titre of antibodies

500 µg of purified protein was used for first injection in New Zealand rabbit to raise antibodies. Same amount of protein was used for booster doses. 100 ng of the purified protein antigen was challenged to different serum dilutions of the 1st bleed such as, 1:1000, 1:5000, 1:10000, 1:20000, 1:40000, 1:80000, 1:100000 and 1:150000. Titer of first bleed of antiserum was determined to be 1:100000 (Fig. 5.10). 1st bleed serum dilution of 1:20000 was used for ELISA studies and 1:10000 was used for immuno-cytolocalization.

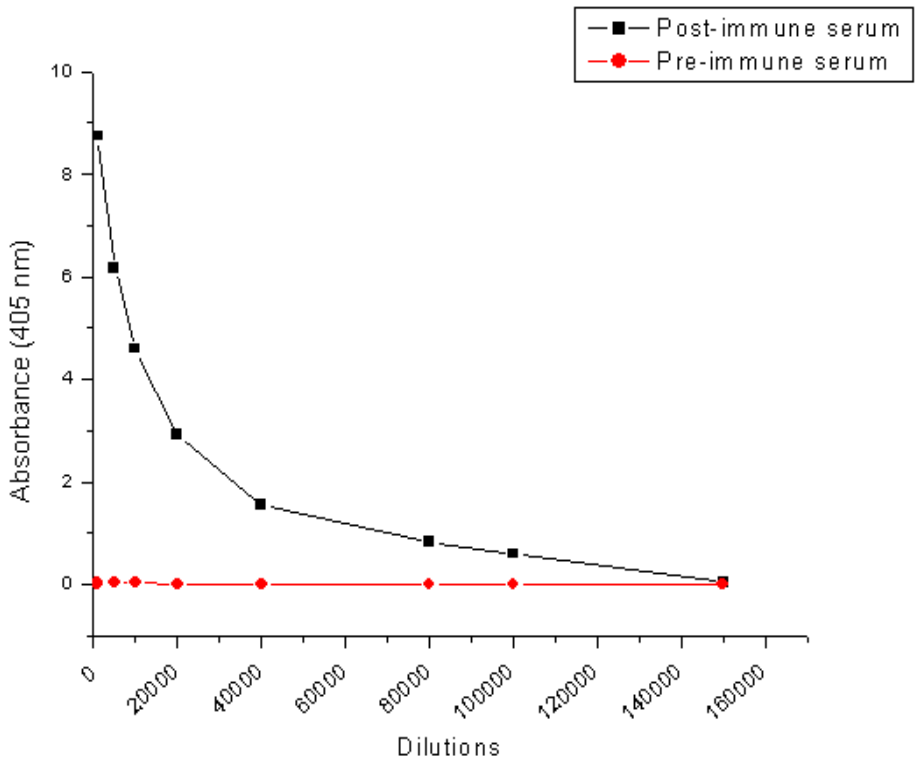


Fig. 5.10: Determination of titre of first bleed serum: Solid square represents A_{405} of serum dilution; a solid circle represents A_{405} of pre-immune serum.

5.3.5 Heterologous expression of pET41a (+)-*Ll*MYB_SSM protein in BL21 cells

E. coli BL21 (DE3) was used as a host for the expression of mature *Ll*MYB_SSM protein fused with GST-tag after have been directionally cloned in pET41a (+) vector as described in section 5.3.1 of this chapter. A pilot experiment with 5 mL culture of cells induced with 1mM IPTG was used to screen the clones showing optimum expression of the recombinant protein. For protein visualization, cells were harvested at 5 h after IPTG induction and isolated by centrifugation. The pelleted cells were resuspended in sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) loading buffer and heated at 95 °C. Proteins were separated by SDS-PAGE on 10% gel and stained with Coomassie blue. Molecular mass of the expressed protein was ~60 kDa (~26 kDa GST tag + ~29 kDa *Ll*MYB_SSM + ~ 5 kDa S-tag and 6-His tag) as confirmed by SDS-PAGE and consistent with the predicted sequence. The expression level of the recombinant *Ll*MYB_SSM protein was more or less same in all the screened clones. In uninduced bacterial clone the *Ll*MYB_SSM expression level was merely detectable (Fig. 5.11).

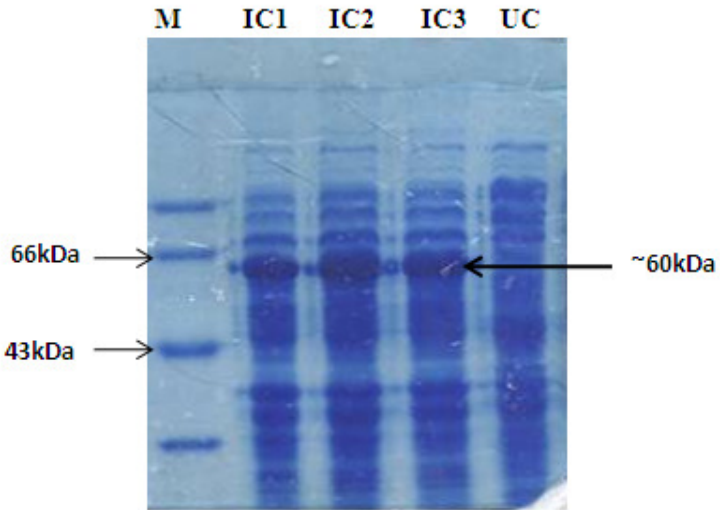


Fig. 5.11: Crude protein lysate from recombinant pET41a (+) *Ll*MYB_SSM clones run on 10% SDS gel; M: Medium Range Protein Molecular Weight Marker, Lanes: IC1, IC2 & IC3: Induced Crude lysate from different clones, showing expression of ~60kDa protein, UC: Uninduced Crude lysate from one of the clones.

5.3.5.1 Purification of recombinant protein in soluble form

Once protein expression was confirmed in pilot experiments, standardization of temperature and IPTG concentration was done to maximize the yield of protein in soluble fraction with 50 mL bacterial cultures. It was found that 14 hours of growth post-induction with 0.5 mM IPTG at 18 °C was enough to achieve the purification of protein in soluble form from the cell lysate (Fig. 5.12 A). Protein isolation from the lysate was done in native conditions to obtain the purified protein in active form. The protein was purified with GST-bind resin (Novagen) by affinity chromatography as described in chapter 2; section 2.9.11.4. Different eluted fractions of *Ll*MYB_SSM protein of ~60 kDa were separated under denaturing condition by 10% SDS-PAGE and visualized with silver staining (Fig. 5.12 B). In spite of using protease inhibitors and performing the purification protocol in cold room conditions (10 °C), the purified protein was associated with slight degraded proteins and impurities along with bacterial GST protein (~29 kDa). Although the yield of purified protein was not very good, it was enough for performing DNA-binding experiments.

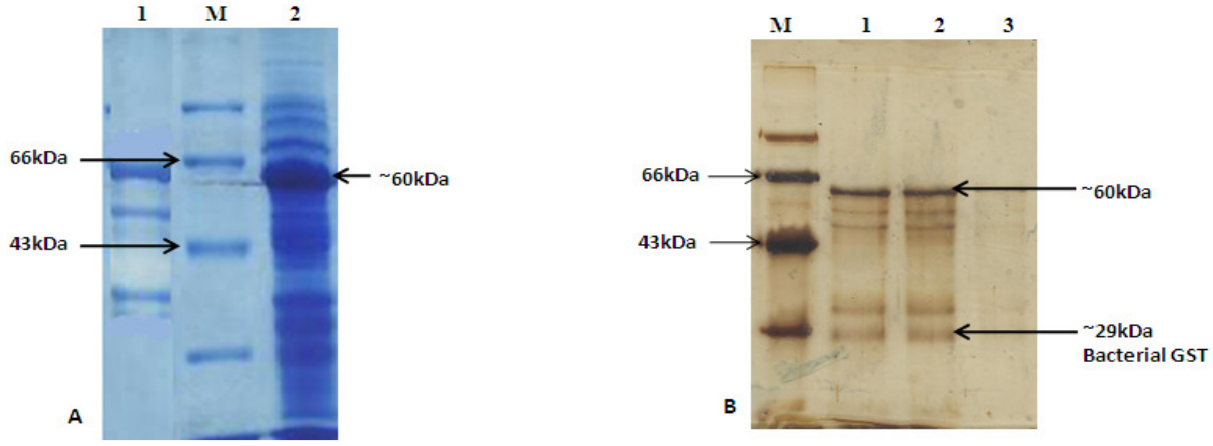


Fig. 5.12(A): Expression profile for one of the pET41 (a)-*Ll*MYB_SSM BL21 clones at 0.5 mM IPTG concentration grown at 18 °C for 14h: Lane 1: Inclusion bodies showing the insoluble fraction of total protein, Lane 2: Lysate showing the soluble fraction of total protein, the expected size of recombinant protein is ~60 kDa, as shown in the two lanes, and Lane M: MR protein marker.

Fig. 5.12 (B): Purification of over-expressed GST-*Ll*MYB_SSM protein from lysate using GST-bind resin: Lane M: MR protein marker, Lanes 1-3: Elution fractions of the purified protein in chronological order.

5.3.5.2 Gel-retardation assay of purified protein with AC-elements

GST-tagged MYB protein was earlier shown to retard AC-element containing promoter regions by Goichoeccha *et al*. (2005). Hence, the purified GST-*Ll*MYB_SSM protein was estimated using Bradford reagent (Bio-Rad, Marnes la Coquette, France) and directly used for EMSA studies with AC-element containing biotin-labeled promoter fragments of C4H, CCoAOMT, CCR and CAD (chapter 3, section 3.2.5). The binding reactions were performed in a total volume of 20 µL, with biotin-labeled promoter fragments (2 ng), ~1µg of purified GST-*Ll*MYB_SSM protein, 1 µg poly dIdC-poly dIdC (Thermo Scientific) and binding buffer (Thermo Scientific). The binding reactions were incubated for 30 min at room temperature as well as on ice and analyzed on a 6 % polyacrylamide gel as described in chapter 2, section 2.9.6 and the transferred blot was detected using chemiluminescence detection kit from Thermo Scientific as mentioned in chapter 2, section 2.9.7. However, no retardation was seen with any of the promoter fragments in the given conditions. The results show that either, *Ll*MYB_SSM protein does not directly interact with the AC-elements or it required some post-translational modifications (not provided by the bacterial host system) for binding to the elements. So, other methods for functional characterisation of the protein like molecular modeling was tried to gain an insight into the structural design of the protein 3-D structure.

5.3.6 Molecular modeling of *Ll*MYB_SSM using I-TASSER tool

Computational methods for predicting three-dimensional (3D) protein structures have been historically divided into three categories, based on the availability of template structures in the PDB library (Murzin, 2001): (1) comparative modeling, (2) fold recognition, and (3) *ab initio* methods. Among the several prediction techniques, the iterative threading assembly refinement (I-TASSER) server is an integrated platform for automated protein structure and function prediction based on the sequence-to structure-to-function paradigm (Roy *et al*., 2010). It is one of the composite approaches in protein structure prediction, which combine various techniques such as threading, *ab initio* modeling and atomic-level structure refinement approaches (Zhang, 2007, 2008; Das *et al.*, 2007, Zhou *et al*., 2007).

5.3.6.1 I-TASSER method for 3-D structure prediction

To determine the 3-D structure, protein sequence of *Ll*MYB_SSM was submitted in the I-TASSER webpage (http://zhanglab.ccmb.med.umich.edu/I-TASSER). Starting from an amino acid sequence I-TASSER first generates three-dimensional (3D) atomic models from multiple threading alignments and iterative structural assembly simulations. The various stages involved in the procedure were as follows:

1. Threading:

(i) The query protein sequence, *Ll*MYB_SSM was matched against a nonredundant sequence database by position-specific iterated BLAST (PSI-BLAST; (Altschul *et al*., 1997)). (ii) Based on multiple alignment of the sequence homologs, a sequence profile was created for prediction of secondary structure using PSIPRED (Jones, 1999). Secondary structure of the protein was predicted with a statistically significant Z-score of >1(Fig. 5.13). (iii) Based on the sequence profile and the predicted secondary structure, *Ll*MYB_SSM sequence was then threaded through a representative PDB structure library using a locally installed meta-threading server (LOMETS) and the threading programs (FUGUE, HHSEARCH, MUSTER, PROSPECT, PPA, SP3 and SPARKS). (iv) Based on the threading results, top template hit was then selected for further modelling. The quality of the template alignment was judged by statistically significant energy (*Z*-score).

```
                    20                  40                  60
                    |                   |                   |
Sequence    MGRAPCCEKAHTNKGAWTKEEDDRLVAYIRAHGEGCWRSLPKAAGLLRCGKSCRLRWINY
Prediction  CCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHCCCCHHHHHHHCCCCCCCCCHHHHHHHC
Conf.Score  998865455678468899899999999999978984999998717885650298877740
                    80                  100                 120
                    |                   |                   |
Sequence    LRPDLKRGNFTEEEDELIIKLHSLLGNKWSLIAGRLPGRTDNEMKNYWNTHIRRKLLSRG
Prediction  CCCCCCCCCCCHHHHHHHHHHHHHHCCHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHCC
Conf.Score  584554688988999999999872569999998788995478888899999999999758
                    140                 160                 180
                    |                   |                   |
Sequence    IDPTTHRPLNDASQNQSHASSASAAKQHDSEILITDTDANSKRPGVFEVCPDLNLELTIS
Prediction  CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
Conf.Score  997778887644457777777665545655533468766777887545689866344567
                    200                 220
                    |                   |
Sequence    LPHHHQHQQPDQISLKSINSESSTAGDGFDFLESKSSVLDCRSSVTQIDWRALTE
Prediction  CCCCCCCCCCCHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHCHHHHHCC
Conf.Score  998765457553121366888886656544456767653535861130287641 9
```

Fig. 5.13: Predicted secondary structure of *Ll*MYB_SSM. The prediction contains α-helix, coil structures are designated as H, C respectively with confidence scores for each residue (between 0 to 9)

2: Structural assembly: *Ab-initio* modeling

(i) Based on the threading alignments, aligned sections that aligned well was excised from the template structure. (ii) The unaligned regions were modeled by *ab-initio* modeling (Jauch *et al*., 2007).

3: Model selection and refinement

(i) Fragment assembly simulation was performed again starting from the selected cluster centroids. (ii) During this second round of simulations, the lowest energy structures were selected as input for REMO (Li and Zhang, 2009), which generates the final structural model through the optimization of hydrogen bonding networks.

5.3.6.2 Generation of the I-TASSER model

Among the top 10 templates used by I-TASSER as threading templates, the PDB hits, tabulated below (Fig. 5.14) showed the best normalized Z-scores with coverage of 45-47% with respect to the query protein (*Ll*MYB_SSM).

PDB hit	Molecule name	Length	Classification
2K9NA	NMR structure of the R2R3 DNA binding domain of Myb1 protein from protozoan parasite *Trichomonas vaginalis*	107	DNA binding protein
1A5J	Chicken b-Myb DNA binding domain, repeat 2 and repeat3, NMR	110	DNA binding protein

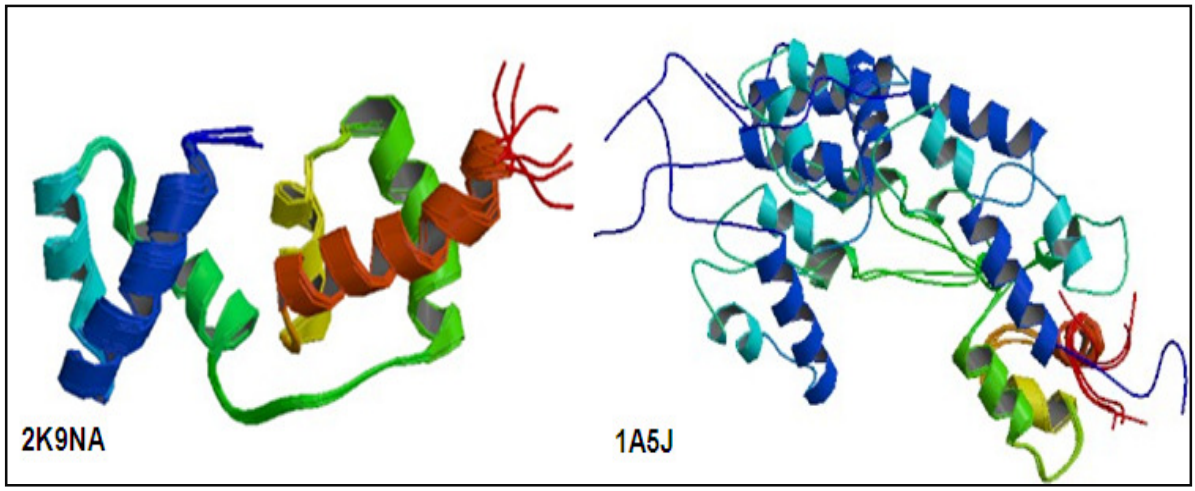


Fig. 5.14: PDB structures of the top two templates used by I-TASSER for model generation.

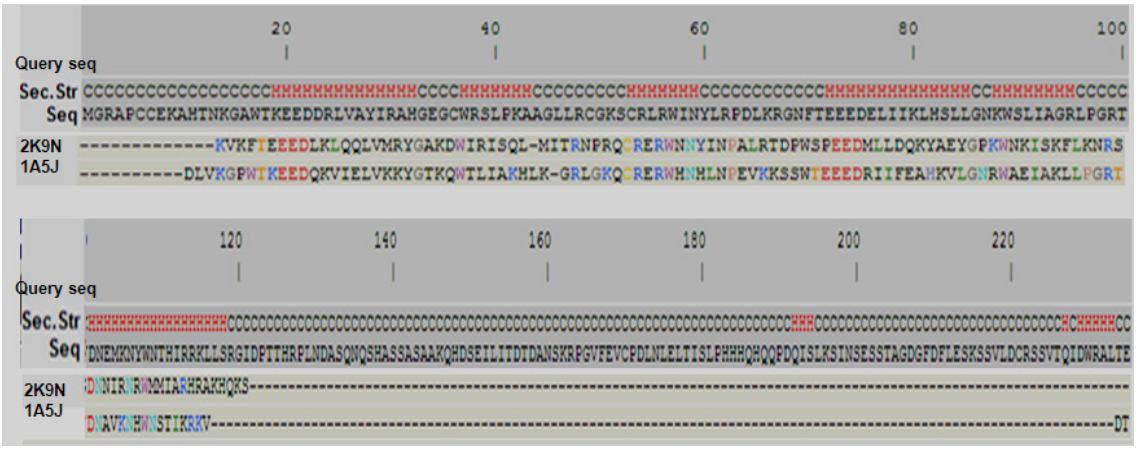


Fig. 5.15: Coverage of threading alignment of query protein (*Ll*MYB_SSM) with 2K9N and 1A5J. Residues in template which are identical to the residue in the query sequence are highlighted in colour.

The alignment of the query sequence along with its secondary structures against the two best threading templates 2K9N and 1A5J (Fig. 5.15), shows the best alignment in the region of R2 and R3 domains of the proteins. The identity is mostly restricted to the 6 helical regions highlighted in red in the secondary structure of *Ll*MYB_SSM.

5.3.6.3 Evaluation of the three dimensional model predicted for *Ll*MYB_SSM

In total five models were generated by I-TASSER server from which the top second model with a C-score: -3.803 (Fig. 5.16) was selected as the best after being subjected to internal evaluation of self-consistency checks. As described in I-TASSER, (Roy *et al*., 2010), C-score (confidence score) is an estimate of the quality of the predicted model and confidence of structure prediction, which is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations.

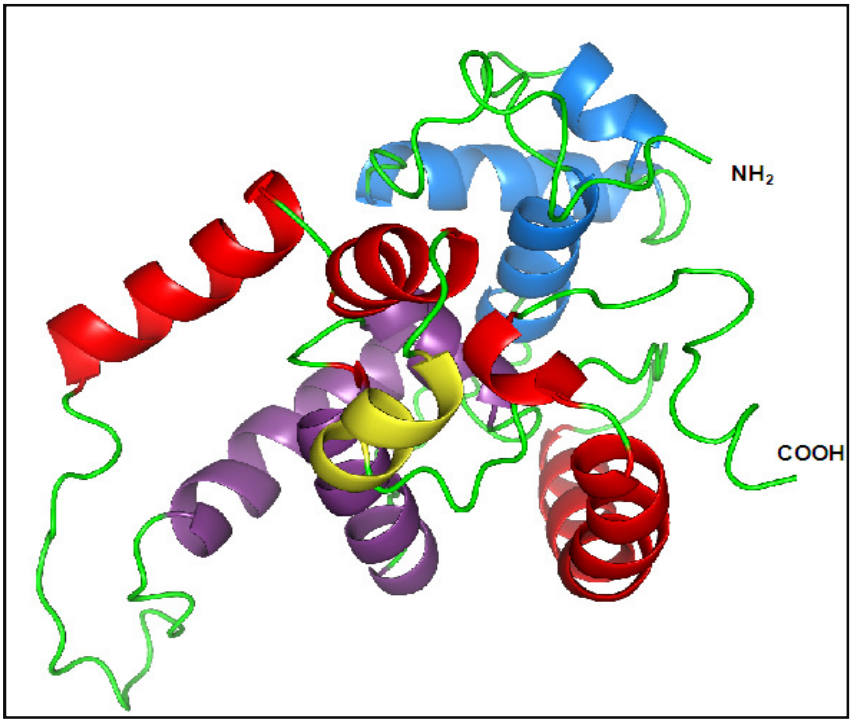


Fig. 5.16: Predicted model of *Li*MYB_SSM generated by I-TASSER and produced using Pymol software. The α-helices in R2 domain (H1, H2& H3) and R3 domain (H4, H5, & H6) are shown in marine blue and purple violet colours respectively; 9th α-helix containing the repressor motif (PDLNL) is shown in yellow colour; rest of the helices lying in the trans-activation C-terminal domain are shown in red colour and the loops are shown in green colour.

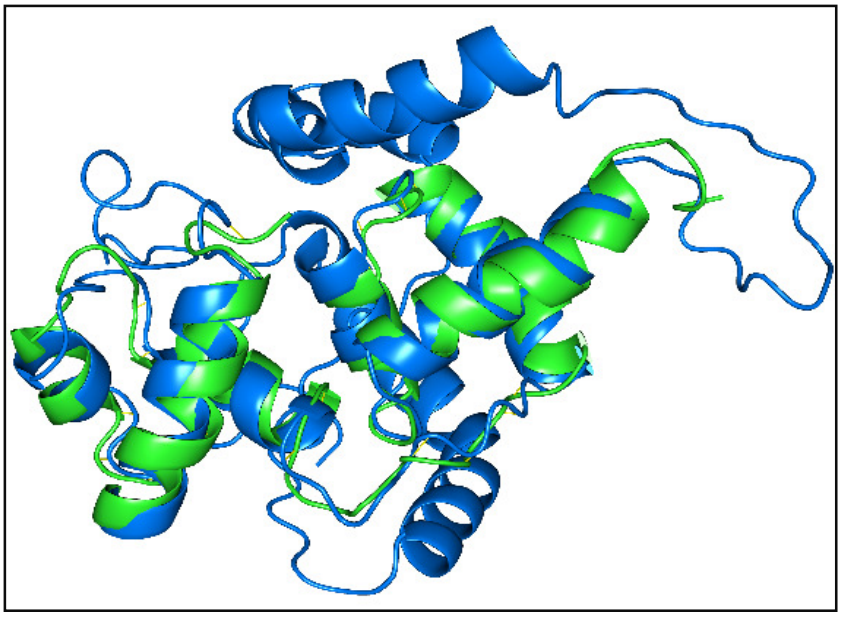


Fig. 5.17: Superposition of the I-TASSER generated *Li*MYB_SSM model with the structurally closest PDB hit: 2K9N identified by TM-align tool. The I-TASSER model is shown in marine blue colour and the NMR based R2 R3 domain of 2K9N model is shown superimposed in green colour.

TM-align tool predicts the TM score and the RMSD for the structurally aligned regions between the proteins, where TM score is the scale proposed to measure the structural similarity between two structures and RMSD is an average distance of all residue pairs in two structures. The TM-score for the model was predicted to be 0.4206 and RMSD of 1.66 with reference to 2K9N model. The TM-score is close to 0.5 which shows the fairly correct topology of the predicted structure. The RMSD score was also confirmed using Chimera tool (Pettersen *et al*., 2004) to be 1.185, which is close to 1.0 and hence shows the reliability of the model. The superimposed region (Fig. 5.17) is mostly seen in the DNA-binding domain of the two proteins. Due to lack of threading templates *ab initio* modeling was done for the non-aligned trans-activation domain.

5.3.6.4 Validation of model using PROCHECK

The stereo chemical quality of the two best scoring models was assessed by PROCHECK (Laskowski *et al*., 1993). The overall stereo chemical quality of the model was assessed by Ramachandran plot analysis, by quantifying the residues in available zones of the Ramachandran plot.

5.3.6.4.1 Ramchandran Plot

The PROCHECK analysis showed 83.3%, 15.3%, and 0.5% residues in most favorable, additionally allowed and generously allowed regions, respectively, of the Ramachandran map (Fig. 5.18). Only two residues were found to lie in the disallowed region. The overall structure assessed based on the plot characteristics, reveals a comparatively good and acceptable stereochemical quality of the predicted model. The Goodness factors (G-factors), from the PROCHECK results showed the quality of covalent and overall bond-angle distances. The overall average of PROCHECK G-factor was -0.23.The comparable Ramachandran plot characteristic and the goodness factors confirm the quality of the modeled structure.

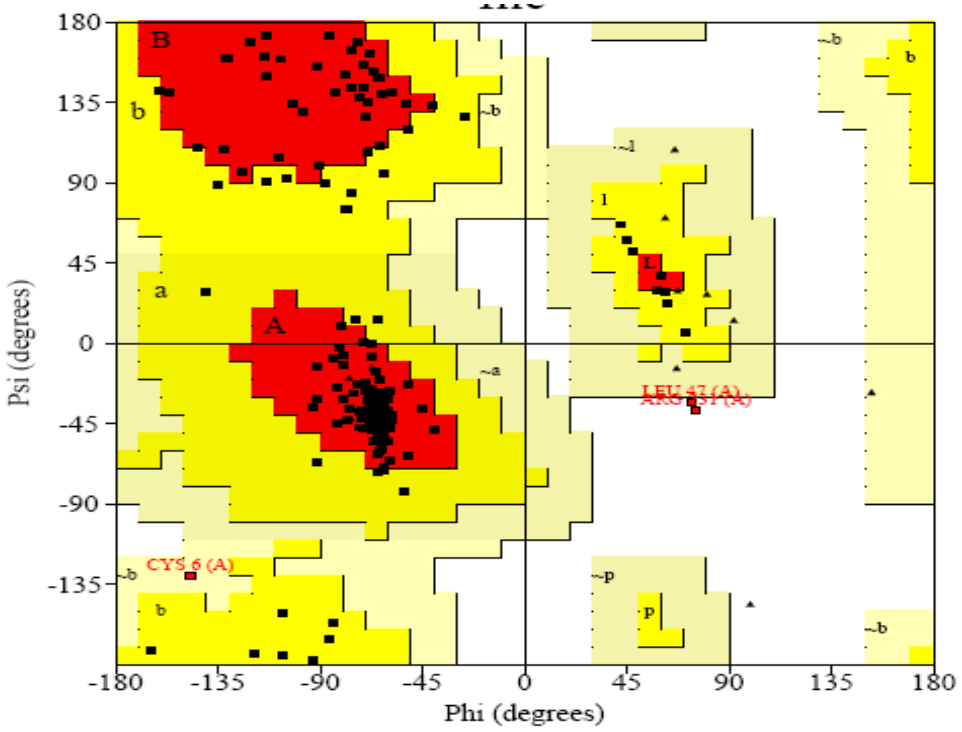


Fig. 5.18: Ramchandran plot statistics for model 1

Residues in most favoured regions [A,B,L]	**174**	**83.3%**
Residues in additional allowed regions [a,b,l,p]	**32**	**15.3%**
Residues in generously allowed regions [~a,~b,~l,~p]	**1**	**0.5%**
Residues in disallowed regions	**2**	**1.0%**
	----	------
Number of non-glycine and non-proline residues	**209**	**100.0%**
Number of end-residues (excl. Gly and Pro)	**2**	
Number of glycine residues (shown as triangles)	**14**	
Number of proline residues	**10**	

Total number of residues	**235**	

5.3.6.4.2 PROMOTIF Scan

The identified structural motifs and their locations in the predicted model were evaluated using PROMOTIF program and presented in Fig. 5.19.

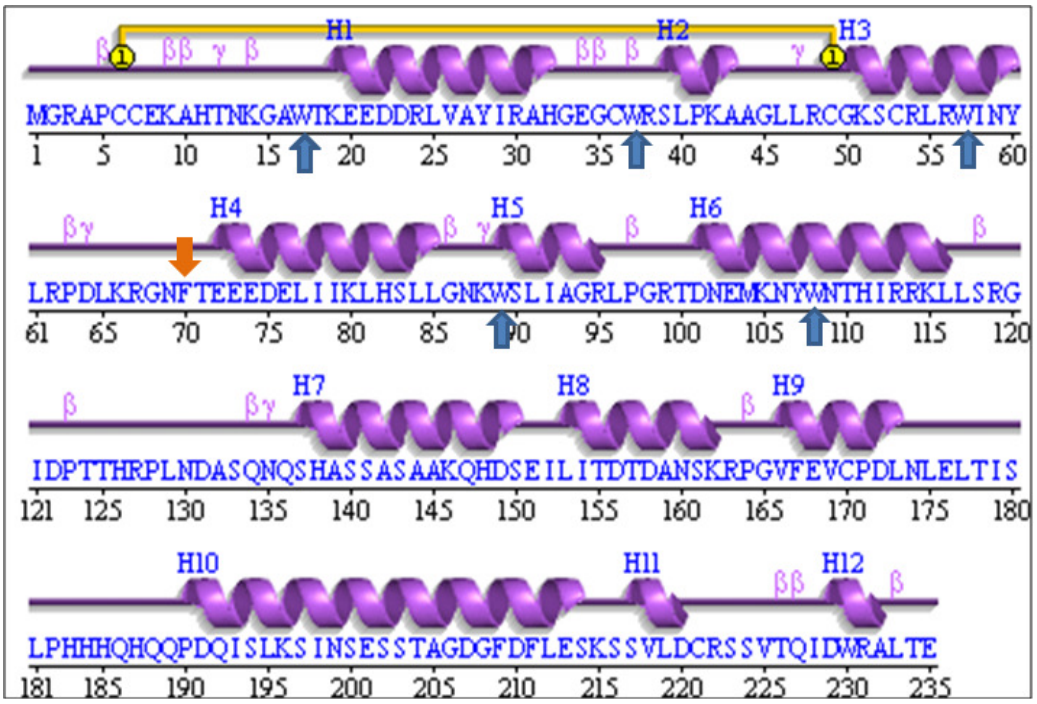


Fig. 5.19: Secondary structures predicted using promotif scan: Helices- 12; **Helix-helix interactions**-19; .**Beta turns (β)** -17; **Gamma turns (γ)** -5; **Disulphide bond is formed between** Cys6 and Cys49. The blue arrows indicate the Tryptophan residues (W) and the orange arrow indicates the Phenyl-alanine residue (F).

The *Ll*MYB_SSM structure contains 12 α-helices, 17 β-turns and 5 γ- turn motifs. The predicted protein structure contains two imperfect tandem repeats of 53 and 49 amino acids, designated as R2 and R3. As shown in Table 5.1, R2 and R3 repeat regions consists of three well defined α-helices and two less-well defined turns respectively forming a helix-turn-helix (HLH) structure (Fig. 5.19). Three regularly spaced (19 residues apart) tryptophan residues were present in R2 repeat and two regularly spaced (18 residues apart) tryptophan residues were present in R3 repeat (Fig. 5.19), which forms a tryptophan cluster in three-dimensional HLH structure (Table 5.1) and critical in sequence-specific binding. These structures are in accordance with the reported characteristics of typical plant R2R3 MYB proteins (Ogata *et al.*, 1992; Ying *et al.*, 2000). The third α-helix predicted in R2 and R3 repeats make contact with the DNA bases and thus determine the binding specificity (Ogata *et al*., 1992; 1994).

Table 5.1: Overall structural architecture of the predicted *Ll*MYB_SSM model

Structural component	Type of structural motif	Residue Locations
R2 repeat	α – Helix (H1)	Lys19 to His32
	β-Turn – Type IV	Gly33 to Ser39
	α – Helix (H2)	Ser39 to Ala43
	γ-turn	Leu46 to Arg48
	α – Helix (H3)	Gly50 to Tyr60
R3 repeat	α – Helix (H4)	Glu72 to Leu85
	β-Turn – Type IV	Leu85 to Lys88
	γ-turn	Asn87 to Trp89
	α – Helix (H5)	Trp89 to arg95
	β-Turn – Type I	Leu96 to Arg99
	α – Helix (H6)	Asp101 to Leu116
EAR repressor motif	α – Helix (H9)	Val166 to Leu173

5.3.6.5 Quality assessment of the model

5.3.6.5.1 Verify-3D

Verify-3D (Eisenberg *et al*., 1997) analyzes the compatibility of an atomic model (3D) with its own amino acid sequence (1D) based on score function. Each residue is assigned a structural class based on its location and environment (alpha, beta, loop, polar, nonpolar, etc). A collection of good structures is used as a reference to obtain a score for each of the 20 amino acids in this structural class. The Verify-3D profile for the final structure showed that 81.3 % residues had an average score over 0.2, thus the predicted *Ll*MYB_SSM model was considered as reliable.

5.3.6.5.2 ProSA analysis

The ProSA program (Protein Structure Analysis) (Wiederstein and Sippl, 2007), is an established tool which is frequently employed in the refinement and validation of experimental protein structures and in structure prediction and modeling. The program evaluates the energy of the structure using a distance-based pair potential. Residues with negative ProSA energies confirm the reliability of the predicted model. The z-score (-7.05) value displayed in plot is within the range of Z-scores of all experimentally determined protein chains in current PDB and indicates overall model quality (Fig. 5.20 A).

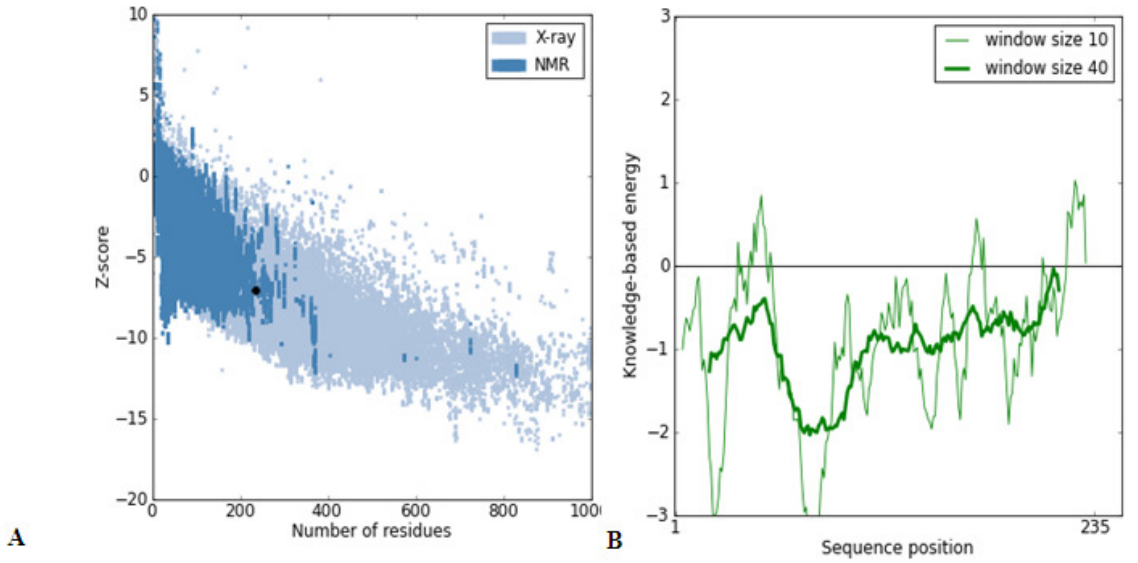


Fig. 5.20: ProSA analysis of the I-TASSER predicted model for *Ll*MYB_SSM. (A): ProSA-web z-scores of all protein chains in PDB determined by X-ray crystallography (light blue) or NMR spectroscopy (dark blue) with respect to their length. The z-score of the model (-7.05) is highlighted as a large dot. **(B): Energy plot of the model.** Residue energies averaged over a sliding window are plotted as a function of the central residue in the window (default size 40).

The energy plot shows the local model quality by plotting energies as a function of amino acid sequence position (Fig. 5.20 B). In general, positive values correspond to problematic or erroneous parts of a model. The plot shows that overall the residue energies are largely negative in window size 40 with the exception of some peaks in the N-and C-terminal parts in window size 10. The overall ProSA score obtained for the related modeled *Ll*MYB_SSM structure is close to the data base average and hence ensures the reliability of the model.

5.3.7 Heterologous expression of *Ll*MYB_SSM in *Nicotiana tobacum*

Redundancy in the R2R3 MYB family (having ~190 members in Poplar species) is the major hurdle in assessment of necessity of a MYB gene in a plant sytem like *Leucaena leucocephala*, which is a tree species. As postulated in the previous chapter (chapter 4) in Southern experiments, there is a possibilty of the presence of 3-4 functional orthologues of *Ll*MYB_SSM in the plant, so directed modification of the gene by either knock-out/knock-down experiments in a a homologous background was likely to prove difficult.

This is because functional redundancy of a gene limits the loss-of function experiments which requires suppression of the putative homologues to observe a phenotype. (Schwechheimer *et al*., 1998; Zhang, 2003). Hence, overexpression in a heterologous plant system is the preferred method, since it is less affected by functional redundancy.

The first line of genetic evidence on the possible involvement of MYBs in the regulation of lignin biosynthesis came from the study of two MYB proteins, AmMYB308 and AmMYB330 from *Antirrhinum*. Overexpression of the *Antirrhinum* MYB proteins in transgenic tobacco plants caused a reduction in the expression of several lignin biosynthetic genes and a decrease in lignin content, suggesting that the *Antirrhinum* MYBs are able to regulate the expression of lignin biosynthetic genes and thereby affect lignin biosynthesis (Tamagnone *et al*., 1998). Since then, several MYBs from *Arabidopsis* (Borewitz *et al.,* 2000), grapes (Deluc *et al*., 2006) and *Pinus* species (Patzlaff *et al*., 2003) have been shown to alter the expression of phenylpropanoid biosynthetic genes and lignin biosynthesis when overexpressed.

Tobacco was chosen as the heterologous plant system for transformation experiments as it provided an expression system that is taxonomically closest model angiosperm species to *Leucaena leucocephala*. However, previous reports have shown that the mechanisms underlying the transcriptional regulation of lignin biosynthesis are conserved between herbaceous species (like *Arabidopsis* and tobacco) and tree species.

5.3.7.1 Construction of plant transformation vector containing *Ll*MYB_SSM gene in sense orientation

The complete ORF of the gene from pGEM-*Ll*MYB_SSM clone was amplified with the introduction of *Kpn*I *and Bam*HI sites at 5' and 3' end of the gene respectively. The amplicon was cloned in pGEM-T easy vector. After digestion with *Kpn*I *and Bam*HI enzymes, it was subcloned in corresponding site of modified binary vector pCAMBIA1301 downstream to CaMV 35S promoter. The recombinant clone was referred as pCAMBIA1301-*Ll*MYB_SSM and was confirmed by restriction digestion which released a ~700 bp insert as depicted in Fig. 5.21. This construct, pCAMBIA1301-*Ll*MYB_SSM (Fig. 5.3) was used for transformation of *Leucaena* plants. The construct

was mobilized in *A. tumefaciens* strain (GV2260) by the freeze thaw method (chapter 2, section 2.9.2.4) for plant transformation.

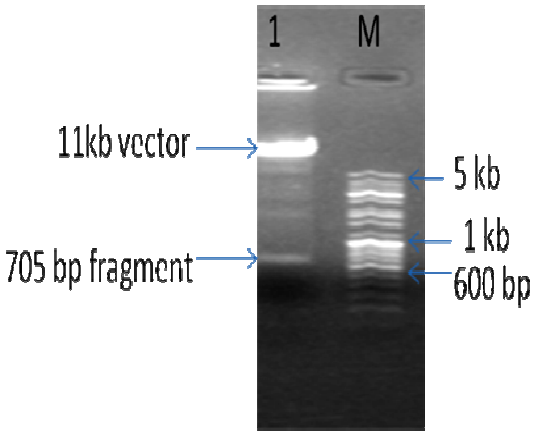


Fig. 5.21: Restriction digestion of recombinant pCAMBIA1301- *Li*MYB_SSM vector clone. Lane M: MR DNA ladder, Lane1: Recombinant plasmid releasing ~700bp insert from 11kb vector on digestion with *BamHI* and *Kpn*I.

5.3.7.2 Tobacco transformation

Tobacco was used as a standard transformation system. Transformation of tobacco was done according to the protocol given by Horsch, *et al*., 1985. The experiment started with preparation of 50 freshly prepared explants from axenic tobacco cultures. After co-cultivation the explants were incubated for 48 hrs in the dark. The explants were then washed in liquid MS media containing cefotaxime to remove *Agrobacterium*. They were then placed in SIM containing cefotaxime. After 1 week the explants were transferred to selection media (containing 5 mgL^{-1} hygromycin) where the transformed explants were selected. After 4-5 weeks of selection about 50% survival of putative transformed explants was obtained and small multiple shoots were seen emerging from the surviving explants. They were removed and further maintained on SIM (Shoot induction Medium) with cefotaxime for another 3-4 weeks to allow the proper growth of multiple shoots [Fig. 5.22 (A)-(E)].

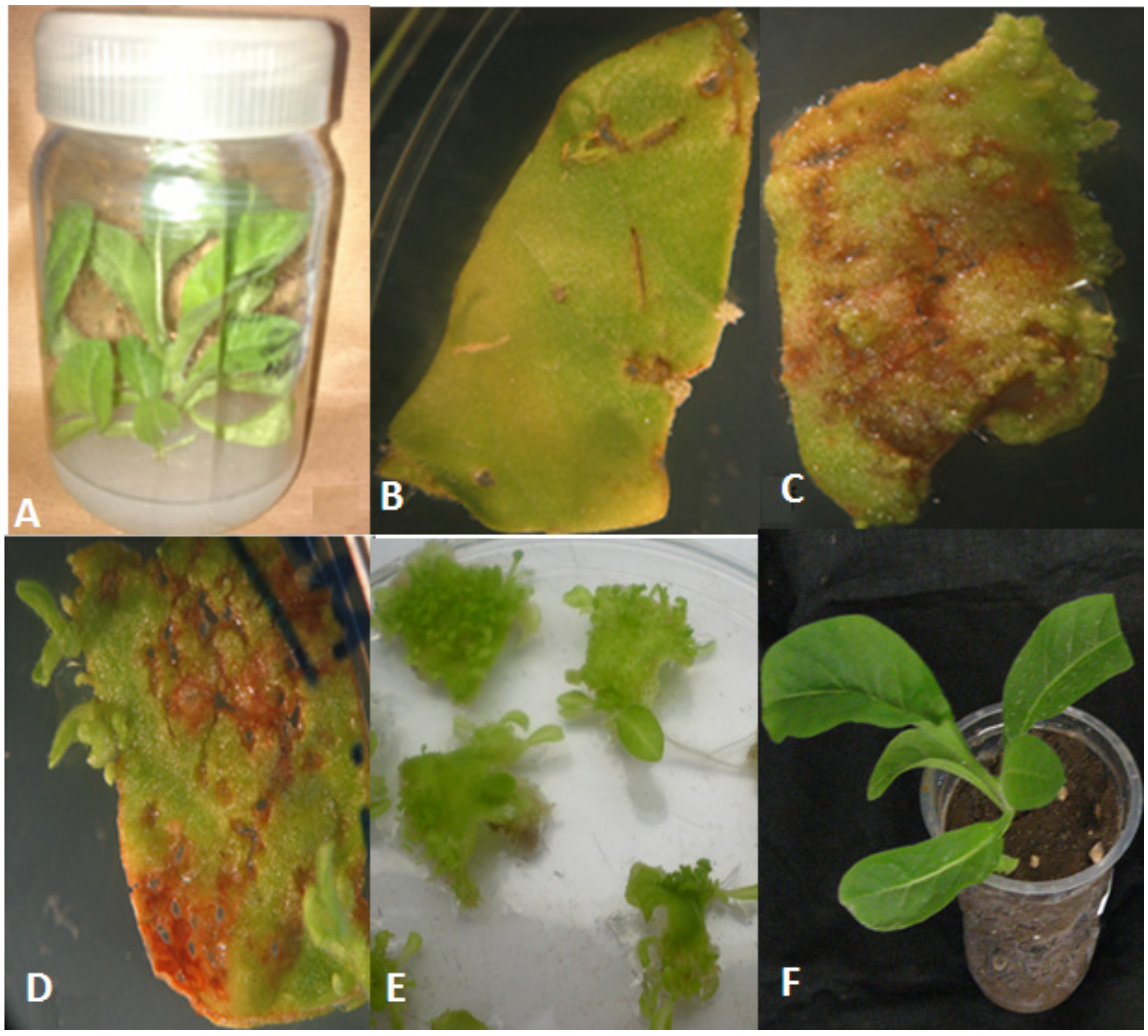


Fig. 5.22: (A) Axenic culture used for excision of explants (B) 10X magnified freshly prepared explant for cocultivation (C) 10X magnified image of a 4 weeks old explant placed on selection media showing the small emerging shoot buds at the edges and few in between (D) 10X image of a 6 weeks old explant showing multiple proliferating buds on the selection media, the non-transformed tissue is seen turning brown (E) 8 weeks old explants placed in shoot induction medium (F) Elongated and rooted shoots excised and placed in soil for proper growth before sending to green house.

Nearly 40 shoots in total were seen emerging from the surviving explants. The sufficiently elongated 22 individual shoots (putative transformants) were excised and shifted to RIM (Root Induction Medium) for further elongation and rooting. The other 9 explants carrying nearly 18 smaller multiple shoots were further maintained on SIM in bottles. Thus the average number of shoots emerging per explants was ~ 1.6. The percentage of putative transformants expected per explant was calculated to be

approximately 80%. However, the 18 smaller shoots did not survive due to contaminations observed and out of the 22 shoots kept for elongation and rooting only 8 showed proper growth and survived in the soil after being kept for hardening [Fig. 5.22(F)]. These rooted explants were used for PCR based confirmation for integration of hygromycin phosphotransferase gene (*hpt*II) and the positive plants were sent to green house for proper development and further analysis .No significant phenotypic difference was detectable at this stage in the putative transformants when compared to the control tobacco plants except for a little retarded growth of a few of them (Fig. 5.23).

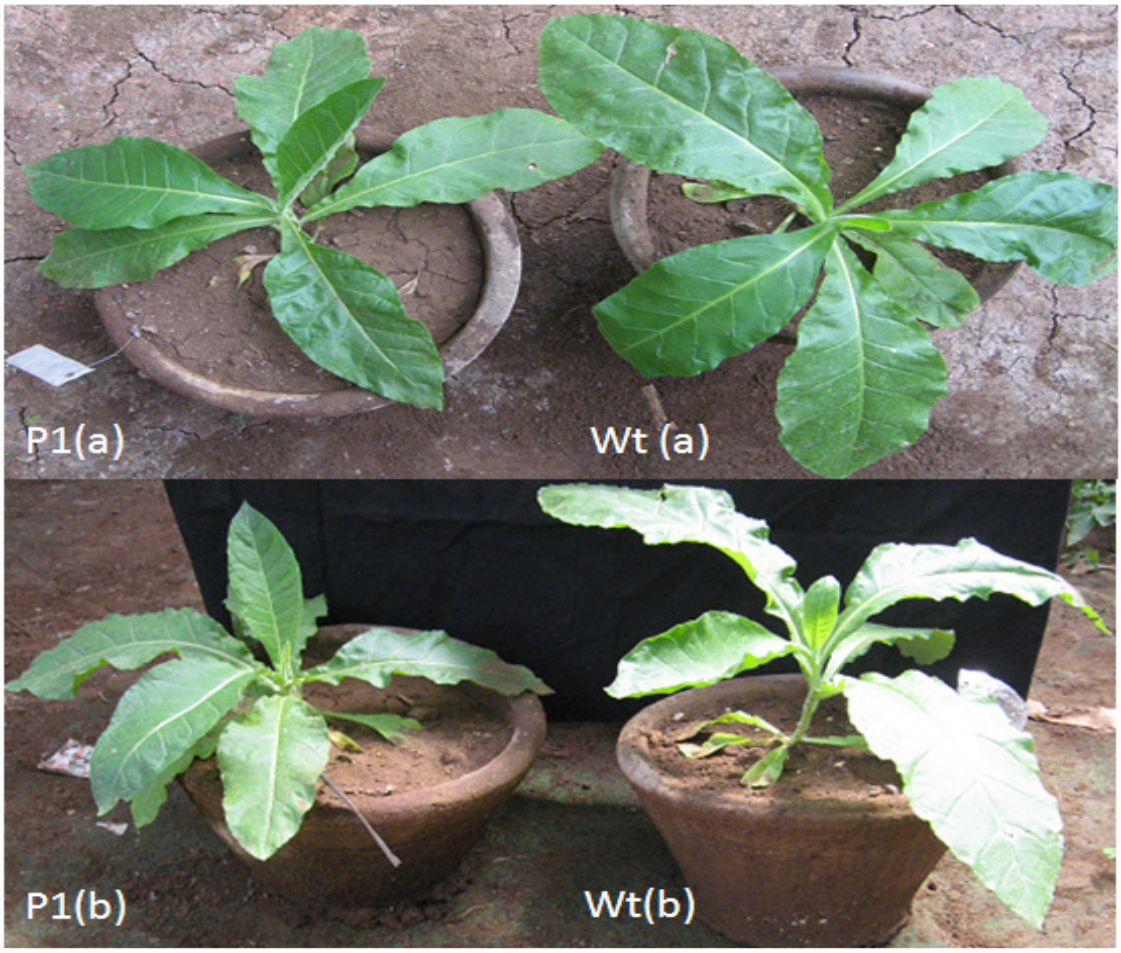


Fig. 5.23: P1(a) and P1(b): Aerial and vertical views of the putative transformant *Li*MYB_SSM-1 respectively, Wt(a) and Wt(b): Aerial and vertical views of the wild type or non-transformed control plant respectively.

5.3.7.3 Analysis of tobacco transformants

5.3.7.3.1 Checking the efficiency of transformation using GUS assay

To check the transient expression of *gus* gene, histochemical Gus assay was carried out (chapter 2, section 2.9.16) with 2 day old co-cultivated explants along with a control explant. Development of intense indigo blue colour scattered over the tissue was seen in co-cultivated explants, whereas no colour development was seen in the control explants (Fig.5.24). Thus the efficiency of *Agrobacterium* mediated transformation in tobacco explants was confirmed.

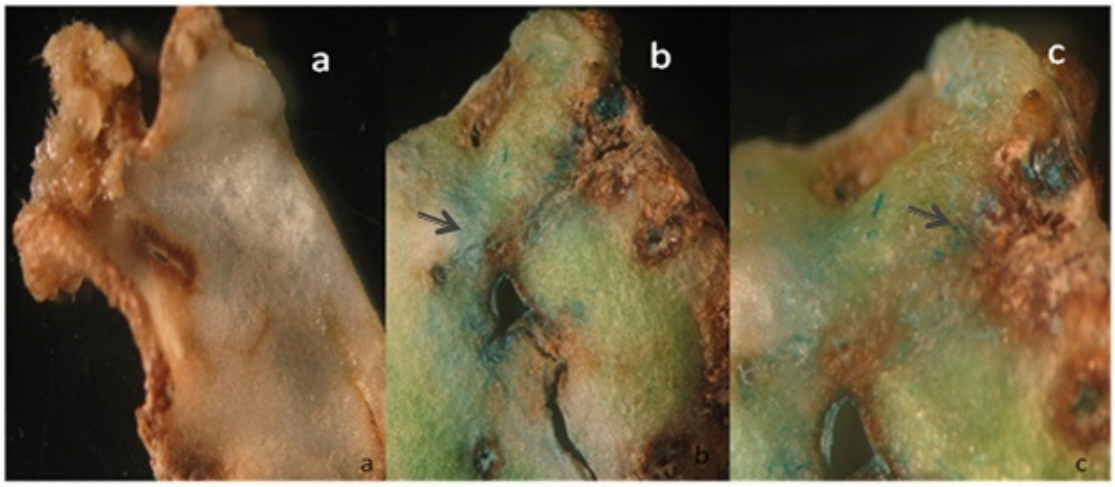


Fig. 5.24: Transient Gus assay of control and transformed tobacco explants after two days of co-cultivation. A) control explant, B and C transfected explants (arrows show the blue patches developed after GUS assay)

5.3.7.3.2 PCR based screening using hygromycin gene specific primers

Very few plants survived after transferring to rooting medium. PCR based screening for putative transformants were done for integration of hygromycin phosphotransferase gene (*hptII*) in the eight putative transformants. DNA was isolated from plants, which survived on hygromycin supplemented media. PCR was performed to amplify *hpt*II gene from putative transgenic lines using *hpt*II gene specific primers listed in chapter 2, section 3.3.7

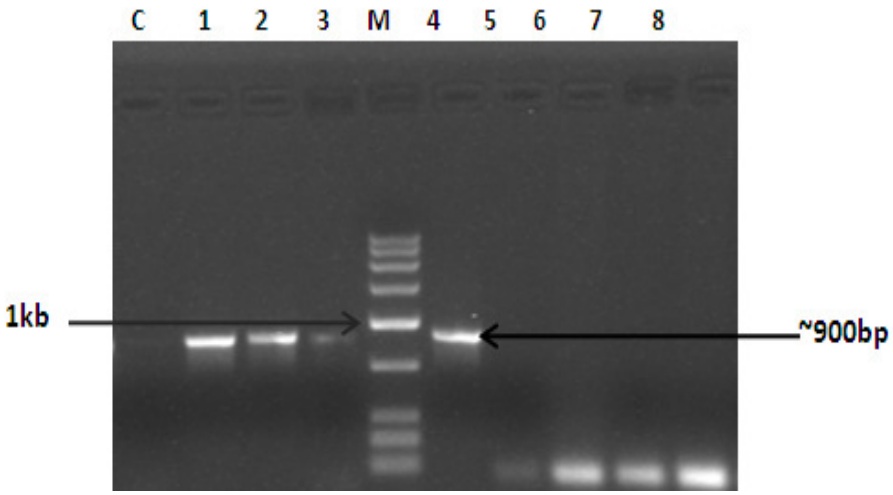


Fig. 5.25: PCR amplification of hygromycin resistance gene from putative transformed tobacco plants separated on 1% agarose gel: Lane C: Control/Wild type plant, showing no amplification, Lanes 1, 2, 3 & 4: Putative transgenics showing an amplification of ~900 bp of *hpt*II gene, Lanes 6, 7 and 8: Putative transgenics selected on hygromycin but not positive for *hpt*II gene.

An amplicon of ~900 bp was amplified from plausible transformed plants (Fig. 5.25). Four putative transformants of tobacco were characterized to be *hpt*II gene positive. After transfer to the glasshouse, *Ll*MYB_SSM plants still showed no visible differences in phenotypic aspect, except for slightly retarded growth relative to control plants. The putative transformants were analysed after two weeks of transfer to the green house.

5.3.7.3.3 Quantitative Real time PCR of putative tobacco transgenics

To gain an insight into the role of *Ll*MYB_SSM *in planta*, the four independent transgenic tobacco plants positive for *hptII* gene, expressing *Ll*MYB_SSM under the control of the 35S promoter, were selected for analysis. The ability of *Ll*MYB_SSM to alter the expression of a few genes involved in lignin biosynthesis (PAL, C4H, 4CL, CCoAOMT and CAD) was assessed by quantitative RT-PCR using RNA isolated from the sixth leaf of every putative transgenic and control plant described above. The tobacco specific primers were designed to hybridize to all genes of the class postulated to have a role in lignin biosynthesis, as most of them belong to multigene families. The assays were normalized to 18s gene transcript levels. The primer sequences used for the study are listed as follows:

Primer name	Primer sequence
***Leu*-18s-F**	5'-CATTCGAACGTCTGCCCTATCA-3'
***Leu*-18s-R**	5'-GATGTGGTAGCCGTTTCTCAGG-3'
PAL-Tob-F	5'-CGGTGGATTTTTCGAGTTGCAGCC-3'
PAL-Tob-R	5'-TGAGCCGCCTTCACATAAGAGCT-3'
C4H-Tob-F	5'-GTGGCAAGTGTAATTGAGGATGTG-3'
C4H-Tob-R	5'-GCCAATCTACTCCTTTCACCATTC-3'
4CL-Tob-F	5'-GGTTACACACTGGCGACATTGG-3'
4CL-Tob-R	5'-GGAACTTCTCCTGCTTGCTCATC-3'
CCoAOMT-Tob-F	5'-ACACCCTATGGAATGGATCAG-3'
CCoAOMT-Tob-R	5'-GCCTTGTTGAGTTCCAATACG-3'
CADTobF	5'-GGGAGTGAAAATAGCAAAGG-3'
CADTobR	5'-GCCAACAGGGACAGTATC-3'

PCR reaction was set up as mentioned in Chapter 2; section 2.9.8.2. Thermo-cycling conditions used for the qRT-PCR was as follows:

Temperature	Time	Cycles
95 °C	10 min	1
94 °C,	10 s	40
55 °C	45 s	
72 °C	30 s	
95 °C	1 min	1
55 °C	30s	
95 °C	30s	

A single fluorescent reading was taken at the end of each cycle as shown in the PCR conditions in the above table. Technical triplicates were performed for each sample to assess variation in QRT-PCR data. Difference between the transcript levels of the respective genes in control plant and the putative transformants was used to calculate fold change ratios. Fig. 5.26 shows the relative levels of transcript accumulation found in greenhouse-grown putative transformant plants relative to control plants. Among the four putative transgenics analyzed, two of them (*Ll*MYB_SSM-P1 and *Ll*MYB_SSM-P3) showed different levels of transcripts of *Ll*MYB_SSM gene, as well as a considerable change in the level of transcript accumulation for PAL, C4H, 4CL, CCoAOMT and CAD genes (Fig. 5.26). *Ll*MYB_SSM transcript level was higher in P3 (4 fold higher) as

compared to P1. However, negligible transcript level of *Ll*MYB_SSM was seen in the other two putative transformants (P2 and P4) and hence the data is not shown here.

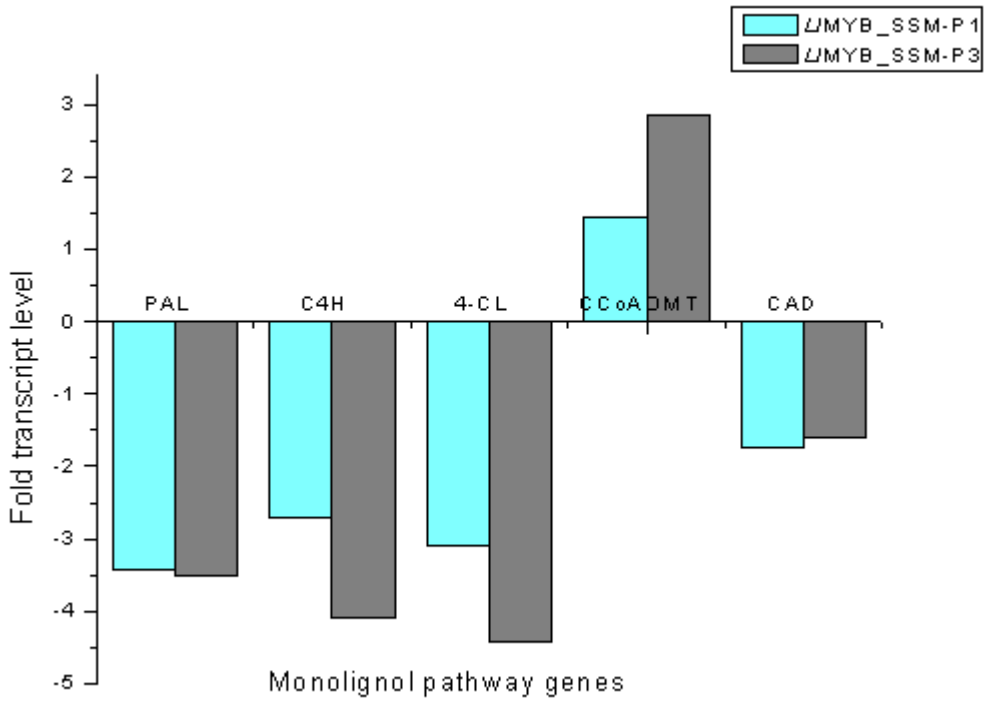


Fig. 5.26: Targeted QRT-PCR analysis in *Ll*MYB_SSM-P1 and *Ll*MYB_SSM-P3 putative transformants. Fold expression of the transcript levels of the monolignol biosynthetic pathway genes (PAL, C4H, 4-CL, CCoAOMT and CAD) have been calculated relative to their expression in control plants.

The level of transcript accumulation for the five genes analysed showed a definite correlation in the two transgenics, based on the relative levels of expression of the introduced *Ll*MYB_SSM gene as seen from the Fig. 5.26. The penetrance of the effect of *Ll*MYB_SSM transcript level was most pronounced for C4H, 4CL and CCoAOMT genes. The level of transcript accumulation for C4H and 4CL was significantly decreased (upto four fold) in *Ll*MYB_SSM-P3, which showed higher expression for *Ll*MYB_SSM as compared to *Ll*MYB_SSM-P1. On the other hand an upregulation of the transcript level was seen for CCoAOMT gene, which was increased in P3 (upto 2.5 fold) as compared to P1. Reduced transcript levels were seen for both PAL and CAD genes in the

two transformants with respect to control but it did not correlate with penetrance of *Ll*MYB_SSM expression level. PAL transcript level showed a reduction of upto 3.5 fold but in case of CAD it was only 1.5 fold. So, it can be inferred that CAD is least effected by overexpression of *Ll*MYB_SSM. The less number of transformants and the less biomass, limited further confirmation of the positive transformants either by ELISA or by slot blots.

5.4 Discussion

R2R3MYB transcription factors comprising the largest family of transcription factors in plants are shown to be involved in a wide spectrum of plant-specific processes. Most of them have been shown to be transcriptional activators as well as repressors of the phenylpropanoid metabolism in various plants like *Arabidopsis, Antirrhinum*, *Eucalyptus*, *Pinus* and *Poplar* species. Studies on the function of MYB transcription factors have been performed *in vitro* by heterologous expression of the protein in *E. coli* for EMSA and *in vivo* studies by yeast one-hybrid assays or by over expression studies in plant species like *Arabidopsis, Antirrhinum* and *Nicotiana tabacum L.*

*Ll*MYB_SSM gene isolated from *L. leucocephala* was cloned in pET 28 b (+) vector in fusion with N-terminal and C-terminal His tags and pET41a (+) vector in fusion with N-terminal His-tag, S-tag and GST-tag and expressed in *E. coli* strain BL21(DE3) cells for functional characterization of the protein. The 6X His affinity tag facilitated binding of protein in inclusion bodies to Ni-NTA agarose beads and thereby its purification in denaturing conditions. MALDI MS/MS analysis spectrum and coverage map of the purified protein confirmed the protein to be *Ll*MYB_SSM. Another best hit from the Uniprot database was of AmMYB308, which has been shown to be phylogenetically closest member of the protein as mentioned in chapter 4, section 4.3.8.4. The recombinant protein was then used for raising polyclonal antibodies in rabbit (New Zealand White). The antibody was used for further studies like Western, ELISA and immuno-cytolocalization. However, to obtain the protein in soluble and active form the GST-fusion protein was purified in mild conditions using GST-bind resin (Novagen). The partially degraded purified protein (~ 60 kDa) was used for binding studies with the MYBPLANT element present in the C4H, CCoAOMT, CCR and CAD promoters

isolated from *L .leucocephala*. But no retardation was observed in any case, which leads to the conclusion that the transcription factor activity might be dependent on post-translational modifications like phosphorylation, nitrosylation etc. not fulfilled by a bacterial host system. Another possibility was its ability to activate its target genes in close cooperation with DNA-binding proteins of other classes (Ness, 1999) like bHLH proteins.

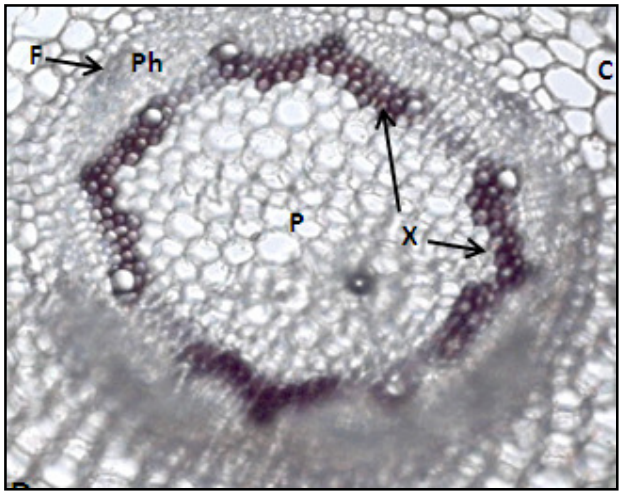
Overexpression of the protein was carried out in a genetically heterologous background best provided by *Nicotiana tabacum* for *in planta* studies of the gene and determining its role in regulation of the lignin biosynthetic pathway genes. The QRT-PCR studies on the putative transgenics for the expression of monolignol biosynthetic pathway genes: PAL, C4H, 4-CL, CCoAOMT and CAD genes was carried out. The results showed that all the genes except CCoAOMT were downregulated in the *Ll*MYB_SSM overexpressing transformants P1 and P3. Similar observations were made in case of orthologs of the gene, AmMYB308 and AtMYB4, where downregulation of C4H and 4-CL was common while CCoAOMT was upregulated only in case of AtMYB4 (Tamagnone *et* al., 1998; Jin *et al*., 2000). Although very little phenotypic difference of slight growth retardation was observed at the stage of analysis, the putative transgenics appeared to be genetically modified based on the changes in transcript levels of the studied genes.

Molecular modeling of the amino acid sequence of the protein was tried using I-TASSER tool to gain an insight into the structural architecture of *Ll*MYB_SSM. However, due to lack of homology in the trans-activation domain of the protein the generated model was best aligned in the DNA-binding region comprising the R2 and R3 domains. The structural features of the model were in agreement with the domain properties explained by Ogata *et al* (1992, 1994). The model was validated using PROCHECK, ProSA and Verify 3-D to be energetically stable. Future docking studies would further define the bases involved in binding to the DNA double helix and regions for protein-protein interactions.

5.5 Conclusions

The cDNA of *Ll*MYB_SSM isolated from *L. leucocephala* was directionally cloned in pET 28b+ expression system with two 6-His tags at N and C-termini. The recombinant protein was standardized for overexpression and purified from the inclusion bodies using Ni-NTA column. The purified protein of ~30 kDa was analysed using MALDI MS/MS for peptide mass fingerprinting data, which showed the best match with another R2R3MYB AmMYB308. The purified protein was then used to raise antibodies in New Zealand rabbit for future studies like immuno-cytolocalization and ELISA. As the protein could not be obtained in active form from the inclusion bodies it was expressed in pET41a+ vector in fusion with GST-tag and purified using GST-bind resin from the soluble fraction of the cell lysate. However, no binding of the protein was observed with the MYBPLANT element *in vitro*, so *in planta* overexpression of the gene was studied in tobacco. The study showed the changes in transcript levels of the few analysed lignin biosynthetic genes as discussed earlier in the overexpressing lines. The decrease in transcript levels of most of the genes validates the gene to be a negative regulator of lignin biosynthetic pathway, which is also supported by the presence of EAR repressor motif in the trans-activation domain of the protein discussed in chapter 4. I-TASSER modeling of the protein was useful in predicting the topology of the protein structure which was found to be similar to the reported R2R3 MYB domain structures in the Uniprot database.

CHAPTER 6



Spatio-Temporal expression Pattern of *LM*YB_SSM gene in *Leucaena leucocephala*

6.1 Introduction

Spatial and temporal control of lignin deposition relies on the antagonistic action of both repressors and activators to ensure the precise level of lignin accumulation in specific cells, tissues or organs. Although the roles of most genes of the monolignol biosynthesis pathway in determining lignin amount and composition have been elucidated, our knowledge is still scarce about their transcriptional regulation, which is mediated by lignin regulators.

Lignin deposition is also known to be activated in response to environmental cues like salt, wounding, pathogen attack, UV, jasmonic acid stress etc., for plant defense, which can also be correlated to the activation or suppression of specific transcription factors involved in lignification. Transcriptional repressors play a key role in controlling stress-related gene expression by acting as molecular brakes for slowing down the action of activators during the defense to prevent self-elicited permanent damage to the cellular components. Most of the lignin repressors known till date belong to subgroup4 of the R2R3MYB family (Zhao and Dixon, 2010). However, spatio-temporal expression studies have not been performed for any of these at gene / transcript level in the plants from which they have been derived.

The present chapter focuses on elucidating the involvement of a novel R2R3MYB transcription factor, *Ll*MYB_SSM from *L. leucocephala* in lignification by determining the level of its transcript accumulation in different tissues at developing stages of seedlings. The C-terminal region of the translated gene sequence is shown to have a conserved repressor motif (pdLNLD/ELxiG/S) having similarities to a characterized repressor motif called ERF-associated amphiphilic repression (EAR) motif, present in class II AP2/ERF transcriptional repressors and also in zinc finger proteins which function as repressors (Ohta *et al.*, 2001). Based on the implications of these motifs in controlling various stress responses (Kazan, 2006), known to influence lignification as well, *L. leucocepohala* seedlings were subjected to certain stress conditions to see the influence on the transcript level of *Ll*MYB_SSM. Although, the measurement of transcripts can be done with Northern blot analysis, it is not the most suitable experimental approach to semi-quantify RNA especially when very low concentrations

are expected. As, *Ll*MYB_SSM is a putative transcriptional repressor as shown in previous chapter, therefore abundance of its transcript level is expected to be low and therefore, real-time PCR has been used as the method of choice for studying its transcript level spatio-temporally. Moreover, real-time PCR has been used as a precise method of quantifying nucleic acids when exact quantification is sought, in conjunction with high sensitivity (Bustin, 2002, 2004). Anti-*Ll*MYB_SSM antibodies raised in rabbit (Chapter 5) has been used for studying the gene expression at protein level and for immuno-cytolocalisation studies with stem and root tissue sections in relation to lignin staining.

6.2: Materials and methods

6.2.1 Plant material

Seeds of *L. leucocephala* (K-636) were treated as described in chapter 2: section 2.1.1. The imbibed seeds were transferred to ½MS basal medium. The day of inoculation was considered as the zero day. Seedlings were harvested at 5, 10, 15 and 20 days respectively (Fig. 6.2). For 10, 15 and 20 day seedlings root, shoot and leaves were harvested separately and were used for Real-Time PCR analysis, ELISA and immuno-cytolocalization studies.

6.2.2 Chemicals used for stress induction

Methyl jasmonate - 95% (Sigma Aldrich); Sodium-chloride (HiMedia).

6.2.3 Primers and TaqMan® Probes

Gene specific forward and reverse primers and TaqMan probe were designed and synthesized by Eurogentec for QRT-PCR experiments.

6.2.4 Mastermix for QRT-PCR

Brilliant QPCR master mix (Stratagene) with low ROX (Reference dye) was used for qPCR experiments.

6.2.5 Primary antibodies for anti-LlMYB_SSM protein

Purified *Ll*MYB_SSM protein was used to raise antibodies in rabbit as discussed in chapter 5, section 5.2.2.12.

6.2.6 Secondary antibodies

Alkaline phosphatase tagged goat anti- rabbit IgG antibodies were purchased from Merck.

6.2.7 Methods

6.2.7.1 Total RNA extraction and cDNA synthesis

Total RNA was isolated from experimental germinated seedlings according to the TRIzol reagent (Sigma, Germany) or hot phenol method at different time intervals. First strand cDNA synthesis, primed with an oligo (dT)15 primer, was performed with Avian Myeloblastosis Virus Reverse Transcriptase (AMV-RT) according to the manufacturer's protocol (Promega Corp., Madison, USA).

6.2.7.2 Quantitative Real Time PCR (QRT-PCR)

Quantitative Real Time PCR is a powerful tool for gene expression analysis and was first demonstrated by Higuchi *et al.* (1992, 1993). QRT-PCR quantifies the initial amount of the template most specifically and is a preferable alternative to other forms of quantitative reverse transcriptase PCR that detect the amount of final amplified product at the end point (Freeman *et al.*, 1999). QRT-PCR monitors the fluorescence emitted during the reaction as an indicator of amplicon production during each PCR cycle (i.e., in real time) as opposed to the endpoint detection. In the present study the Brilliant® II QPCR Master Mix (Stratagene, USA) with Low ROX was used. This kit supports quantitative amplification and detection systems. The kit supports PCR amplifications and detection of a variety of DNA targets, including genomic DNA, plasmid DNA, and cDNA. The Brilliant II QPCR master mix includes SureStart® Taq DNA polymerase, a modified version of Taq2000™ DNA polymerase with hot start capability. A passive reference dye

(ROX) is provided so as to control the final dye concentration, increasing the flexibility of the reagents for use with different platforms.

6.2.7.3 TaqMan® Probes (Hydrolysis Probes)

TaqMan probes are linear molecules with a fluorophore usually at the 5´ end of the probe, and the quencher is at the 3´ end. As long as the probe is intact, regardless of whether it is hybridized with the target or free in solution, no fluorescence is observed from the fluorophore. During the combined annealing-extension step of PCR, the primers and the TaqMan probe hybridize with the target. The DNA polymerase displaces the TaqMan probe by 3 or 4 nucleotides, and the 5´-nuclease activity of the DNA polymerase separates the fluorophore from the quencher. Fluorescence can be detected during each PCR cycle, and fluorescence accumulates during the course of PCR. The present study uses 5'-6-FAM dye as the fluorophore and 3'-BHQ-1 as the quencher in the TaqMan probe.

6.2.7.4 Q-PCR considerations: As discussed in chapter 2, section 2.9.8.1

6.2.7.5 Preparing the QRT-PCR reactions: As discussed in chapter 2, section 2.9.8.2

6.2.7.6 Quantitative Real-time PCR for *Ll*MYB_SSM gene

Total RNA was extracted individually from leaves, roots and shoots at different time intervals from developing seedling of *L. leucocephala.* One µg of total RNA was used for making cDNA using ImProm cDNA synthesis kit (Promega, Madison, USA). Brilliant QPCR kit (Stratagene, USA) and Stratagene Mx3000P real time machine were used for all reactions. The primer sequences that were designed for *Leucaena Ll*MYB_SSM gene and 18s rRNA are listed below:

Primer/ Probe	**Sequence**	**Tm**
Leu-18s-F	5'-CATTCGAACGTCTGCCCTATCA-3'	**60.3 ^{o}C**
Leu-18s-R	5'-GATGTGGTAGCCGTTTCTCAGG-3'	**62.1 ^{o}C**
Leu-18s-probe	5'-*FAM*-ATTCTCCGTCACCCGTCACCACCAT-*BHQ*-3'	**55.9 ^{o}C**

SSM-MYB-F	5'- GCCTCCCAACGATGCTTCTCAG -3'	**51.6 ^{o}C**
SSM-MYB-R	5'- GTCCGTAATCAGAATCTCCGAGTC- 3'	**52.3 ^{o}C**
SSM-MYB probe	5'-*FAM*-TCATGCCTCATCTGCTTCTGCTGCT-*BHQ*-3'	**54.2 ^{o}C**

Optimal numbers of PCR cycles within the linear range of amplification for each gene were determined in preliminary experiments. RT-PCR reactions were performed following conditions mentioned in Chapter 2, section 2.9.8.2 with annealing temperature of 55 ^{o}C. The reaction was run in triplicates and repeated twice. It was ensured that equal quantity of RNA template was used for each reaction (Pfaffl, 2001; Freeman *et al*., 1999; Edyta Zdunek-Zastocka, 2008).

6.2.7.7 Salt, jasmonic acid and UV stress

L. leucocephala seedlings, which were of 15 day old age after germination were subjected to salt and jasmonic acid stress, by dipping the roots of the seedling completely in MS liquid medium containing 100 mM NaCl and 100 µM methyl-jasmonate respectively in standard tissue culture conditions. After 24 h the leaf, stem and root tissues were harvested for RNA isolation.

UV-stress was given to the 15day old seedlings growing in MS semi-solid medium by keeping the bottles with mouth opened in laminar airflow cabinets and exposed to the germicidal UV-C radiation (280 nm-100 nm) produced by medium-pressure lamps for 15 min. After exposure the seedlings were further grown for 24 h under standard tissue culture conditions to allow the accumulation of UV-responsive transcripts in the plant tissue and harvested for RNA extraction.

6.2.7.8 Relative quantification method

Relative (comparative) quantification relates the PCR signal of the target transcript in a treatment group to that of another such as untreated control. The comparative Ct method is also known as the $2^{-\Delta\Delta Ct}$ method, where:

$\Delta\Delta$ Ct = Δ Ct $_{\text{calibrator}}$ - Δ Ct $_{\text{treated sample}}$

Here, Δ Ct $_{calibrator}$ is the Ct value for any untreated sample normalized to the endogenous housekeeping gene and Δ Ct $_{treated\ sample}$ is the Ct value for the treated sample also normalized to the endogenous housekeeping gene. Means,

Δ Ct $_{calibrator}$ = Ct $_{untreated\ sample}$ − Ct $_{endogenous\ control}$ (18s rRNA gene or any other housekeeping gene)

Δ Ct $_{treated\ sample}$= Δ Ct $_{treated\ sample}$ − Ct $_{endogenous\ control}$

For the $\Delta\Delta$Ct calculation to be valid, the amplification efficiencies of the target and the endogenous reference must be approximately equal. This can be established by looking at how ΔCt varies with template dilution. If the plot of cDNA dilution versus delta Ct is close to zero, it implies that the efficiencies of the target and housekeeping genes are very similar. If a housekeeping gene cannot be found whose amplification efficiency is similar to the target, then the standard curve method is preferred

6.2.7.9 Western blotting using anti-*Ll*MYB_SSM antibodies

Polyclonal antibodies raised against *Ll*MYB_SSM in New Zealand rabbit as mentioned in previous chapter was confirmed for its specificity for the protein and if any cross-reaction with another MYB protein from *L. leucocephala* (*Ll*MYB_ATS, GenBank: GU901208) by using its 1:10000 times dilution.

iBlot Gel Transfer System was used to transfer the protein from SDS-PAGE gels on PVDF membrane by strictly following manufacturer's (Invitrogen, USA) instructions. The PVDF membrane containing transferred protein samples were processed as per standard procedure, blocking, treatment with primary antibody (anti- *Ll*MYB_SSM antibody), washing followed by treatment with secondary antibody which is ALP conjugated, followed by washing and colour development using BCIP/NBT as substrate. The reagents used for the analysis were same as those used for ELISA experiments as described in chapter 2, section 2.8.8.

6.2.7.10 Nuclear protein extraction from mature plant tissues

Nuclear protein was isolated from matured plant leaf and stem tissues of *L.leucocephala* as per the protocol described in chapter2, section 2.9.5.

6.2.5.11 ELISA for *Ll*MYB_SSM protein in nuclear protein extract

The protocol was followed as mentioned in chapter 2, section 2.9.14.2.

6.2.7.12 Immuno-cytolocalization of *Ll*MYB_SSM protein in *Leucaena leucocephala*

The The solutions and reagents for immuno-cytolocalisation were used as listed in chapter 2, section 2.8.8, and the protocol was followed as mentioned in chapter 2, section 2.9.12.

6.2.7.13 Histochemical staining of lignin

Free hand transverse sections were prepared for 10 day old stem and root tissues of *L.leucocephala* as for immuno-cytolocalisation experiments and phloroglucinol staining was done as described in chapter 2, section 2.9.13. The slides were viewed under a light microscope and pictures captured (Axioplan 2, Zeiss).

6.3 Results

6.3.1 Designing *L. leucocephala* specific 18s probe

Analysis and comparison of Real-Time qPCR assays is susceptible to various uncontrolled variables like the amount of starting material, enzymatic efficiencies and differences between tissues. Normalization of the data to an endogenous housekeeping gene is the most commonly used method for overcoming these drawbacks. However, the reference gene has to be chosen so that its expression does not change under the experimental conditions or between different tissues (Cook NL *et al*., 2009). Various control kits are commercially available for commonly used reference genes like: 18s rRNA, 28s rRNA, Beta actin, Ubiquitin C etc., which are known to show constant level of expression between individuals, among different tissues of an organism and at all stages of development. Most of the kits have been standardized for animal systems and hence using them for plant system requires proper validation of results, as there are differences at genetic level between plants and animals.

Another important factor in selecting a reference gene is that its expression level should be comparable to the gene of interest. Keeping this in mind a few pairs of primers for different internal controls were tested for their expression levels in different *L. leucocephala* tissues like leaf, root and stem at different developmental stages. The results showed that 18s rRNA had the most stable expression level, comparable to the expression level of *Ll*MYB_SSM in the plant. As, the 18s rRNA sequence for *L. leucocephala* was not known, the 18s rRNA control kit, having Yakima Yellow®-Eclipse® Dark Quencher probe (Eurogentec) was used for normalization of data. No signal was detected for *L. leucocephala* cDNA with the probe in the course of reaction by the instrument. However, the probe showed signals when used with the Human Xsomal DNA provided in the kit as a positive control. As the gel analysis of both the reaction products showed an amplification of the desired fragment (121 bp), the fragment was cloned and sequenced to confirm if there was a difference in the sequence in the probe annealing region in *L. leucocephala* genome. The retrieved sequences were aligned to find out the regions of dissimilarities (Fig. 6.1).

```
18SHu           CATTCGAACGTCTGCCCTATCAACTTTCGATGGTAGTCGCCGTGCCTACCATGGTGACCA
18sLeu          CATTCGAACGTCTGCCCTATCAACTTTCGATGGTAGGATAGTGGCCTACCATGGTGGTGA
                ************************************       *************   *

18SHu           CGGGTGACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACAT
18sLeu          CGGGTGACGGAGAATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACAT
                ********** **** ****************************************

18SHu           C
18sLeu          C
                *
```

Fig. 6.1: Probe designing region identified for *Leucaena* 18s rRNA gene compared to the *Homo sapiens* 18s gene rRNA gene. The highlighted portion shows the *Leu*-18S-probe designing region.

As evident from the alignment results, the probe annealing region highlighted in the above figure was different and not the primer annealing regions. As the amplicon of expected size was obtained, but the probe did not hybridize to the *L. leucocephala* 18s rRNA sequence, no signal was detected in QRT-PCR. BLAST results also showed that

the 18s rRNA gene reported for other plants shared 100% identity with the *L. leucocephala* sequence. This leads to the conclusion that the control kits standardized for animal systems do not necessarily work for plant systems. Based on the sequence an 18s rRNA probe specific to *L. leucocephala* was designed (sequence mentioned in section 6.2.7.6 of this chapter) and used for normalizing the QRT-PCR data.

6.3.2 Spatio-Temporal expression pattern of *Ll*MYB_SSM gene using Real-time PCR

Spatio-temporal expression of *Ll*MYB_SSM was carried out by isolating total RNA from leaf, stem and root tissues of 5, 10, 15 and 20 day old developing seedlings of *L. leucocephala* (Fig. 6.2). 30 day and 60 day old seedlings were also studied for expression in leaf and stem tissues. TaqMan probe in combination with gene specific primers designed from the most non-conserved C-terminal region of the gene was used for the QRT-PCR experiments. The primer and probe sequences are listed in section 6.2.7.6 of this chapter. The cDNA was normalized using the *Leu*-18s primers and *Leu*-18s-probe designed in section 6.3.1. The PCR cycling conditions were optimized for both the control (18S rRNA) and target (*Ll*MYB_SSM) genes. Mx3000P (Stratagene, USA) was used for Real time quantitative PCR studies.

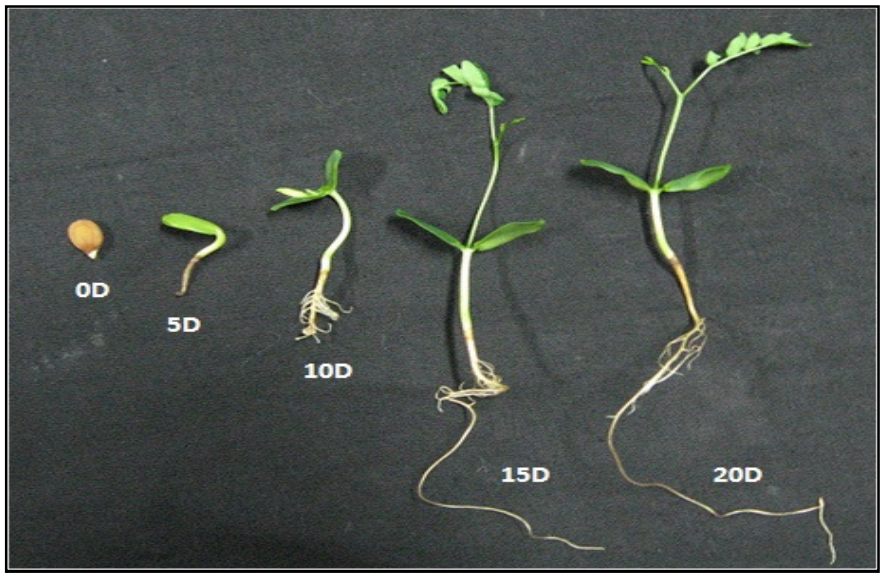


Fig. 6.2: Different stages of *L. leucocephala* seedling development

The mean Ct value was calculated from three technical replicates along with standard deviation of each sample (target as well as internal control gene) run individually. Based on the Ct values, the expression patterns in different tissue types of different age groups were analyzed.

6.3.2.1 Temporal expression pattern in leaf tissue for *Ll*MYB_SSM gene

Fold expression was calculated for the gene using cDNA prepared from leaf tissues of 10, 15, 20, 30 and 60 day old seedlings, with respect to the Ct value for 15 day old leaf showing the lowest expression.

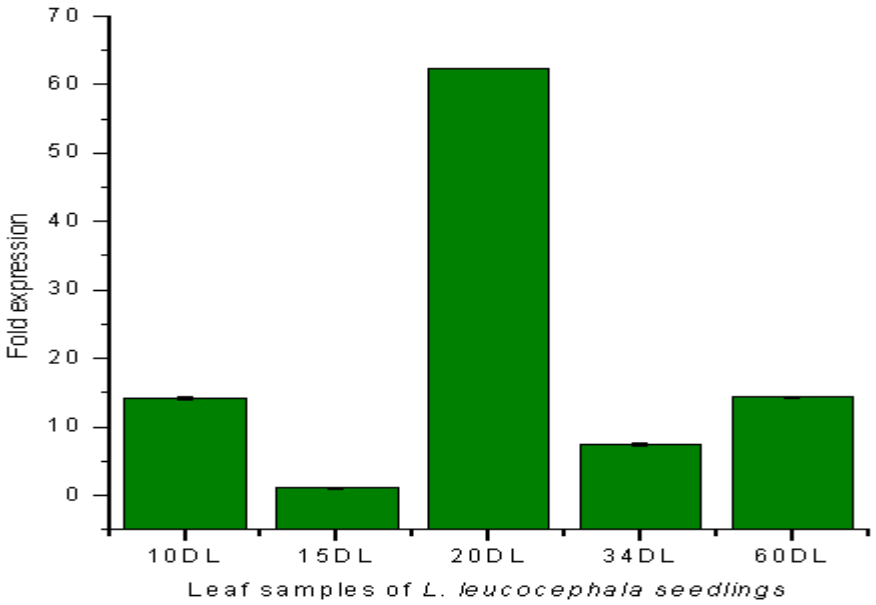


Fig. 6.3: Fold expression of *Ll*MYB_SSM transcript level in leaf tissue at different times

As seen from Fig. 6.3, there is an irregular pattern of the level of *Ll*MYB_SSM transcript level in the leaf tissue at different developmental stages of the *L. leucocephala* seedlings. The expression was observed to be highest i.e. almost 62.26 fold of the 15 day old leaf tissue in 20 day leaf tissue. However, a decrease in expression was seen with the matured leaf tissues of 30 and 60 day.

6.3.2.2 Temporal expression in root tissue for *Ll*MYB_SSM gene

Fold expression was calculated for the gene using cDNA prepared from root tissues of 0, 5, 10, 15 and 20 day old seedlings, with respect to the Ct value for 0 day old root (radicle) showing the lowest expression.

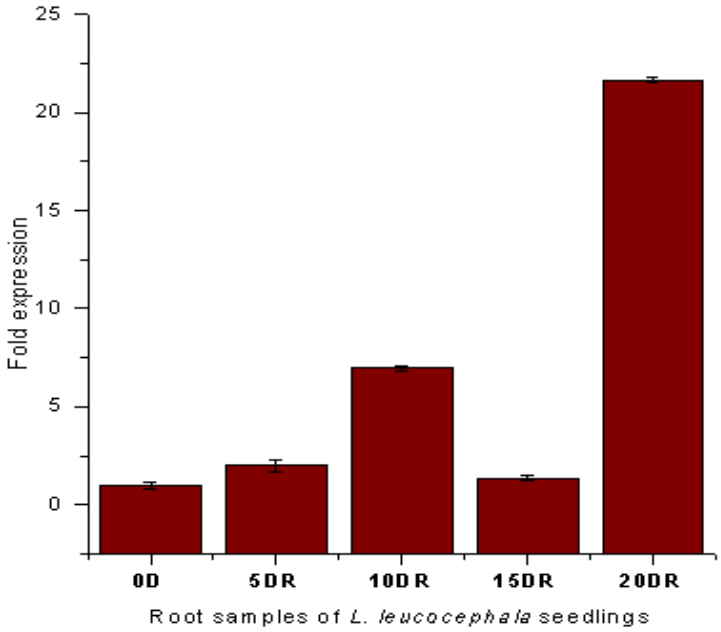


Fig. 6.4: Fold expression of *Ll*MYB_SSM transcript level in root tissue at different times

Expression level of the gene was highest in 20 day root (21.65 fold) with respect to 0 day seedling stage (Fig. 6.4). Again a decrease in the transcript level was observed in 15 day root with respect to 10 day root, as observed in case of leaf tissue. However, 30 and 60 day root tissues could not be analysed.

6.3.2.3 Temporal expression in stem tissue for *Ll*MYB_SSM gene

Fold expression was calculated for the gene using cDNA prepared from stem tissues of 5, 10, 15 20, 30 and 60 day old seedlings, with respect to the Ct value for 60 day old stem showing the lowest expression.

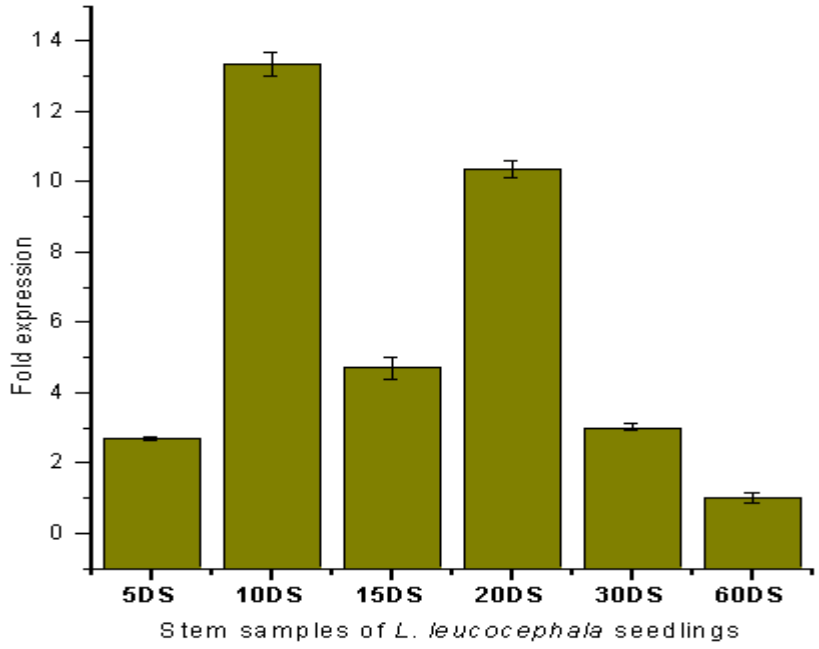


Fig. 6.5: Fold expression of *Ll*MYB_SSM transcript level in stem tissue at different times

The pattern of *Ll*MYB_SSM gene transcript abundance in stem tissue was again similar to that observed in case of leaf and root tissues. There was a decline in the fold expression in 15 day stem tissue as compared to the 10 day and 20 day tissues (Fig. 6.5). However, the fold expression decreased with maturity of the stem tissue (for 30 and 60 day old seedlings).

6.3.2.4 An overview of spatio-temporal expression in leaf, root and stem tissues for *Ll*MYB_SSM gene

Analysis of the data individually for different tissues at different developmental stages showed an irregular pattern of transcript accumulation of *Ll*MYB_SSM gene. One characteristic feature was a decline in transcript level in 15 day seedling stage as observed for all the three tissues. To determine a comparative data for the tissue showing the best spatial expression at transcript level, fold expression was calculated based on the lowest Ct value observed for 15 day leaf tissue.

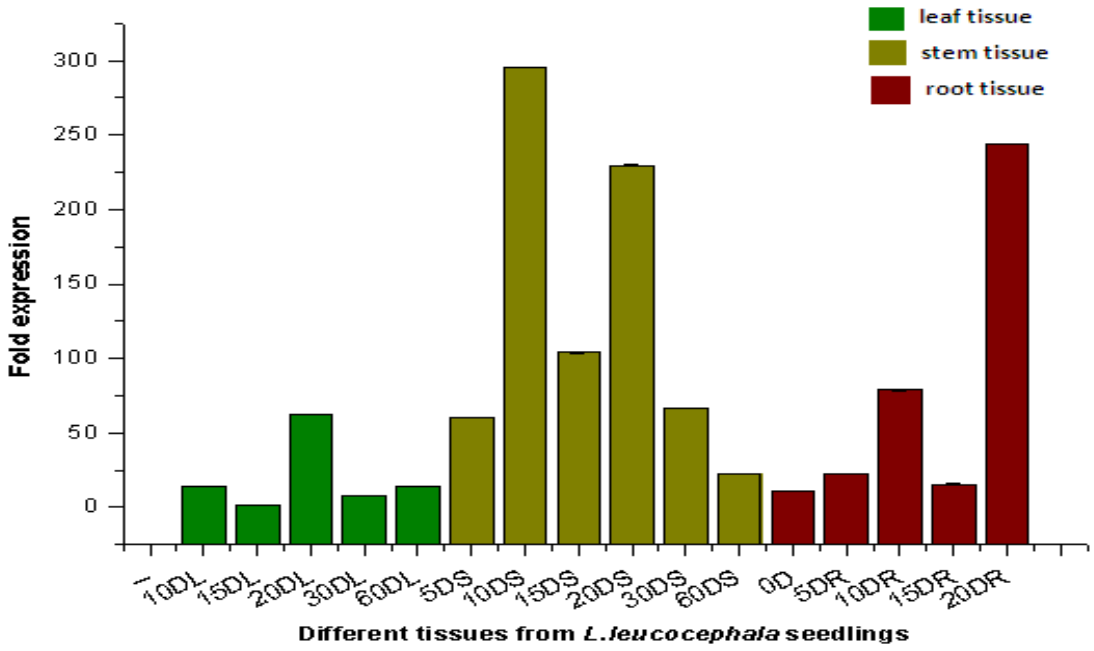


Fig. 6.6: Fold expression of *Ll*MYB_SSM transcript in different tissues at different times

The above graph (Fig.6.6) shows that *Ll*MYB_SSM transcript level is highest in 10 day stem tissue (295.43 fold) as compared to the expression in 15 day leaf tissue, followed by 20 day root (243.87 fold) and 20 day stem (229.65 fold). A general increased transcript level was seen in case as of stem tissue at every point of time. Thus, the expression of gene is possibly more in stem under unstressed conditions, which shows its best expression level in 10 and 20 day stem tissues.

6.3.3 Relative quantification of *Ll*MYB_SSM transcripts in *L. leucocephala* leaf and stem tissues grown on salt, methyl jasmonate and UV stress

Spatio-temporal studies showed that although overall transcript level of *Ll*MYB_SSM gene was variable in most of the seedling stages, it was lowest in 15 day old seedlings. As the gene expression was expected to be influenced (mostly up-regulated) due to stress conditions by virtue of the presence of EAR motif (Kazan, 2006) in its C-terminal domain, 15 day old seedlings were subjected to salt, methyl jasmonate and UV stress as

mentioned in section 6.2.5.7 of this chapter. Control or untreated sample was maintained for every stress treatment of the seedlings. Relative quantification method (section 6.2.7.8) was used to calculate the fold expression of *Ll*MYB_SSM gene in treated samples with respect to the control sample. However, effect of all the stress conditions could not be analyzed for root tissue.

6.3.3.1 Fold expression of *Ll*MYB_SSM in leaf tissues subjected to salt, methyl jasmonate and UV stress

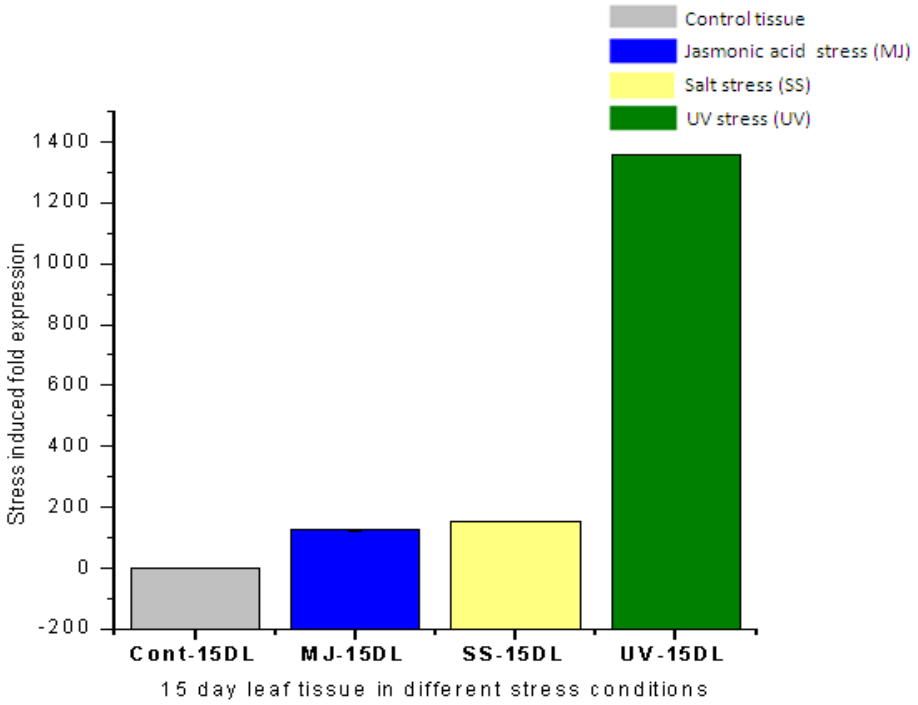


Fig. 6.7: Fold expression of *Ll*MYB_SSM transcript in different stress conditions in 15 day leaf

The above graph (Fig. 6.7) depicts a drastic increase in fold expression in 15 day leaf tissue subjected to UV stress as compared to the unstressed control leaf tissue (1357 fold). An increase in fold expression of the gene was also seen in methyl jasmonate (124 fold) and salt stress (153 fold). The high values of fold increase can be attributed to the extremely low transcript level of *Ll*MYB_SSM observed in control 15 day leaf tissues. Overall, the results indicate a positive effect of all the three stress conditions on the

expression level of the gene in 15 day leaf tissues of the plant, where UV stress showed the highest influence on gene transcript level.

6.3.3.2 Fold expression of *Ll*MYB_SSM in stem tissues subjected to salt, methyl jasmonate and UV stress

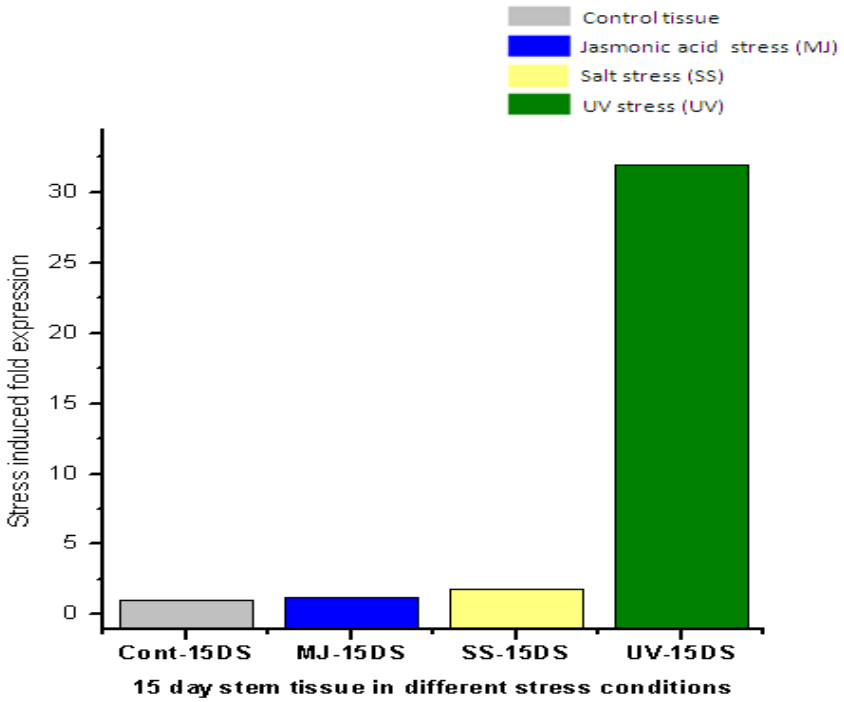


Fig. 6.8: Fold expression of *Ll*MYB_SSM transcript in different stress conditions in 15 day stem

The results from Fig. 6.8 depict an increase of fold expression of *Ll*MYB_SSM transcript levels in stem tissue exposed to UV stress with respect to the 15 day old control seedling (32 fold). However, the influence of methyl jasmonate (1.18 fold increase) and salt stress (1.89 fold increase) was not so pronounced in stem tissue as in the case of leaf tissue. The results show that *Ll*MYB_SSM gene expression is highly influenced by UV stress.

6.3.4 Western blot for detection of anti-*Ll*MYB_SSM antibody specificity

Polyclonal antibodies were raised against *Ll*MYB_SSM protein in New Zealand rabbit as mentioned in chapter 5 and the titre of first bleed antiserum was estimated to be 1:100000. However, it was necessary to check the specificity of the antibodies as

*Ll*MYB_SSM gene encoded protein belongs to R2R3MYB family constituting more than 100 members (*Arabidopsis* and *Populus*) with a conserved R2R3 DNA-binding domain. Western blotting is one such method which helps in determining specificity of the antibody for the antigen. A confirmatory experiment was conducted by running 10 % SDS gel with *E. coli* expressed *Ll*MYB_SSM protein and *E. coli* expressed another MYB protein from *L.leucocephala* (*Ll*MYB_ATS, GenBank: GU901208). The Western blotting protocol was followed as described in section 6.2.7.9 of this chapter.

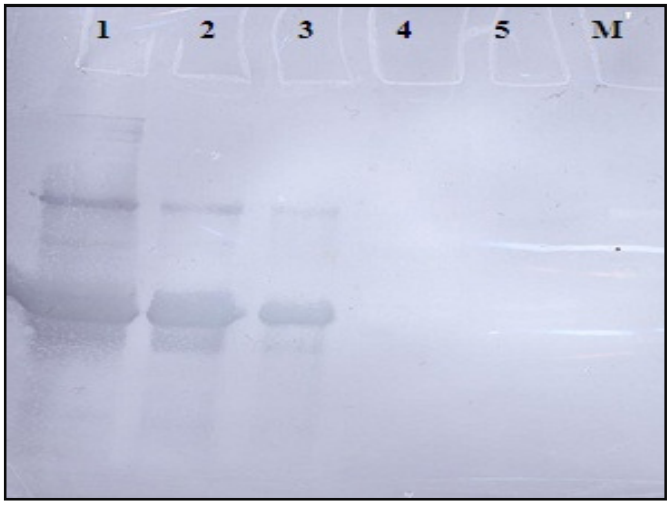


Fig. 6.9: Western blot showing detection of *Ll*MYB_SSM protein expressed in *E. coli*
Lane 1: *Ll*MYB_SSM in Inclusion bodies, Lanes 2 & 3: Purified *Ll*MYB_SSM protein (100 ng & 50 ng respectively), Lanes 4 & 5: *Ll*MYB_ATS in inclusion bodies and purified *Ll*MYB_ATS protein (100 ng) respectively not detected by Western experiment and Lane M: Marker lane

The results from Western hybridization (Fig. 6.9) show that the polyclonal anti-*Ll*MYB_SSM antibodies are specific for *Ll*MYB_SSM protein (although a few non-specific bands corresponding to *E. coli* protein are seen), as no signal was detected in another MYB protein obtained from *L. leucocephala* (Lanes 4 & 5). So, adequate dilutions of the antiserum were used for the subsequent ELISA and immuno-cytolocalisation experiments.

6.3.5 ELISA studies with *L. leucocephala* leaf and stem nuclear proteins

Western blotting with plant protein required large quantity of nuclear proteins for getting signals for *Ll*MYB_SSM protein and the transcript level of the gene was already

determined to be quite low in mature tissues. So, a more sensitive method of ELISA was used for detection of even low expression levels of the protein in mature plant tissues because of availability of larger biomass from mature plant tissues. Nuclear protein was isolated from leaf and stem tissues of matured plant as per the protocol described in chapter 2, section 2.9.5. A standard curve was plotted using serial dilutions of purified *Ll*MYB_SSM protein (5 ng, 10 ng, 20 ng, 50 ng, 100 ng and 200 ng) along with the unknown nuclear protein samples for ELISA analysis. All the samples were loaded in triplicates on the ELISA plate to avoid pipeting or experimental errors. The ELISA protocol was followed as mentioned in chapter 2, section 2.9.14.2 and the readings were taken at 405 nm using ELISA plate reader.

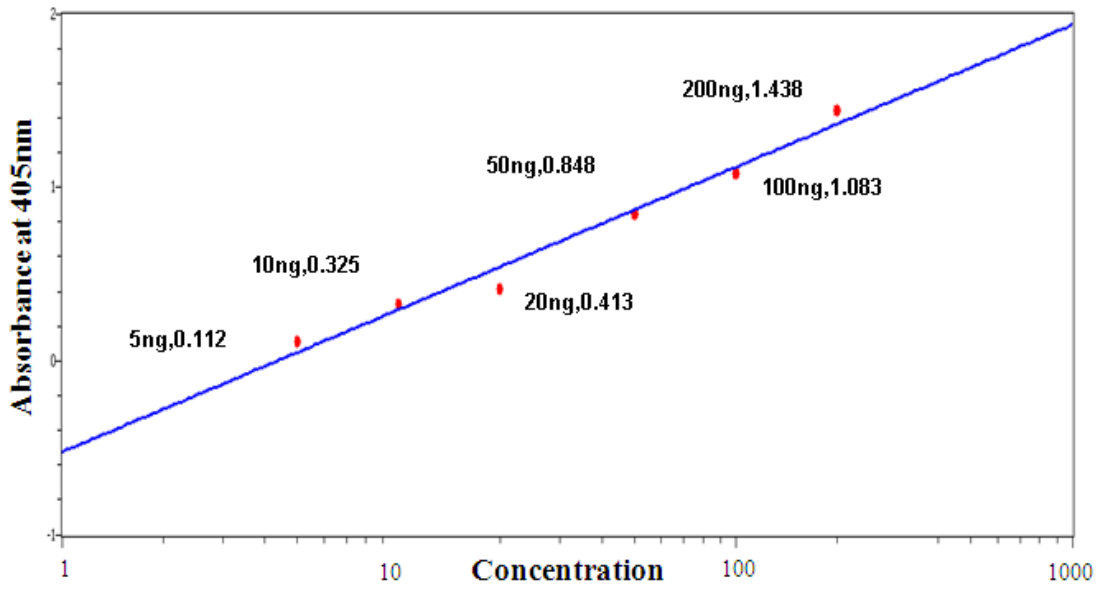


Fig. 6.10: Semilog plot showing the standard curve for different concentrations of *Ll*MYB_SSM. The values shown beside the points on the graph indicate the values of X, Y axes respectively. Values on **X-axis** represent the different concentrations of purified protein and values on **Y-axis** represent O.D at 405 nm in ELISA assay using pNPP (p-Nitro phenyl phosphate) as substrate.

The fig. 6.10 shows that most of the points corresponding to the X and Y coordinates fit well in the semilog graph plotted using the ELISA plate reader software. The absorbance obtained for the unknown (leaf and stem nuclear protein) samples was correlated with

amount of *Ll*MYB_SSM protein detected in the tissue. The results obtained for the expression based on the semilog relationship are as follows:

Sample loaded	**Mean absorbance(A_{405})**	**Concentration (ng)**
Stem nuclear protein (20 µg)	0.037	4.876ng
Leaf nuclear protein (20 µg)	0.04	negligible

The results for ELISA showed that overall the expression level of the protein was quite low in mature tissue of *L. leucocephala*. However, the protein expression was detectable in the stem nuclear protein and almost not seen in the leaf nuclear protein. Probably, the expression of *Ll*MYB _SSM is largely dependent on the stress conditions to which a plant is exposed as evident from the stress induced transcript level data for the gene mentioned in section 6.3.3.

6.3.6 Histology and Immuno-cytolocalization

10 day old stem and root samples were used for immuno-cytolocalization of *Ll*MYB_SSM protein as 10 day tissues showed the best spatial expression of the gene (Fig. 6.6)**.** Phloroglucinol lignin staining was also performed to compare the localization of the protein with tissues undergoing lignification.

The deposition of blue-black to purple precipitate, which appears brownish [Fig. 6.11(A) and 6.12(C)] after incubating with BCIP/NBT mix confirms the presence of *Ll*MYB_SSM protein near the sites of lignification, mostly the xylem tissues and phloem fibres both in the stem and root transverse sections. However, phloroglucinol stain is mainly taken up by xylem tissues [Fig. 6.11(B) and 6.12(D)]. Being a regulatory protein, the role of *Ll*MYB_SSM can be considered in regulating the genes involved in lignifying tissues, as apparent from the immuno-cytolocalisation experiment.

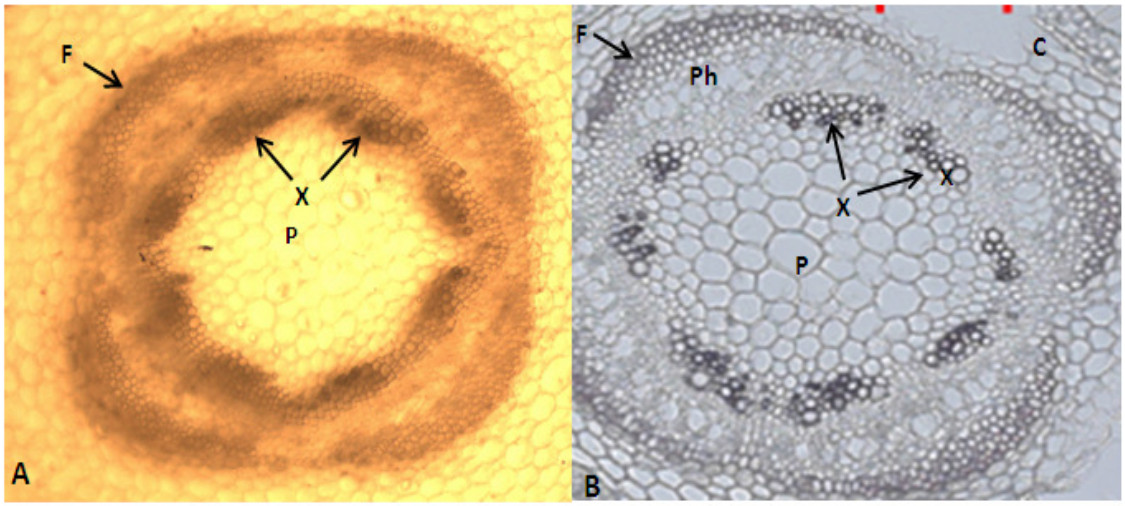


Fig. 6.11: 10X visualized transverse section of 10 day old <u>stem</u> of *L. leucocephala* (A): showing immunocytolocalisation of *Ll*MYB_SSM and (B):corresponding phloroglucinol stained section.

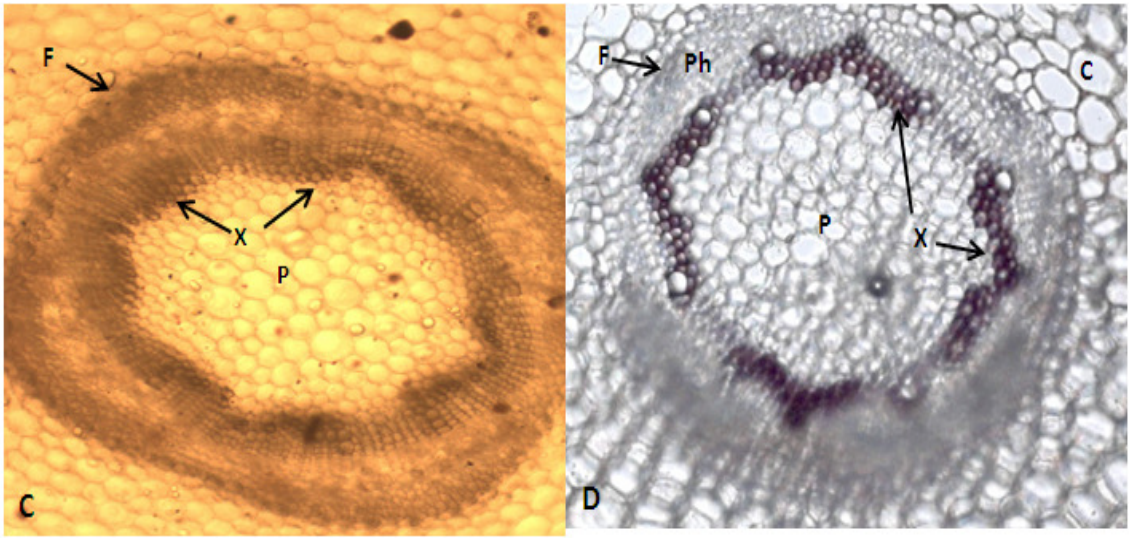


Fig. 6.12: 10X visualized transverse section of 10 day old <u>root</u> of *L. leucocephala* (C) showing immunocytolocalisation of *Ll*MYB_SSM and (D): corresponding phloroglucinol stained section. (X- Xylem, Ph- Phloem, F- Phloem fiber, C- Cortex, P- Pith)

6.4 Discussion

In this chapter an attempt was made to investigate the putative role of *Ll*MYB_SSM in lignin regulation by studying its spatio-temporal expression in the leaf, stem and root tissues of *L. leucocephala* in its different developmental stages. Expression pattern of the gene was monitored from 0 to 20 days in leaf, stem and root tissues of growing seedling using QRT-PCR. The transcript levels were also studied in leaf and stem tissues for 30 and 60 day seedlings. The Ct values obtained from real-time PCR reflect the highest

expression of the gene in stem followed by root and then in leaf tissue. However, a non-homogenous level of expression was seen in different developmental stages of the seedlings which showed a decline in expression of the gene in the 15th day seedling followed by an increase in 20th day seedling for every tissue type. A similar pattern of expression was observed in earlier studies on *L. leucocephala* stem tissue for 4-CL (Gupta, 2008) and recently in C4H. This pattern of expression can be related to the biomass accumulation in the seedling while its transition from 10th day to 15th day and 15th day to 20th day as seen from the Fig. 6.5. It can be postulated that the there is a significant increase in size of the seedling from 10th day to 15th day, which shows that greater amount of the plant's energy is involved in the primary metabolism rather than secondary metabolism (which involves lignification) and hence certain genes involved in lignification are temporarily suppressed for this transition stage. However, not much biomass accumulation is apparent between 15th and 20th day seedlings leading to further activation of those suppressed genes. Moreover, *Ll*MYB_SSM which seems to be a negative regulator of lignin biosynthesis is likely to follow a similar pattern, as its activity is not required in a suppressed state of these monolignol biosynthetic genes. Another important observation is a drastic decline in the fold expression of the candidate gene after 20 days (leaf and stem tissues) which is justifiable as the seedling shows active lignification from 30 days onwards and so most of the monolignol biosynthetic genes become actively involved in the process, so that *Ll*MYB_SSM (a postulated repressor) remains less active unless the seedling is exposed to any stress.

As per previous reports on characteristics of MYB genes belonging to subgroup4 *Ll*MYB_SSM also carries an EAR repressor motif (chapter 4, section 4.3.8.5), responsible for stress responsiveness of the genes carrying this motif. Most of the transcription factors carrying this motif are known to respond positively to abiotic stress like, salt, drought, jasmonic acid, wounding, UV etc., (Bedon *et al.*, 2010). The effects of a few stresses like jasmonic acid, salt and UV-C was specifically studied on 15 day old *L. leucocephala* seedlings, as the gene showed lowest expression at this stage. The results showed a drastic increase in the fold expression of *Ll*MYB_SSM in response to UV-C stress (15 min) both in the leaf and stem tissues of the seedling as determined using QRT-PCR experiments. As, shown for AtMYB4 (Jin *et al.*, 2000), the time of exposure to UV

stress also had a great influence on the expression level of the gene in the plants. The expression level of AtMYB4 was shown to be induced by short time exposure to UV stress but prolonged incubation (more than 6 h) in UV-B was shown to totally suppress its expression. Probably, the short time exposure elicits a strong response for the UV-protecting screens (secondary metabolites produced by phenylpropanoid pathway) in the plant. Genes like C4H and PAL which are involved in the earlier steps of this pathway are over-activated under such a stress and hence molecular brakes are required to control their activities. As, *Ll*MYB_SSM has been shown to be a negative regulator of these genes especially, C4H (chapter 5, section 5.3.8.3.3), it is likely to be activated in response to the UV stress to negatively regulate the activity of C4H by acting as a molecular brake.

The gene expression level was also increased in response to salt and jasmonic acid stress, but the fold increase was not significant in stem tissue. This shows that the repressor motif might have a role in responding to different stress conditions in the plant, which in turn have an indirect impact on stress induced lignification as well.

ELISA studies with mature plant nuclear protein showed a merely detectable level of *Ll*MYB_SSM expression in stem tissue as compared to no expression in leaf tissue. This result further confirms the stem specificity of the protein. Moreover, the extremely low level of its expression in mature tissue (which is highly lignified) further confirms it as a negative regulator of lignification. Tissue specificity of the protein was also investigated by immuno-cytolocalization studies with 10 day old stem and root transverse sections, as 10 day seedling showed the best spatio-temporal expression level for *Ll*MYB_SSM. The study revealed the localisation of the protein mainly in xylem and phloem fibres in the vascular bundles of the stem and root sections. However, phloroglucinol stain was mainly localized to the xylem tissues and so the protein has a broader role in lignification rather than being limited to regulating lignin specific genes.

6.5 Conclusions:

Regulatory role of *Ll*MYB_SSM was investigated in *L. leucocephala* seedlings at various developmental stages of the seedlings using QRT-PCR studies. The spatio-temporal expression profile showed the best expression level of the protein in stem followed by root and then leaf tissue, which was the highest (295.43 fold) in 10 day stem tissue with

respect to the lowest expression seen in 15 day leaf tissue. Exposure to UV-C stress was found to drastically increase the fold expression of the gene in leaf and stem tissues of 15 day seedling, which can be attributed to the presence of the EAR repressor motif in the C-terminal domain of the gene. An increase in fold expression was also seen in response to salt and jasmonic acid stress, which shows the broad specificity of repressor motif for different abiotic stress conditions. ELISA with nuclear protein extracts prepared from mature stem and leaf tissues further confirmed the higher expression level of the protein in stem tissue, which was in cognizance with QRT-PCR results, where transcript level of the gene was more in stem tissue as compared to leaf tissue at every developmental stage of the seedling. However, expression level of the protein was quite low in mature tissue, probably due to the lignin repressor function of the gene, as lignification increases with age of the plant. The protein was also shown to be immuno-cytolocalised in lignifying tissues (xylem and phloem fibres) of both stem and root tissues. Overall, *Ll*MYB_SSM can be postulated to have a negative regulatory role in lignification as inferred from the spatio-temporal expression profile of the gene.

SUMMARY

&

FUTURE PROSPECTS

Summary

Leucaena leucocephala is one of the most versatile, fast growing commercially important trees for paper and pulp industry in India, contributing about 25% raw material for paper & pulp industry and also valued as an excellent source of nutritious forage. The major constraint while using *Leucaena* wood as a raw material for the pulp and paper industry or forage is its extent of lignification. Removal of lignin from the pulp entails complex, expensive and environment polluting processes due to the higher proportion of phenolics present in the polymer.

Complexity of the lignin polymer lies in the ratio of its basic three monolignol subunits: p-hydroxyphenyl (H), guaiacyl (G) and syringyl (S), where S and G are largest component in angiosperm species. A higher S/G ratio and lesser total lignin content are the two desirable features for reducing the cost of lignin removal. So, there has been considerable interest in altering lignin content or composition through genetic engineering to achieve efficient and environmentally acceptable pulp production. Preliminary molecular approaches to achieve this aim focused on down-regulation of target genes of the common phenylpropanoid pathway: 4 CL (Hu *et al*., 1999) or of the monolignol specific pathway: CAD (Lapierre *et al*., 1999) and CCR (Piquemal *et al*., 1998), in various plants which demonstrated that the lignin content of the resulting transgenic plants can be strongly decreased (up to 50%). Similar studies in *L. leucocephala* also resulted in decrease in lignin content, which was also associated with phenotypic changes in the transgenics, like extremely stunted growth or bending of plants. Such effects can be a consequence of altered flux of the pathway and the use of a strongly constitutive CaMV35S promoter for driving the gene antisense constructs. To accurately target lignin modification in a tissue or cell specific manner, targeted alteration of the gene transcripts was desired, which requires understanding of the transcriptional control of lignification. R2R3 MYB transcription factors are a major group of regulatory proteins known to play a role in coordinating the regulation of phenylpropanoid pathway genes by interacting with AC-rich cis-elements found on most of their promoters (Zhong and Ye, 2007).

In order to further our understanding on the regulation of lignin biosynthetic pathway in *L. leucocephala*, promoter regions were isolated for C4H, CCoAOMT, CCR and CAD genes using PCR based approach. C4H promoter (195 bp) and CCoAOMT promoter (508 bp) were isolated using TAIL PCR approach and accession numbers were taken as JN656956 and GU13243 respectively. CAD and CCR promoters were fished out based on the sequences available in ncbi (GenBank: GU984572 and EU722904.3) by designing gene specific primers. Signal scan search against the PLACE database was used to identify different putative regulatory elements on the promoter sequences. Several of the identified regulatory elements were common to all the four promoters, like binding sites for certain transcription factors like WRKY and MYC and stress responsive motifs. An important MYBPLANT element, which is a binding site for plant R2R3 Myb transcription factors identified in several phenylpropanoid gene promoters (Douglas, 1996), was found on all the four promoter sequences. Thus, conservation in signal cascade was seen between the four gene promoters, all of them being associated to the lignin biosynthetic pathway. As, the MYBPLANT element is known to coordinate the regulation of most of the lignin biosynthetic genes in other plant species, the promoter regions carrying this element were biotinylated and used for EMSA studies with *L. leucocephala* leaf and stem nuclear proteins. The results showed that all the four promoters were retarded more by stem nuclear protein forming a single complex, while lesser retardation was seen with leaf nuclear proteins forming two complexes. Thus, promoter-specific regulatory proteins were found in both leaf and stem nuclear extracts, with some additional regulatory protein found in leaf extract due to the formation of two complexes. However, affinity of stem nuclear extract appeared to be more for the promoters, because of lignin regulatory proteins being more in stem tissues compared to leaf tissue with respect to the lignin deposition. CCoAOMT promoter driven GUS-gene fusion construct cloned in pCAMBIA-1381Z vector was used to transform *N. tabacum* leaf explants, which were regenerated to give putative transformants confirmed positive for hygromycin resistance gene. GUS assay for the transformed leaf tissues showed that the GUS activity was restricted to the midrib and veins. Thus, a partial promoter was shown sufficient to retain the vascular-tissue specific activity in a transient expression study. As, the other 4 promoter sequences also carried similar promoter elements and

showed similar retardation with the nuclear proteins, they could also have a tissue-specific activity similar to CCoAOMT promoter.

Another part of the objectives was aimed at characterizing 'lignin-specific' trans-activators; in *L. leucocephala*. Considering the potential role for R2R3 Myb transcription factors in regulating the phenylpropanoid pathway genes, a putative R2R3 MYB transcription factor was isolated from xylem tissue cDNA pool of the plant, using degenerate primer based approach. A full length gene with an ORF of 705 bp (GenBank: GU901209), encoding for 235 amino acid long polypeptide was fished out, cloned and named *Ll*MYB_SSM. The 3'UTR (203 bp) and 5' UTR (25 bp) of the gene were fished out using RACE approach. Bioinformatic characterization of the *in silico* translated gene revealed the highly conserved R2 and R3 DNA-binding domains in its N-terminal, showing upto 95% similarity with the R2R3 Myb genes belonging to subgroup 4 of the R2R3 Myb gene family based on the classification criteria of Stracke *et al.*, 2001 and Kranz *et al.*, 1998. However, the C-terminal domain was highly divergent, but retained a few motifs characteristic for subgroup 4 genes which were explored using MEME software. One of the motifs (pdLNL[D/E]LXi[G/S) containing the EAR motif is known to play an essential role in active repression (Kazan, 2006; Hiratsu *et al.*, 2003). Phylogenetic analysis using the full length amino acid sequence of *Ll*MYB_SSM showed that *Am*MYB308, known to be a repressor of lignin pathway genes (Tamagnone *et al.*, 1998), was the closest member of the gene. To identify the copy number of the gene in *L. leucocephala* genome Southern hybridization was performed with the C-terminal sequence of the gene as a probe, which revealed atleast 3-4 copies or isoforms of the gene being present.

To elucidate the function of *Ll*MYB_SSM, the gene was cloned in pET28b (+) and pET 41a (+) vectors and heterologously expressed in *E. coli* (BL21 DE3) strain. Recombinant protein of ~30 kDa fused with 6-Histidine tag was purified using Ni-NTA column in denaturing conditions from pET28b (+) expressed proteins, which was used for raising polyclonal antibodies in NewZealand rabbit. To obtain the protein in soluble fraction for DNA-binding studies, GST-fused protein was expressed using pET 41a (+) vector and purified in mild conditions using GST-bind resin. However, EMSA studies of the purified protein with the MYBPLANT element carrying promoters (C4H,

CCoAOMT, CCR and CAD) did not show any retardation. As per earlier reports, other factors like post-translational modifications or involvement of a partner protein in DNA-binding could be the possible reasons for no binding. Hence, *in planta* gain of function approach was used to determine the role of *Ll*MYB_SSM, by heterologous over-expression in *N. tabacum* driven by CaMV35S promoter. Considering the putative role of the gene in regulating the phenylpropanoid pathway, the putative transformants were analysed for the alteration in the transcript levels of few genes: PAL, C4H, 4-CL, CCoAOMT and CAD using qRT-PCR. Two *Ll*MYB_SSM over-expressing transformants (P1 and P3), which virtually showed no phenotypic difference with respect to control plant, showed a decline in the transcript levels of all the genes except CCoAOMT. The suppression of transcripts was most prominent for C4H and PAL genes, which was similar to the effect of AmMYB308 (Tamagnone *et al*., 1998) and AtMYB4 (Jin *et al*., 2000) over-expression produced in tobacco plants. 4-CL transcript level was also decreased to a similar level as PAL and C4H; whereas, CAD was the least influenced at transcript level. The results indicate that *Ll*MYB_SSM might act as a repressor for general phenylpropanoid pathway genes (PAL, C4H and 4-CL) with lesser effects on the downstream genes of the monolignol biosynthetic pathway (CAD). Molecular modeling of the gene was also performed using I-TASSER tool to gain further insight into the topology of the protein structure. The 3-D model was validated using PROCHECK and Verify 3-D tools and it helped in identifying the folds and motifs responsible for DNA-binding in the protein.

Quantitative real time-PCR was used to investigate the spatio-temporal expression of the gene in the stem, root and leaf tissues of 0-20 day of growing seedlings and also in leaf and stem tissues of 30 and 60 day old seedlings. The fold expression profile showed the best transcript level of the gene in stem tissues, followed by root and then leaf tissue. Temporal expression of the gene was highest in 10 day stem tissue and least in 15 day leaf tissue. A drastic decline in the fold expression of the gene was observed after 20 days (leaf and stem tissues) which is justifiable as the seedling shows active lignification from 30 days onwards and so most of the monolignol biosynthetic genes become actively involved in the process, so that *Ll*MYB_SSM (a postulated repressor) remains less active. To determine the role of the putative repressor motif (pdLNL[D/E]LXi[G/S) found on C-

terminal region of the gene, known to be elicited in abiotic stress conditions, 15 day *L. leucocephala* seedlings were subjected to various stress like UV, salt and jasmonic acid. A drastic increase in the gene transcript level was observed in response to the UV stress especially in the leaf and stem tissues with respect to the unstressed seedling. Based on *in silico* analysis using ProtComp version9 server, subcellular localization of *Ll*MYB_SSM was predicted to be a nuclear, with a comfortably high score. So, ELISA was performed with nuclear protein extracts prepared from mature stem and leaf tissues, which further confirmed that the protein is preferably expressed in stem as compared to leaf tissue, although the expression was quite low because of the lower expression of the protein in mature tissues. The protein was also shown to be immuno-cytolocalised in lignifying tissues (xylem and phloem fibres) of both stem and root tissues of 10 day old seedling.

The findings of the thesis can be summarized as, promoter analysis for few of the lignin biosynthetic genes from *L. leucocephala* revealed the presence of potential AC-rich elements involved in co-regulating their expression by interacting with certain R2R3Myb transcription factors. To understand the transcriptional regulation of these genes, *Ll*MYB_SSM, an R2R3MYB gene encoding a transcription factor was isolated from *L. leucocephala*, which can be postulated as a potential negative regulator of a few lignin biosynthetic genes based on bioinformatic and molecular biology experimental data.

Future prospects

- The promoters isolated in the study can be utilized for competitive EMSA in combination with deletion analysis of the promoter driven reporter gene constructs to determine the minimal functional elements required for their tissue specific expression

- The role of *Ll*MYB_SSM as a putative negative regulator of lignin biosynthesis can be further confirmed by mutagenesis of the repressor motif found in the C-terminal domain of the protein and then using it for *in planta* over-expression studies.

- Being a negative regulator of lignin biosynthetic genes, *Ll*MYB_SSM can be over-expressed in *L. leucocephala* driven by a vascular tissue-specific promoter like CCoAOMT to reduce the lignin content of the plant without affecting the metabolic flux of the pathway.

REFERENCES

Abdalla, K. O., Thomson, J. A., & Rafudeen, M. S. (2009). Protocols for nuclei isolation and nuclear protein extraction from the resurrection plant *Xerophyta viscosa* for proteomic studies. *Analytical Biochemistry*, *384*(2), 365-367.

Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2003). *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *The Plant Cell*, *15*(1), 63-78.

Abhilash, O. U. (2009). Studies on peroxidases and their potential role in lignin metabolism of *Leucaena leucocephala, PhD thesis,* 96-97. University of Pune, NCL.

Aloni, R., Tollier, M. T., & Monties, B. (1990). The role of auxin and gibberellin in controlling lignin formation in primary phloem fibers and in xylem of *Coleus blumei* stems. *Plant Physiology*, *94*(4), 1743-1747.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389-3402.

Alwen, A., R.M.B. Moreno. 0., Vicente and E. Hebcrle-Bars, (1992). Plant endogenous β-glucuronidase activity: how to avoid interference with the use of the *E. coli* β -glucuronidase as a reporter gene in transgenic plants. *Transgenic Res*.,1, 63-70.

Amthor, J. S. (2003). Efficiency of lignin biosynthesis: a quantitative analysis. *Annals of Botany*, *91*(6), 673-695.

An, G. (1987). Binary Ti vectors for plant transformation and promoter analysis. *Methods in Enzymology*, *153*, 292-305.

Argyropoulos, D. S., Jurasek, L., Kristofová, L., Xia, Z., Sun, Y., & Palus, E. (2002). Abundance and reactivity of dibenzodioxocins in softwood lignin. *Journal of Agricultural and Food Chemistry*, *50*(4), 658-666.

Arnheim, N., & Erlich, H. (1992). Polymerase chain reaction strategy. *Annual Review of Biochemistry*, *61*, 131-156.

Avila, J., Nieto, C., Canas, L., et al. (1993). *Petunia hybrida* genes related to the maize regulatory C1 gene and to animal myb proto-onco gene. *Plant journal, 3*, 553-562

Azeemoddin, G., Rao, J. M., Thirumala, S., Rao S.D. (1988). Amino acid composition of subabul (*Leucaena leucocephala*) seed kernel proteins. *J Food Sci Technol*, *25*, 158-163.

Bailey, T. L., & Elkan, C. (1994a). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings International Conference on Intelligent Systems for Molecular Biology ISMB International Conference on Intelligent Systems for Molecular Biology*, *2*(6), 28-36.

Bailey, T. L., & Elkan, C. (1994b). Fitting a mixture model by expectation maximisation to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, 2*(6), 28-36.

Baranowskij, N., Frohberg, C., Prat, S., & Willmitzer, L. (1994). A novel DNA binding protein with homology to MYB oncoproteins containing only one repeat can

function as a transcriptional activator. *the The European Molecular Biology Organization Journal*, *13*(22), 5383-5392.

Bedon, F., Bomal, C., Caron, S *et al*. (2010). Subgroup 4 R2R3-MYBs in conifer trees: gene family expansion and contribution to the isoprenoid- and flavonoid- oriented responses. *Journal of Experimental Botany, 61*(14), 3847-3864

Bedon, F., Grima-Pettenati, Jacqueline, & Mackay, John. (2007). Conifer R2R3-MYB transcription factors: sequence analyses and gene expression in wood-forming tissues of white spruce (*Picea glauca*). *BMC Plant Biology*, *7*, 1-17.

Bell-Lelong, D. A., Cusumano, J. C., Meyer, K., & Chapple, C. (1997). Cinnamate-4-hydroxylase expression in *Arabidopsis*. Regulation in response to development and the environment. *Plant Physiology*, *113*(3), 729-738.

Bender, J., & Fink, G. R. (1998). A Myb homologue, ATR1, activates tryptophan gene expression in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(10), 5655-5660.

Biemelt, S., Tschiersch, H., & Sonnewald, U. (2004). Impact of altered gibberellin metabolism on biomass accumulation, lignin biosynthesis, and photosynthesis in transgenic tobacco plants. *Plant Physiology*, *135*(1), 254-265.

Blanchette, M., & Sinha, S. (2001). Separating real motifs from their artifacts. *Bioinformatics*, *17 Suppl 1*(1), 30-38.

Blount, J. W., Korth, K. L., Masoud, S. A., Rasmussen, S., Lamb, C., & Dixon, Richard A. (2000). Altering expression of cinnamic acid 4-hydroxylase in transgenic plants provides evidence for a feedback loop at the entry point into the phenylpropanoid pathway. *Plant Physiology*, *122*(1), 107-116.

Blumberg, D. D. (1987). Creating a ribonuclease-free environment. *Methods in Enzymology*, *152*, 20-24.

Boerjan, Wout, Ralph, John, & Baucher, M. (2003). Lignin biosynthesis. *Annual Review of Plant Biology*, *54*(1), 519-546.

Boija, E., Lundquist, A., Edwards, K., & Johansson, G. (2007). Evaluation of bilayer disks as plant cell membrane models in partition studies. *Analytical Biochemistry*, *364*(2), 145-152.

Boija, E., Lundquist, A., Nilsson, M., Edwards, K., Isaksson, R., & Johansson, G. (2008). Bilayer disk capillary electrophoresis: a novel method to study drug partitioning into membranes. *Electrophoresis*, *29*(16), 3377-3383.

Bolwell, G.P., Cramer, C.L., Lamb, C.J., Schuch, W., Dixon, R.A. (1986). L-phenylalanine ammonia lyase from *Phaseolus vulgaris*: modulation of the levels of active enzyme by transcinnamic acid. *Planta*, *169*, 97-107.

Bomal, C., Bedon, F., Caron, S., Mansfield, Shawn D, Levasseur, C., Cooke, J. E. K., Blais, S., et al. (2008). Involvement of *Pinus taeda* MYB1 and MYB8 in phenylpropanoid metabolism and secondary cell wall biogenesis: a comparative in planta analysis. *Journal of Experimental Botany*, *59*(14), 3925-3939.

Borevitz, J. O., Xia, Y., Blount, J., Dixon, R. A., & Lamb, C. (2000). Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *The Plant Cell*, *12*(12), 2383-2393.

Bowles, D., Lim, E.-K., Poppenberger, B., & Vaistij, F. E. (2006). Glycosyltransferases of lipophilic small molecules. *Annual Review of Plant Biology*, *57*(1), 567-597.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, *72*, 248-254.

Braun, E L, & Grotewold, E. (1999). Newly discovered plant c-myb-like genes rewrite the evolution of the plant MYB gene family. *Plant Physiology, 121*(1), 21-24.

Bray, R. A. et al., (1988). Performance of 14 *Leucaena* lines at five sites in Queensland. *Aust. J. exp. Agric*, 28: 69-76.

Brewbaker, J.L., and Sorenson, C.T. (1990). *Leucaena*: New tree crops from interspecific hybrids. In: Advances in new crops. Timber Press. Oregon, pp 238

Bustin, S A. (2002). Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *Journal of Molecular Endocrinology*, *29*(1), 23-39.

Bustin, Stephen A, & Nolan, T. (2004). Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *Journal of biomolecular techniques*, *15*(3), 155-166.

Cano-Delgado, A., Penfield, S., Smith, Caroline, Catley, M., & Bevan, Michael. (2003). Reduced cellulose synthesis invokes lignification and defense responses in *Arabidopsis thaliana*. *The Plant Journal*, *34*(3), 351-362.

Capellades, M, Torres, M. A., Bastisch, I., Stiefel, V., Vignols, F., Bruce, W. B., Peterson, D., et al. (1996). The maize caffeic acid O-methyltransferase gene promoter is active in transgenic tobacco and maize plant tissues. *Plant Molecular Biology*, *31*(2), 307-322.

Caño-Delgado, A. I., Metzlaff, K., & Bevan, M W. (2000). The *eli*1 mutation reveals a link between cell expansion and secondary cell wall formation in *Arabidopsis thaliana*. *Development Cambridge England*, *127*(15), 3395-3405.

Carle Urioste, J.C., Marrs, K., Bodeau, J., & Walbot. V. (1995). Direct gene transfer to protoplasts. Gene transfer to protoplasts: transient gene expression analysis. pp 106-111.

Chabannes, M, Barakate, A., Lapierre, C, Marita, J M, Ralph, J, Pean, M., Danoun, S, et al. (2001). Strong decrease in lignin content without significant alteration of plant development is induced by simultaneous down-regulation of cinnamoyl CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD) in tobacco plants. *The Plant Journal*, *28*(3), 257-270.

Chandrasekhara, R.T., Lakshminarayana, G., Prasad, N. B. L., Rao, S. M., S. Azeemoddin G. Atchynta R. D., & Thirumala Rao SD. (1984). Characteristics and

compositions of *Carissa spinarum*, *Leucaena leucocephala* and *Physalis minima* seeds and oils. *J Am Oil Chem Soc*. *61*, 1472-1483.

Cheng, H., Song, S., Xiao, L., Soo, H. M., Cheng, Z., Xie, D., & Peng, J. (2009). Gibberellin acts through jasmonate to control the expression of MYB21, MYB24, and MYB57 to promote stamen filament growth in *Arabidopsis*. (G. P. Copenhaver, Ed.) *PLoS Genetics*, *5*(3), 13.

Chumakov, K. M. (1994). Reverse transcriptase can inhibit PCR and stimulate primer-dimer formation. *PCR Methods And Applications*, *4*(1), 62-64.

Cocciolone, S. M., Chopra, S, Flint-Garcia, S. A., McMullen, M D, & Peterson, T. (2001). Tissue-specific patterns of a maize MYB transcription factor are epigenetically regulated. *The Plant Journal*, *27*(5), 467-478.

Coley, P. D., Bryant, J. P., & Chapin, F. S. I. I. I. (1985). Resource availability and plant antiherbivore defense. *Science*, *230* (4728), 895-899.

Cominelli, E., & Tonelli, Chiara. (2009). A new role for plant R2R3-MYB transcription factors in cell cycle regulation. *Cell research*, *19*(11), 1231-1242.

Cominelli, E., Galbiati, M., Vavasseur, A., Conti, L., Sala, T., Vuylsteke, M., Leonhardt, N., et al. (2005). A guard-cell-specific MYB transcription factor regulates stomatal movements and plant drought tolerance. *Current Biology*, *15*(13), 1196-1200.

Cominelli, E., Sala, T., Calvi, D., Gusmaroli, G., & Tonelli, Chiara. (2008). Over-expression of the *Arabidopsis* AtMYB41 gene alters cell expansion and leaf surface permeability. *The Plant Journal*, *53*(1), 53-64.

Cook, N. L., Vink, R., Donkin, J. J., & Van Den Heuvel, C. (2009). Validation of reference genes for normalization of real-time quantitative RT-PCR data in traumatic brain injury. *Journal of Neuroscience Research*, *87*(1), 34-41.

Daniel, X., Sugano, Shoji, & Tobin, E. M. (2004). CK2 phosphorylation of CCA1 is necessary for its circadian oscillator function in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(9), 3292-3297.

Das, A. (1990). Overproduction of proteins in *Escherichia coli*: vectors, hosts, and strategies. *Methods in Enzymology*, *182*(1985), 93-112.

Das, R., Qian, B., Raman, S., Vernon, R., Thompson, J., Bradley, P., Khare, S., et al. (2007). Structure prediction for CASP7 targets using extensive all-atom refinement with Rosetta@home. *Proteins*, *69*(8), 118-128.

Deluc, L, Barrieu, F, Marchive, C, Lauvergeat, V, Decendit, A, Richard, T, Carde, J. P., et al. (2006). Characterization of a grapevine R2R3-MYB transcription factor that regulates the phenylpropanoid pathway. *Plant Physiology*, *140*(2), 499-511.

Demura, T., & Fukuda, H. (2007). Transcriptional regulation in wood formation. *Trends in Plant Science*, *12*(2), 64-70.

Devaiah, B. N., Madhuvanthi, R., Karthikeyan, A. S., & Raghothama, K. G. (2009). Phosphate starvation responses and gibberellic acid biosynthesis are regulated by the myb62 transcription factor in *Arabidopsis*. *Molecular plant*, *2*(1), 43-58.

Dharmawardhana, D. P., Ellis, B E, & Carlson, J. E. (1999). cDNA cloning and heterologous expression of coniferin beta-glucosidase. *Plant Molecular Biology*, *40*(2), 365-372.

Dharmawardhana, D. P., Ellis, B E, & Carlson, J. E. (1995). A beta-glucosidase from lodgepole pine xylem specific for the lignin precursor coniferin. *Plant Physiology*, *107*(2), 331-339.

Dias, A. P., Braun, Edward L, McMullen, Michael D, & Grotewold, Erich. (2003). Recently duplicated maize R2R3 MYB genes provide evidence for distinct mechanisms of evolutionary divergence after duplication. *Plant Physiology, 131*(2), 610-620.

Douglas, C J, Hauffe, K. D., Ites-Morales, M. E., Ellard, M., Paszkowski, U., Hahlbrock, K., & Dangl, J. L. (1991). Exonic sequences are required for elicitor and light activation of a plant defense gene, but promoter sequences are sufficient for tissue specific expression. *The European Molecular Biology Organization Journal*, *10*(7), 1767-1775.

Du, H., Zhang, L., Liu, L., Tang, X.-F., Yang, W.-J., Wu, Y.-M., Huang, Y.-B., et al. (2009). Biochemical and molecular characterization of plant MYB transcription factor family. *Biochemistry*, *74*(1), 1-11.

Dubos, C., Stracke, Ralf, Grotewold, Erich, Weisshaar, Bernd, Martin, Cathie, & Lepiniec, L. (2010). MYB transcription factors in *Arabidopsis*. *Trends in Plant Science*, *15*(10), 573-581.

Duprey, S. P., & Boettiger, D. (1985). Developmental regulation of c-MYB in normal myeloid progenitor cells. *Proceedings of the National Academy of Sciences of the United States of America*, *82*(20), 6937-41.

Duthie, G., & Crozier, A. (2000). Plant-derived phenolic antioxidants. *Current Opinion in Clinical Nutrition and Metabolic Care*, *3*(6), 447-451.

Eckert, K. A., & Kunkel, T. A. (1990). High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nucleic Acids Research*, *18*(13), 3739-3744.

Ehlting, J., Mattheus, N., Aeschliman, D. S., Li, E., Hamberger, B., Cullis, I. F., Zhuang, J., et al. (2005). Global transcript profiling of primary stems from *Arabidopsis thaliana* identifies candidate genes for missing links in lignin biosynthesis and transcriptional regulators of fiber differentiation. *The Plant Journal*, *42*(5), 618-640.

Ellis, C., Karafyllidis, I., Wasternack, C., & Turner, J. G. (2002). The *Arabidopsis* mutant *cev1* links cell wall signaling to jasmonate and ethylene responses. *Plant Physiology*, *14*(7), 1557-1566.

Eriksson, M E, Israelsson, M, Olsson, O., & Moritz, T. (2000). Increased gibberellin biosynthesis in transgenic trees promotes growth, biomass production and xylem fiber length. *Nature Biotechnology*, *18*(7), 784-788.

Escamilla-Treviño, L. L., Chen, W., Card, M. L., Shih, M.-C., Cheng, C.-L., & Poulton, J. E. (2006). *Arabidopsis thaliana* beta-Glucosidases BGLU45 and BGLU46 hydrolyse monolignol glucosides. *Phytochemistry*, *67*(15), 1651-1660.

Feinberg, A. P., & Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry*, *132*(1), 6-13.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, *39*(4), 783-791.

Feng, C., Andreasson, E., Maslak, A., Peter, H., Mattsson, O., & Mundy, John. (2004). *Arabidopsis* MYB68 in development and responses to environmental cues. *Plant Science*, *167*(5), 1099-1107.

Feuillet, C., Lauvergeat, V, Deswarte, C., Pilate, G, Boudet, A., & Grima-Pettenati, J. (1995). Tissue- and cell-specific expression of a cinnamyl alcohol dehydrogenase promoter in transgenic poplar plants. *Plant Molecular Biology*, *27*(4), 651-667.

Fornalé, S., Sonbol, F.-M., Maes, T., Capellades, Montserrat, Puigdomènech, Pere, Rigau, Joan, & Caparrós-Ruiz, D. (2006). Down-regulation of the maize and *Arabidopsis thaliana* caffeic acid O-methyl-transferase genes by two new maize R2R3-MYB transcription factors. *Plant Molecular Biology*, *62*(6), 809-823.

Freeman, W. M., Walker, S. J., & Vrana, K. E. (1999). Quantitative RT-PCR: pitfalls and potential. *Biotechniques*, *26*(1), 112-122.

Fritz, C., Palacios-Rojas, N., Feil, R., & Stitt, M. (2006). Regulation of secondary metabolism by the carbon-nitrogen status in tobacco: nitrate inhibits large sectors of phenylpropanoid metabolism. *The Plant Journal*, *46*(4), 533-548.

Frohman, M. A., Dush, M. K., & Martin, G. R. (1988). Rapid production of full length cDNAs from rare transcripts: Amplification using a single gene specific oligonucleotide primer. *Proceedings of the National Academy of Sciences of the United States of America*, 85, 8998-9002.

Fukuda, H., & Komamine, A. (1982). Lignin synthesis and its related enzymes as markers of tracheary element differentiation in single cells isolated from the mesophyll of *Zinnia elegans*. *Planta*, *155*, 423-430.

Fütterer, J. (1995a). Genetics markers and expression signals. Expression signals and vectors. *Gene transfer to plants* pp 311-324.

Gabaldón, C., López-Serrano, M., Pedreño, M. A., & Barceló, A. R. (2005). Cloning and molecular characterization of the basic peroxidase isoenzyme from *Zinnia elegans*, an enzyme involved in lignin biosynthesis. *Plant Physiology*, *139*(3), 1138-1154.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook* pp. 571-607.

Gavnholt, B., & Larsen, K. (2002). Molecular biology of plant laccases in relation to lignin formation. *Physiologia Plantarum*, *116*(3), 273-280.

Gebauer, R. L. E., Strain, B. R., & Reynolds, J. F. (1998). The effect of elevated CO_2 and N availability on tissue concentrations and whole plant pools of carbon-based secondary compounds in loblolly pine (*Pinus taeda*). *Oecologia*, *113*, 29-36.

Gerard, G. F., D'Alessio, J. M., Kotewicz, M. L., & Noon, M. C. (1986). Influence on stability in *Escherichia coli* of the carboxy-terminal structure of cloned Moloney murine leukemia virus reverse transcriptase. *Dna Mary Ann Liebert Inc*, *5*(4), 271-279.

Gigolashvili, T., Berger, B., Mock, H.-P., Müller, C., Weisshaar, Bernd, & Flügge, U.-I. (2007). The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. *The Plant Journal*, *50*(5), 886-901.

Gigolashvili, T., Engqvist, M., & Yatusevich, R. (2006). HAG2 / MYB76 and HAG3 / MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate biosynthesis in *Arabidopsis thaliana*. *New Phytologist*, 627-642.

Gigolashvili, T., Yatusevich, R., Berger, B., Müller, C., & Flügge, U.-I. (2007). The R2R3-MYB transcription factor HAG1/MYB28 is a regulator of methionine-derived glucosinolate biosynthesis in *Arabidopsis thaliana*. *The Plant Journal*, *51*(2), 247-261.

Goffner, D., Campbell, M. M., Campargue, C., Clastre, M., Borderies, G., Boudet, a, & Boudet, a M. (1994). Purification and characterization of Cinnamoyl-Coenzyme A:NADP Oxidoreductase in *Eucalyptus gunnii*. *Plant physiology*, *106*(2), 625-632.

Goicoechea, Monica, Lacombe, Eric, Legay, Sylvain, Mihaljevic, S., Rech, P., Jauneau, A., Lapierre, Catherine, et al. (2005). EgMYB2, a new transcriptional activator from *Eucalyptus* xylem, regulates secondary cell wall formation and lignin biosynthesis. *The Plant Journal*, *43*(4), 553-567.

Gómez Ros, L V, Gabaldón, C., Pomar, F., Merino, F., Pedreño, M. A., & Barceló, A. R. (2007). Structural motifs of syringyl peroxidases predate not only the gymnosperm-angiosperm divergence but also the radiation of tracheophytes. *New Phytologist*, *173*(1), 63-78.

Gong, Z. Z., Yamagishi, E., Yamazaki, M., & Saito, K. (1999). A constitutively expressed Myc-like gene involved in anthocyanin biosynthesis from *Perilla frutescens*: molecular characterization, heterologous expression in transgenic plants and transactivation in yeast cells. *Plant Molecular Biology*, *41*(1), 33-44.

Gonzalez, A., Zhao, M., Leavitt, J. M., & Lloyd, A. M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *The Plant Journal*, *53*(5), 814-827.

Gonzalez, D., Bowen, A. J., Carroll, T. S., & Conlan, R. S. (2007). The transcription corepressor LEUNIG interacts with the histone deacetylase HDA19 and mediator components MED14 (SWP) and CDK8 (HEN3) to repress transcription. *Molecular and Cellular Biology*, *27*(15), 5306-5315.

Goodenow, M., Huet, T., Saurin, W., Kwok, S., Sninsky, J., & Wain-Hobson, S. (1989). HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions. *Journal of Acquired Immune Deficiency Syndromes*, *2*(4), 344-352.

Grabber, J. H., Hatfield, R. D., Lu, F., & Ralph, John. (2008). Coniferyl ferulate incorporation into lignin enhances the alkaline delignification and enzymatic degradation of cell walls. *Biomacromolecules*, *9*(9), 2510-2516.

Grimmig, B., & Matern, U. (1997). Structure of the parsley caffeoyl-CoA O-methyltransferase gene, harbouring a novel elicitor responsive cis-acting element. *Plant Molecular Biology*, *33*(2), 323-341.

Groover, A. T., Mansfield, Shawn D, DiFazio, S. P., Dupper, G., Fontana, J. R., Millar, R., & Wang, Y. (2006). The Populus homeobox gene ARBORKNOX1 reveals overlapping mechanisms regulating the shoot apical meristem and the vascular cambium. *Plant Molecular Biology*, *61*(6), 917-932.

Groover, A., & Robischon, M. (2006). Developmental mechanisms regulating secondary growth in woody plants. *Current Opinion in Plant Biology*, *9*(1), 55-58.

Gubler, F, Raventos, D., Keys, M., Watts, R., Mundy, J, & Jacobsen, J V. (1999). Target genes and regulatory domains of the GAMYB transcriptional activator in cereal aleurone. *The Plant Journal*, *17*(1), 1-9.

Gupta, S.K. (2008). Isolation, cloning and characterization of lignin biosynthesis pathway gene(s) 4-Coumarate Co A Ligase (4CL) from *Leucaena leucocephala*., *PhD thesis,* 141-146.University of Pune, NCL.

Gutteridge, R.C., & Shelton, H. M. (1998). *Forage Tree Legumes in Tropical Agriculture*. Tropical Grassland Society of Australia, Inc., 2.1 "*Leucaena leucocephala* - the Most Widely Used Forage Tree Legume"

Haga, N., Kato, K., Murase, M., Araki, S., Kubo, M., Demura, T., Suzuki, K., et al. (2007). R1R2R3-Myb proteins positively regulate cytokinesis through activation of KNOLLE transcription in *Arabidopsis thaliana*. *Development Cambridge England*, *134*(6), 1101-1110.

Hajdukiewicz, P., Svab, Z., & Maliga, P. (1994). The small, versatile pPZP family of *Agrobacterium* binary vectors for plant transformation. *Plant Molecular Biology*, *25*(6), 989-994.

Hammond, A. C., Allison, M. J., Williams, M. J., Prine, G. M., & Bates, D. B. (1989a). Prevention of leucaena toxicosis of cattle in Florida by ruminal inoculation with 3-hydroxy-4-(1H)-pyridone-degrading bacteria. *American Journal Of Veterinary Research*, *50*(12), 2176-2180.

Hammond, A.C., Allison, M.J., Williams, M.J. (1989b). Persistence of DHP-degrading bacteria between growing seasons in subtropical Florida. *Leucaena Reseach Report*. *10*, 66.

Harris, P. J., Kelderman, M. R., Kendon, M. F., & McKenzie, R. J. (1997). Monosaccharide compositions of unlignified cell walls of monocotyledons in relation to the occurrence of wall-bound ferulic acid. *Biochemical Systematics and Ecology*, *25*(2), 167-179.

Hatton, D., Sablowski, R., Yung, M. H., Smith, C, Schuch, W, & Bevan, M. (1995). Two classes of cis sequences contribute to tissue-specific expression of a *PAL2* promoter in transgenic tobacco. *The Plant Journal*, *7*(6), 859-876.

Hauffe, K. D., Lee, S. P., Subramaniam, R., & Douglas, C J. (1993). Combinatorial interactions between positive and negative cis-acting elements control spatial patterns of *4CL*-1 expression in transgenic tobacco. *The Plant Journal*, *4*(2), 235-253.

Hauffe, K. D., Paszkowski, U., Schulze-Lefert, P., Hahlbrock, K., Dangl, J. L., & Douglas, C J. (1991). A parsley *4CL*-1 promoter fragment specifies complex expression patterns in transgenic tobacco. *The Plant Cell*, *3*(5), 435-443.

Hegarty, M.P., Court, R.D. & Thorne, P.M. (1964).The determination of mimosine and 3, 4-DHP in biological material. *Aust. J. Agric.Res., 15*, 168.

Hehl, R., & Wingender, E. (2001). Database-assisted promoter analysis. *Trends in Plant Science*, *6*(6), 251-255.

Heid, C. A., Stevens, J., Livak, K. J., & Williams, P. M. (1996). Real time quantitative PCR. *Genome Research*, *6*(10), 986-994.

Heine, G. F., Hernandez, J. M., & Grotewold, Erich. (2004). Two cysteines in plant R2R3 MYB domains participate in REDOX-dependent DNA binding. *The Journal of biological chemistry*, *279*(36), 37878-85.

Hendrickson, W. (1985). Protein-DNA interactions studied by the gel electrophoresis-DNA binding assay. *BioTechniques*, 3, 198-207.

Hernandez, J. M., Feller, A., Morohashi, K., Frame, K., & Grotewold, Erich. (2007). The basic helix loop helix domain of maize R links transcriptional regulation and histone modifications by recruitment of an EMSY-related factor. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(43), 17222-17227.

Hewitt, L., & McDonnell, J. M. (2004). Screening and optimizing protein production in *E. coli*. *Methods In Molecular Biology*, *278*, 1-16.

Higo, K., Ugawa, Y., Iwamoto, M., & Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Research*, *27*(1), 297-300.

Higuchi, R., Dollinger, G., Walsh, P. S., & Griffith, R. (1992). Simultaneous amplification and detection of specific DNA sequences. *Biotechnology, 10*(8), 413-417.

Higuchi, R., Fockler, C., Dollinger, G., & Watson, R. (1993). Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology*, *11*(9), 1026-1030.

Hiratsu, K, Matsui, K, Koyama, T., & Ohme-Takagi, M. (2003). Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in *Arabidopsis*. *The Plant Journal*, *34*(5), 733-739.

Hiratsu, Keiichiro, Mitsuda, N., Matsui, Kyoko, & Ohme-Takagi, Masaru. (2004). Identification of the minimal repression domain of SUPERMAN shows that the DLELRL hexapeptide is both necessary and sufficient for repression of transcription in *Arabidopsis*. *Biochemical and Biophysical Research Communications*, *321*(1), 172-178.

Hodal, L., Bochardt, A., Nielsen, J.E., Mattsson, O., & Okkels, F.T. (1992). Detection, expression and specific elimination of endogenous β-glucuronidase activity in transgenic and non-transgenic plants. *Plant Sci*., 8, 115-122.

Hong, S. H., Kim, H. J., Ryu, J. S., Choi, H., Jeong, S., Shin, J., Choi, G., et al. (2008). CRY1 inhibits COP1-mediated degradation of BIT1, a MYB transcription factor, to activate blue light-dependent gene expression in *Arabidopsis*. *The Plant Journal*, *55*(3), 361-371.

Horch, R., Fry, J., Hoffmann, N., Eichholtz, P., Rogers, R., Fratey, T. (1985) A simple method for transferring genes into plants. *Science,* 227, 1229–1231.

Hösel, W., Surholt, E., & Borgmann, E. (1978). Characterization of beta-glucosidase isoenzymes possibly involved in lignification from chick pea (*Cicer arietinum* L.) cell suspension cultures. *The Federation of European Biochemical Societies Journal*, *84*(2), 487-492.

Hösel, W., and Todenhagen, R. (1980). Characterization of a beta-glucosidase from *Glycine max* which hydrolyses coniferin and syringin. *Phytochemistry, 19*, 1349-1353.

Hu, C.-Y., Chee, P. P., Chesney, R. H., Zhou, J. H., Miller, P. D., & Brien, W. T. O. (1990). Intrinsic GUS-like activities in seed plants. *Plant Cel Reports, 9,* 1-5.

Hu, W. J., Harding, S. A., Lung, J., Popko, J. L., Ralph, J, Stokke, D. D., Tsai, C. J., et al. (1999). Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. *Nature Biotechnology*, *17*(8), 808-812.

Hughes, C. E. (1998). *Monograph of Leucaena (Leguminosae-Mimosoideae)*. Systematic botany monographs v. 55.

Hui, E. K.-W., Wang, P.-C., & Lo, S. J. (2002). PCR-based strategies to clone unknown DNA regions from known foreign integrants. An overview. *Methods In Molecular Biology*, *192*, 249-274.

Humphreys, J. M., and Chapple, C. (2002). Rewriting the lignin roadmap. *Curr. Opin. Plant Biol. 5*, 224-229.

Ibrahim, R.K. (1977). Glucosylation of lignin precursors by uridine diphosphate glucose: coniferyl alcohol glucosyltransferase in higher plants. *Z. Pflanze n physiol.*, *85*, 253-262.

Israelsson, M., Eriksson, M. E., Hertzberg, M., Aspeborg, H., Nilsson, P., & Moritz, Thomas. (2003). Changes in gene expression in the wood-forming tissue of transgenic hybrid aspen with increased secondary growth. *Plant Molecular Biology*, *52*(4), 893-903.

Ito, M. (2005). Conservation and diversification of three-repeat MYB transcription factors in plants. *Journal of Plant Research*, *118*(1), 61-69.

Jana, S., & Deb, J. K. (2005). Strategies for efficient production of heterologous proteins in *Escherichia coli*. *Applied microbiology and biotechnology*, *67*(3), 289-98.

Jauch, R., Yeo, H. C., Kolatkar, P. R., & Clarke, N. D. (2007). Assessment of CASP7 structure predictions for template free targets. *Proteins*, *69* (8), 57-67.

Jefferson, R. A. (1987). Assaying chimeric genes in plants: The GUS gene fusion system. *Plant Molecular Biology Reporter*, *5*(4), 387-405

Jiang, C., Gu, J., Chopra, Surinder, Gu, X., & Peterson, Thomas. (2004). Ordered origin of the typical two- and three-repeat MYB genes. *Gene*, *326*(12), 13-22.

Jiang, C., Gu, X., & Peterson, Thomas. (2004). Identification of conserved gene structures and carboxy-terminal motifs in the MYB gene family of *Arabidopsis* and *Oryza sativa* L. ssp. indica. *Genome Biology*, *5*(7), 40-46.

Jin, Hailing, Cominelli, E., Bailey, P., Parr, Adrian, Mehrtens, F., Jones, J., Tonelli, Chiara, et al. (2000). Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in *Arabidopsis*. *The European Molecular Biology Organization Journal*, *19*(22), 6150-6161.

Jones, D. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*, *292*(2), 195-202.

Jones, R.J. (1979).The value of *Leucaena leucocephala* feed for ruminants in the tropics. *World Animal Review*. *32*(10), 10-21.

Kacser, H., & Burns, J. A. (1973). The control of flux. *Symposia of the Society for Experimental Biology*, *27*(2), 65-104.

Kaneda, M., Rensing, Kim H, Wong, J. C. T., Banno, B., Mansfield, Shawn D, & Samuels, A Lacey. (2008). Tracking monolignols during wood development in lodgepole pine. *Plant Physiology*, *147*(4), 1750-1760.

Kang, Y. H., Kirik, V., Hulskamp, M., Nam, K. H., Hagely, K., Lee, M. M., & Schiefelbein, J. (2009). The MYB23 gene provides a positive feedback loop for cell fate specification in the *Arabidopsis* root epidermis. *The Plant Cell*, *21*(4), 1080-1094.

Karpinska, B., Karlsson, M., Srivastava, M., Stenberg, A., Schrader, J., Sterky, F., Bhalerao, R., et al. (2004). MYB transcription factors are differentially expressed and regulated during secondary vascular tissue development in hybrid aspen. *Plant Molecular Biology*, *56*(2), 255-270.

Kawaoka, A., Kaothien, P., Yoshida, K., Endo, S., Yamada, K., & Ebinuma, H. (2000). Functional analysis of tobacco LIM protein Ntlim1 involved in lignin biosynthesis. *The Plant Journal*, *22*(4), 289-301.

Kazan, K. (2006). Negative regulation of defence and stress genes by EAR-motif-containing repressors. *Trends in Plant Science*, *11*(3), 109-112.

Keller, T., Abbott, J., Moritz, Thomas, & Doerner, P. (2006). *Arabidopsis* regulators of axillary meristem controls a leaf axil stem cell niche and modulates vegetative development. *The Plant Cell*, *18*(3), 598-611.

Kimmel, A. R., & Berger, S. L. (1987). Preparation of cDNA and the generation of cDNA libraries: overview. *Methods in Enzymology*, *152*, 307-316.

Kirik, V., Lee, M. M., Wester, K., Herrmann, U., Zheng, Z., Oppenheimer, D., Schiefelbein, J., et al. (2005). Functional diversification of MYB23 and GL1 genes in trichome morphogenesis and initiation. *Development Cambridge England*, *132*(7), 1477-1485.

Klempnauer, K. H., Gonda, T. J., & Bishop, J. M. (1982). Nucleotide sequence of the retroviral leukemia gene v-myb and its cellular progenitor c-MYB: the architecture of a transduced oncogene. *Cell*, *31*(2 Pt 1), 453-463.

Kobayashi, S., Ishimaru, M., Hiraoka, K., & Honda, C. (2002). MYB-related genes of the Kyoho grape (*Vitis labruscana*) regulate anthocyanin biosynthesis. *Planta*, *215*(6), 924-933.

Komamine, A. (1980). Establishment of an experimental system for the study of tracheary element differentiation from single cells isolated from the mesophyll of *Zinnia elegans*. *Plant Physiology*, *10*(1), 57-60.

Kotewicz, M.L., D'Alessio, J.M., Driftmier, K.M., Blodgett, K.P., & Gerard, G.F. (1985). Cloning and overexpression of Moloney murine leukemia virus reverse transcriptase in *Escherichia coli. Gene*, 35, 249-255.

Kranz, H. D., Denekamp, M, Greco, R., Jin, H, Leyva, A., Meissner, R. C., Petroni, K., et al. (1998). Towards functional characterisation of the members of the R2R3-MYB gene family from *Arabidopsis thaliana*. *The Plant Journal*, *16*(2), 263-276.

Kranz, H., Scholz, K., & Weisshaar, B. (2000). c-MYB oncogene-like genes encoding three MYB repeats occur in all major plant lineages. *The Plant Journal*, *21*(2), 231-5.

Krishnan, S., & Lalitha, K. (1990). Interactive metabolic regulations during biomethanation of *Leucaena leucocephala*. *Appl Biochem Biotechnol*. *26*, 7384-7390.

Kwok, S., Kellogg, D. E., McKinney, N., Spasic, D., Goda, L., Levenson, C., & Sninsky, J. J. (1990). Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Research*, *18*(4), 999-1005.

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, *157*(1), 105-132.

Lacombe, E, Van Doorsselaere, J., Boerjan, W, Boudet, A. M., & Grima-Pettenati, J. (2000). Characterization of cis-elements required for vascular expression of the cinnamoyl CoA reductase gene and for protein-DNA complex formation. *The Plant Journal*, *23*(5), 663-676.

Laemmli, U. K. (1970). Cleavage of structural proteins during assembly of head of bacteriophage-T4. *Nature*, *227*(5259), 680-690

Lai, L. B., Nadeau, J. A., Lucas, J., Lee, E.-K., Nakagawa, T., Zhao, L., Geisler, M., et al. (2005). The *Arabidopsis* R2R3 MYB proteins four lips and MYB88 restrict divisions late in the stomatal cell lineage. *The Plant Cell*, *17*(10), 2754-2767.

Lanot, A., Hodge, D., Jackson, R. G., George, G. L., Elias, L., Lim, E.-K., Vaistij, F. E., et al. (2006). The glucosyltransferase UGT72E2 is responsible for monolignol 4-O-glucoside production in *Arabidopsis thaliana*. *The Plant Journal*, *48*(2), 286-295.

Lapierre, Catherine, Pollet, B., Petit-Conil, M., Toval, G., Romero, J., Pilate, Gilles, Leplé, J.-C., et al. (1999). Structural alterations of lignins in transgenic Poplars with depressed cinnamyl alcohol dehydrogenase or caffeic acid O-methyltransferase activity have an opposite impact on the efficiency of industrial kraft pulping. *Plant Physiology*, *119*(1), 153-164.

Laskowski, R. A., MacArthur, Malcolm W, Moss, D. S., & Thornton, Janet M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*(2), 283-291.

Lauvergeat, Virginie, Rech, P., Jauneau, A., Guez, C., Coutos-Thevenot, P., & Grima-Pettenati, Jacqueline. (2002). The vascular expression pattern directed by the *Eucalyptus gunnii* cinnamyl alcohol dehydrogenase EgCAD2 promoter is conserved among woody and herbaceous plant species. *Plant Molecular Biology*, *50*(3), 497-509.

Lee, L. G., Connell, C. R., & Bloch, W. (1993). Allelic discrimination by nick-translation PCR with fluorgenic probes. *Nucleic Acids Research*, *21*(16), 3761-3766.

Legay, S, Lacombe, E, Goicoechea, M, Briere, C., Mackay, John, Mackay, J, & Se, A. (2007). Molecular characterization of EgMYB1, a putative transcriptional repressor of the lignin biosynthetic pathway. *Plant Science*, *173*(5), 542-549.

Leinhos, V., Udagama-Randeniya, P. V., & Savidge, R. A. (1994). Purification of an acidic coniferin-hydrolysing beta-glucosidase from developing xylem of *Pinus banksiana*. *Phytochemistry*, *37*(2), 311-315.

Lepiniec, L., Debeaujon, I., Routaboul, J.-M., Baudry, A., Pourcel, L., Nesi, N., & Caboche, M. (2006). Genetics and biochemistry of seed flavonoids. *Annual Review of Plant Biology*, *57*(1), 405-430.

Lewis, N. G. (1999). A 20(th) century roller coaster ride: a short account of lignification. *Current Opinion in Plant Biology*, *2*(2), 153-162.

Leyva, A., Liang, X., Pintor-Toro, J. A., Dixon, R A, & Lamb, C. J. (1992). cis-Element combinations determine phenylalanine ammonia-lyase gene tissue-specific expression patterns. *The Plant Cell*, *4*(3), 263-271.

Li, L., Yu, X., Thompson, A., Guo, M., Yoshida, S., Asami, T., Chory, Joanne, et al. (2009). *Arabidopsis* MYB30 is a direct target of BES1 and cooperates with BES1 to regulate brassinosteroid-induced gene expression. *The Plant Journal*, *58*(2), 275-286.

Li, S. F., Milliken, O. N., Pham, H., Seyit, R., Napoli, R., Preston, J., Koltunow, A. M., et al. (2009). The *Arabidopsis* MYB5 transcription factor regulates mucilage synthesis, seed coat development, and trichome morphogenesis. *The Plant Cell*, *21*(1), 72-89.

Li, X., Weng, J.-K., & Chapple, Clint. (2008). Improvement of biomass through lignin modification. *The Plant Journal*, *54*(4), 569-581.

Li, Y., & Zhang, Yang. (2009). REMO: A new protocol to refine full atomic protein models from C-alpha traces by optimizing hydrogen-bonding networks. *Proteins*, *76*(3), 665-676.

Liljegren, S. J., Ditta, G. S., Eshed, Y., Savidge, B., Bowman, J. L., & Yanofsky, M. F. (2000). Shatterproof MADS-box genes control seed dispersal in *Arabidopsis*. *Nature*, *404*(6779), 766-770.

Lippold, F., Sanchez, D. H., Musialak, M., Schlereth, A., Scheible, W.-R., Hincha, D. K., & Udvardi, M. K. (2009). AtMyb41 regulates transcriptional and metabolic responses to osmotic stress in *Arabidopsis*. *Plant Physiology*, *149*(4), 1761-1772.

Lipsick, J. S. (1996). One billion years of MYB. *Oncogene*, *13*(2), 223-235.

Liu, C.-J., Miao, Y.-C., & Zhang, K.-W. (2011). Sequestration and transport of lignin monomeric precursors. *Molecules Basel Switzerland*, *16*(1), 710-727.

Liu, Y. G., & Whittier, R. F. (1995). Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics*, *25*(3), 674-681.

Livak, K. J., Flood, S. J., Marmaro, J., Giusti, W., & Deetz, K. (1995). Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *Pcr Methods And Applications*, *4*(6), 357-362.

Lodhi, M. A., Ye, G.-N., Weeden, N. F., & Reisch, B. I. (1994). A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. *Plant Molecular Biology Reporter*, *12*(1), 6-13.

Logemann, E., Parniske, M., & Hahlbrock, K. (1995). Modes of expression and common structural features of the complete phenylalanine ammonia-lyase gene family in parsley. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(13), 5905-5909.

Lois, R., Dietrich, A., Hahlbrock, K., & Schulz, W. (1989). A phenylalanine ammonia-lyase gene from parsley: structure, regulation and identification of elicitor and light

responsive cis-acting elements. *the The European Molecular Biology Organization Journal*, *8*(6), 1641-1648.

Malik, R.S., Dutt, D., Tyagi, C.H., Jindal, A.K., & Lakharia, L.K.(2004). Morphological, anatomical and chemical characteristics of *Leucaena leucocephala* and its impact on pulp and paper making properties. *J Sci Ind Res*, *63*, 125-130.

Mandaokar, A., & Browse, J. (2009). MYB108 acts together with MYB24 to regulate jasmonate-mediated stamen maturation in *Arabidopsis*. *Plant Physiology*, *149*(2), 851-862.

Marcinowski, S., & Grisebach, H. (1978). Enzymology of lignification. Cell-wall bound beta-glucosidase for coniferin from spruce (*Picea abies*) seedlings. *The Federation of European Biochemical Societies Journal*, *87*(1), 37-44.

Marjamaa, K., Kukkola, E., Lundell, T., Karhunen, P., Saranpää, P., & Fagerstedt, K. V. (2006). Monolignol oxidation by xylem peroxidase isoforms of Norway spruce (*Picea abies*) and silver birch (*Betula pendula*). *Tree Physiology*, *26*(5), 605-611.

Martin, C, & Paz-Ares, J. (1997). MYB transcription factors in plants. *Trends in Genetics*, *13*(2), 67-73.

Maruyama, K., & Sugano, S. (1994). Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene*, *138*(1-2), 171-174.

Matus, J. T., Aquea, F., & Arce-Johnson, P. (2008). Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across Vitis and *Arabidopsis* genomes. *BMC Plant Biology*, *8*(1), 83.

Mele, G., Ori, N., Sato, Y., & Hake, S. (2003). The knotted1-like homeobox gene brevipedicellus regulates cell differentiation by modulating metabolic pathways. *Genes & Development*, *17*(17), 2088-2093.

Mengiste, T., Chen, X., Salmeron, J., & Dietrich, R. (2003). The botrytis susceptible1 gene encodes an R2R3myb transcription factor protein that is required for biotic and abiotic stress responses in *Arabidopsis*. *Plant Physiology*, *15*(11), 2551-2565.

Mitsuda, N., Iwase, A., Yamamoto, H., Yoshida, M., Seki, M., Shinozaki, K., & Ohme-Takagi, Masaru. (2007). NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of *Arabidopsis*. *The Plant Cell*, *19*(1), 270-280.

Mizutani, M., Ohta, D., & Sato, R. (1997). Isolation of a cDNA and a genomic clone encoding cinnamate 4-hydroxylase from *Arabidopsis* and its expression manner in planta. *Plant Physiology*, *113*(3), 755-763.

Moyano, E., Martínez-Garcia, J. F., & Martin, C. (1996). Apparent redundancy in MYB gene function provides gearing for the control of flavonoid biosynthesis in *Antirrhinum* flowers. *The Plant Cell*, *8*(9), 1519-1532.

Mu, R.-L., Cao, Y.-R., Liu, Y.-F., Lei, G., Zou, H.-F., Liao, Y., Wang, H.-W., et al. (2009). An R2R3-type transcription factor gene AtMYB59 regulates root growth and cell cycle progression in *Arabidopsis*. *Cell Research*, *19*(11), 1291-304.

Mullis, K. B., & Faloona, F. A. (1987). Specific synthesis of DNA *in vitro* via a polymerase-catalysed chain reaction. *Methods in Enzymology*, *155*, 335-350.

Murashige, T., & Skoog, F. (1962). A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia Plantarum*, *15*(3), 473-497.

Mähönen, A. P., Bishopp, A., Higuchi, M., Nieminen, K. M., Kinoshita, K., Törmäkangas, K., Ikeda, Y., et al. (2006). Cytokinin signaling and its inhibitor AHP6 regulate cell fate during vascular development. *Science*, *311*(5757), 94-98.

Müller, D., Schmitz, G., & Theres, K. (2006). Blind homologous R2R3 Myb genes control the pattern of lateral meristem initiation in *Arabidopsis*. *The Plant Cell*, *18*(3), 586-597.

Nakashima, J., Mizuno, T., Takabe, K., Fujita, M., Saiki, H. (1997). Direct visualization of lignifying secondary wall thickenings in *Zinna elegans* cells in culture. *Plant Cell Physiol*, *38*, 818-827.

Nei, Masatoshi. (1996). Phylogenetic analysis in molecular evolutionary genetics. *Annual Review of Genetics*, *30*(1), 371-403.

Ness, S. A. (1999). MYB binding proteins: regulators and cohorts in transformation. *Oncogene*, *18*(19), 3039-3046.

Newman, L. J., Perazza, D. E., Juda, L., & Campbell, Malcolm M. (2004). Involvement of the R2R3-MYB, AtMYB61, in the ectopic lignification and dark-photomorphogenic components of the det3 mutant phenotype. *The Plant Journal*, *37*(2), 239-250.

NRC (National Reseach Council) (1984). Nutritional requirements of goats. *National Academy Press. pp250*

Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., Ishii, S., et al. (1992). Solution structure of a DNA-binding unit of Myb: a helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(14), 6428-6432.

Ogata, K., Kanei-Ishii, C., Sasaki, M., Hatanaka, H., Nagadoi, A., Enari, M., Nakamura, H., et al. (1996). The cavity in the hydrophobic core of MYB DNA-binding domain is reserved for DNA recognition and trans-activation. *Nature Structural Biology*, *3*(2), 178-187.

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., et al. (1994). Solution structure of a specific DNA complex of the MYB DNA-binding domain with cooperative recognition helices. *Cell*, *79*(4), 639-648.

Ohl, S., Hedrick, S. A., Chory, J, & Lamb, C. J. (1990). Functional properties of a phenylalanine ammonia-lyase promoter from *Arabidopsis*. *The Plant Cell*, *2*(9), 837-848.

Ohta, M., Matsui, Kyoko, Hiratsu, Keiichiro, Shinshi, H., & Ohme-Takagi, Masaru. (2001). Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. *The Plant Cell*, *13*(8), 1959-1968.

Pabo, C. O., & Sauer, R. T. (1992). Transcription factors: structural families and principles of DNA recognition. *Annual Review of Biochemistry*, *61*, 1053-1095.

Patzlaff, A., McInnis, S., Courtenay, A., Surman, C., Newman, L. J., Smith, Caroline, Bevan, Michael W, et al. (2003). Characterisation of a pine MYB that regulates lignification. *The Plant Journal*, *36*(6), 743-754.

Patzlaff, A., Newman, L. J., Dubos, C., Whetten, R. W., Smith, Caroline, McInnis, S., Bevan, Michael W, et al. (2003). Characterisation of Pt MYB1, an R2R3-MYB from pine xylem. *Plant Molecular Biology*, *53*(4), 597-608.

Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P. A., & Saedler, H. (1987). The regulatory c1 locus of Zea mays encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. *the The European Molecular Biology Organization Journal*, *6*(12), 3553-3558.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605-1612.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic acids research*, *29*(9), 45-48.

Piquemal, J., Lapierre, Catherine, Myton, K., O'connell, A., Schuch, Wolfgang, Grima-Pettenati, Jacqueline, & Boudet, A.-M. (2002). Down-regulation of cinnamoyl-CoA reductase induces significant changes of lignin profiles in transgenic tobacco plants. *The Plant Journal*, *13*(1), 71-83.

Pottinger, A.J., & Hughes, C.E. (1995). A review of wood quality in *Leucaena*. In: *Leucaena* opportunities and limitations: proceedings of a workshop held in Bogor, Indonesia, ACIAR Proceedings, Canberra. pp 98-102.

Prabu, G. R., & D, T. P. (2011). Structure of DNA binding MYB transcription factor protein (sc MYBas1-3) from sugarcane – threading and ab initio. *Methodology*, *3*(3), 77-82.

Puhringer, H., Moll, D., Hoffmann-Sommergruber, K., Watillon, B., Katinger, H., & Machado, M. L. D. (2000). The promoter of an apple Ypr10 gene, encoding the major allergen Mal d 1, is stress- and pathogen-inducible. *Plant Science*,152, 35-50

Punwani, J. A., Rabiger, D. S., Lloyd, A., & Drews, G. N. (2008). The MYB98 subcircuit of the synergid gene regulatory network includes genes directly and indirectly regulated by MYB98. *The Plant Journal*, *55*(3), 406-414.

Puzio, P. S., Lausen, J., Heinen, P., & Grundler, F. M. W. (2000). Promoter analysis of pyk20, a gene from *Arabidopsis thaliana*. *Plant Science*, *157*(2), 245-255.

Quaedvlieg, N., Dockx, J., Keultjes, G., Kock, P., Wilmering, J., Weisbeek, P., & Smeekens, S. (1996). Identification of a light-regulated MYB gene from an

Arabidopsis transcription factor gene collection. *Plant Molecular Biology*, *32*(5), 987-993.

Quattrocchio, F., Verweij, W., Kroon, A., Spelt, C., Mol, J., & Koes, R. (2006). PH4 of *Petunia* is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway. *Plant Cell*, *18*(5), 1274-1291.

Rabinowicz, P. D., Braun, E L, Wolfe, A. D., Bowen, B., & Grotewold, E. (1999). Maize R2R3 Myb genes: Sequence analysis reveals amplification in the higher plants. *Genetics*, *153*(1), 427-444.

Raes, J., Rohde, A., Christensen, J. H., Van De Peer, Y., & Boerjan, Wout. (2003). Genome-wide characterization of the lignification toolbox in *Arabidopsis*. *Plant Physiology*, *133*(3), 1051-1071.

Raeymaekers, L. (2000). Basic principles of quantitative PCR. *Molecular Biotechnology*, *15*(2), 115-122.

Raffaele, S., Vailleau, F., Léger, A., Joubès, J., Miersch, O., Huard, C., Blée, E., et al. (2008). A MYB transcription factor regulates very-long-chain fatty acid biosynthesis for activation of the hypersensitive cell death response in *Arabidopsis*. *The Plant Cell*, *20*(3), 752-767.

Ralph, J. (2006). What makes a good monolignol substitute? In: *The Science and Lore of the Plant Cell Wall: Biosynthesis, Structure and Function*, pp 285–293.

Ralph, J, Lundquist, K., Brunow, G., Lu, F., Kim, H., Schatz, P. F., Marita, Jane M, et al. (2004). Lignins: Natural polymers from oxidative coupling of 4-hydroxyphenyl-propanoids. *Phytochemistry Reviews*, *3*(1), 29-60.

Ramsay, N. A., & Glover, B. J. (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends in Plant Science*, *10*(2), 63-70.

Ramsay, R. G., Ishii, S., & Gonda, T. J. (1992). Interaction of the MYB protein with specific DNA binding sites. *The Journal of Biological Chemistry*, *267*(8), 5656-5662.

Ranocha, P., Chabannes, M., Chamayou, S., Danoun, S., Jauneau, A., Boudet, A.-M., & Goffner, D. (2002). Laccase down-regulation causes alterations in phenolic metabolism and cell wall structure in Poplar. *Plant Physiology*, *129*(1), 145-155.

Rao, Y.V. (1984). *Leucaena* plantations: a farming experience. *Leucaena Res Rep*, *5*, 489-494.

Raven, P.H., Evert, R.F., & Eichhorn, S.E.(1999). *Biology of plants*, 6th ed. New York, NY: 139-156.

Reale, S., Di Tullio, A., Spreti, N., & De Angelis, F. (2010). Mass spectrometry in the biosynthetic and structural investigation of lignins. *Mass Spectrometry Reviews*, *23*(2), 87-126.

Reischl, U., Wittwer, C.T., & Cockerill, F. (2002). Rapid cycle real-time PCR: Methods and Applications. New York: Springer-Verlag, pp 45

Relwani, L.L., Rangnekar, D.V. (1981). Forests for fuel- social forestry-the second green revolution. *Sci Today*, *15*, 48-50.

Revzin, A. (1989). Gel electrophoresis assays for DNA-protein interactions. *Biotechniques*, *7*(4), 346-355.

Riechmann, J. L. (2000). *Arabidopsis* transcription factors: Genome-wide comparative analysis among eukaryotes. *Science*, *290*(5499), 2105-2110.

Rogers, L. A., & Campbell, Malcolm M. (2004). The genetic control of lignin deposition during plant growth and development. *New Phytologist*, *164*(1), 17-30.

Rombauts, S., Déhais, P., Van Montagu, M., & Rouzé, P. (1999). PlantCARE, a plant cis-acting regulatory element database. *Nucleic Acids Research*, *27*(1), 295-296.

Romero, I., Fuertes, A., Benito, M. J., Malpica, J. M., Leyva, A., & Paz-Ares, J. (1998). More than 80 R2R3-MYB regulatory genes in the genome of *Arabidopsis thaliana*. *The Plant Journal*, *14*(3), 273-284.

Rosinski, J. A., & Atchley, W. R. (1998). Molecular evolution of the MYB family of transcription factors: evidence for polyphyletic origin. *Journal of Molecular Evolution*, *46*(1), 74-83.

Roy, A., Kucukural, A., & Zhang, Yang. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols*, *5*(4), 725-38.

Russell, D.A., & Fromm, M.E. (1995). Biolistic plant transformation: Transient gene expression studies with the biolistic system. *Gene transfer to plants*. pp 118-126.

Sablowski, R. W., Moyano, E., Culianez-Macia, F. A., Schuch, W, Martin, C, & Bevan, M. (1994). A flower-specific MYB protein activates transcription of phenylpropanoid biosynthetic genes. *The European Molecular Biology Organization Journal*, *13*(1), 128-137.

Saha, B. K., Tian, B., & Bucy, R. P. (2001). Quantitation of HIV-1 by real-time PCR with a unique fluorogenic probe. *Journal of Virological Methods*, *93*(1-2), 33-42.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406-425.

Sambrook, J., & Fritsch, E. F. (1989). *Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press Cold Spring Harbor NY.*

Samuels, A L, Rensing, K H, Douglas, C J, Mansfield, S D, Dharmawardhana, D. P., & Ellis, B E. (2002). Cellular machinery of wood production: differentiation of secondary xylem in *Pinus contorta* var. latifolia. *Planta*, *216*(1), 72-82.

Schaffer, R., Landgraf, J., Accerbi, M., Simon, V., Larson, M., & Wisman, E. (2001). Microarray analysis of diurnal and circadian-regulated genes in *Arabidopsis*. *The Plant Cell*, *13*(1), 113-123.

Scharf, S. J., Horn, G. T., & Erlich, H. A. (1986). Direct cloning and sequence analysis of enzymatically amplified genomic sequences. *Science*, *233*(4768), 1076-1078.

Schumacher, K., Vafeados, D., McCarthy, M., Sze, H., Wilkins, T., & Chory, Joanne. (1999). The *Arabidopsis* det3 mutant reveals a central role for the vacuolar H+–

ATPase in plant growth and development. *Genes & Development*, *13*(24), 3259-3270.

Schwechheimer, C., Zourelidou, M., & Bevan, M W. (1998). Plant transcription factor studies. *Annual Review of Plant Physiology and Plant Molecular Biology*, *49*(1), 127-150.

Seo, P. J., & Park, C.-M. (2010). MYB96-mediated abscisic acid signals induce pathogen resistance response by promoting salicylic acid biosynthesis in *Arabidopsis*. *New Phytologist*, *186*(2), 471-483.

Seo, P. J., Xiang, F., Qiao, M., Park, J.-Y., Lee, Y. N., Kim, S.-G., Lee, Y.-H., et al. (2009). The MYB96 transcription factor mediates abscisic acid signaling during drought stress response in *Arabidopsis*. *Plant Physiology*, *151*(1), 275-289.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology*, *98*(3), 503-517.

Speer, E. O. (1987). A method of retaining phloroglucinol proof of lignin. *Stain Technology*, *62*(4), 279-280.

Srivastava, S. (2008). Molecular characterization of cinnamoyl CoA reductase (CCR) gene in *Leucaena leucocephala., PhD thesis,* University of Pune, NCL.

Steeves, V., Förster, H., Pommer, U., & Savidge, R. (2001). Coniferyl alcohol metabolism in conifers. Glucosidic turnover of cinnamyl aldehydes by UDPG: coniferyl alcohol glucosyltransferase from pine cambium. *Phytochemistry*, *57*(7), 1085-1093.

Stewart, J. J., Akiyama, T., Chapple, Clint, Ralph, John, & Mansfield, Shawn D. (2009). The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid poplar. *Plant Physiology*, *150*(2), 621-635.

Stracke, R, Werber, M., & Weisshaar, Bernd. (2001). The R2R3-MYB gene family in *Arabidopsis thaliana*. *Current Opinion in Plant Biology*, *4*(5), 447-456.

Stracke, Ralf, Ishihara, H., Huep, G., Barsch, A., Mehrtens, F., Niehaus, K., & Weisshaar, Bernd. (2007). Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *The Plant Journal*, *50*(4), 660-677.

Suo, J. (2003). Identification of GhMYB109 encoding a R2R3 MYB transcription factor that expressed specifically in fiber initials and elongating fibers of cotton (*Gossypium hirsutum* L.). *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, *1630*(1), 25-34.

Séguin, a, Laible, G., Leyva, a, Dixon, R. a, & Lamb, C. J. (1997). Characterization of a gene encoding a DNA-binding protein that interacts *in vitro* with vascular specific cis elements of the phenylalanine ammonia-lyase promoter. *Plant molecular biology*, *35*(3), 281-91.

Takabe, K., Fujita, M., Harada, H., & Saiki, H. (1985). Autoradiographic investigations of lignification in the cell walls of cryptomeria (*Cryptomeria japonica d*. Don). *Mokuzai Gakkaishi*, *31*, 613-619.

Tamagnone, L., Merida, A., Parr, A, Mackay, S., Culianez-Macia, F., Roberts, K., & Martin, C. (1998). The AmMYB308 and AmMYB330 transcription factors from *Antirrhinum* regulate phenylpropanoid and lignin biosynthesis in transgenic tobacco. *The Plant Cell*, *10*(2), 135-154.

Tamura, K., Dudley, J., Nei, Masatoshi, & Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, *24*(8), 1596-1599.

Tan, W., Wang, K., & Drake, T. J. (2004). Molecular beacons. *Current Opinion in Chemical Biology*, *8*(5), 547-553.

Terry, C. F., Shanahan, D. J., Ballam, L. D., Harris, N., McDowell, D. G., & Parkes, H. C. (2002). Real-time detection of genetically modified soya using Lightcycler and ABI 7700 platforms with TaqMan, Scorpion, and SYBR Green I chemistries. *Journal of AOAC International*, *85*(4), 938-944.

Thapar, G.(2009). The increasing importance of India in the global pulp and paper sector, (RISI Indian seminar, New Delhi, December 2009).

Todd, J. A., Bell, J. I., & McDevitt, H. O. (1987). HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature*, *329*(6140), 599-604.

Torane, J.V., Lokhande, C.D., & Pawar, S.H. (1990). Some studies on anaerobic decomposition of *Leucaena leucocephala* leaves. *Energy Sources*, *12*, 25-32.

Uggla, C., Mellerowicz, E. J., & Sundberg, B. (1998). Indole-3-acetic acid controls cambial growth in scots pine by positional signaling. *Plant Physiology*, *117*(1), 113-121.

Uzal, E. N., Gómez Ros, Laura V, Pomar, F., Bernal, M. A., Paradela, A., Albar, J. P., & Ros Barceló, A. (2009). The presence of sinapyl lignin in *Ginkgo biloba* cell cultures changes our views of the evolution of lignin biosynthesis. *Physiologia Plantarum*, *135*(2), 196-213.

Van Der Velden, V. H. J., Hochhaus, A., Cazzaniga, G., Szczepanski, T., Gabert, J., & Van Dongen, J. J. M. (2003). Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects. *Leukemia official journal of the Leukemia Society of America Leukemia Research Fund UK*, *17*(6), 1013-1034.

Vanholme, R., Morreel, K., Ralph, John, & Boerjan, Wout. (2008). Lignin engineering. *Current Opinion in Plant Biology*, *11*(3), 278-285.

Vannini, C., Locatelli, F., Bracale, M., Magnani, E., Marsoni, M., Osnato, M., Mattana, M., et al. (2004). Overexpression of the rice Osmyb4 gene increases chilling and

freezing tolerance of *Arabidopsis thaliana* plants. *The Plant Journal*, *37*(1), 115-127.

Vet, J. A. M., & Marras, S. A. E. (2005). Design and optimization of molecular beacon real-time polymerase chain reaction assays. *Methods In Molecular Biology Clifton Nj*, *288*, 273-290.

De Vos, M., Denekamp, Marten, Dicke, M., Vuylsteke, M., Van Loon, L., Smeekens, S. C., & Pieterse, C. M. (2006). The *Arabidopsis thaliana* transcription factor AtMYB102 functions in defense against the insect herbivore *Pieris rapae*. *Plant signaling behavior*, *1*(6), 305-311.

Wagner, A., Donaldson, L., Kim, H., Phillips, L., Flint, H., Steward, D., Torr, K., et al. (2009). Suppression of 4-coumarate-CoA ligase in the coniferous gymnosperm *Pinus radiata*. *Plant Physiology*, *149*(1), 370-383.

Wang, X., Niu, Q.-W., Teng, C., Li, C., Mu, J., Chua, N.-H., & Zuo, J. (2009). Overexpression of PGA37/MYB118 and MYB115 promotes vegetative-to-embryonic transition in *Arabidopsis*. *Cell Research*, *19*(2), 224-235.

Watson, J. C., & Thompson, W. F. (1986). Purification and restriction endonuclease analysis of plant nuclear DNA. *Methods in Enzymology*, *118*, 57-75.

Whetten, R., & Sederoff, R. (1995). Lignin Biosynthesis. *The Plant Cell*, *7*(7), 1001-1013.

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, *35*, 407-410.

Wilkins, O., Nahal, H., Foong, J., Provart, N. J., & Campbell, Malcolm M. (2009). Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. *Plant Physiology*, *149*(2), 981-993.

Wittwer, C. T., Herrmann, M. G., Moss, A. A., & Rasmussen, R. P. (1997). Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques*, *22*(1), 130-131.

Wittwer, C. T., Ririe, K. M., Andrew, R. V., David, D. A., Gundry, R. A., & Balis, U. J. (1997). The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. *Biotechniques*, *22*(1), 176-181.

Woodger, F. J., Gubler, Frank, Pogson, B. J., & Jacobsen, John V. (2003). A Mak-like kinase is a repressor of GAMYB in barley aleurone. *The Plant Journal*, *33*(4), 707-717.

Xie, Z., Lee, E., Lucas, J. R., Morohashi, K., Li, D., Murray, J. A. H., Sack, F. D., et al. (2010). Regulation of cell proliferation in the stomatal lineage by the *Arabidopsis* MYB four lips via direct targeting of core cell cycle genes. *The Plant Cell*, *22*(7), 2306-2321.

Yamaguchi, M., Kubo, M., Fukuda, H., & Demura, T. (2008). Vascular-related NAC-domain7 is involved in the differentiation of all types of xylem vessels in *Arabidopsis* roots and shoots. *The Plant Journal*, *55*(4), 652-664.

Yang, C., Xu, Z., Song, J., Conner, K., Vizcay Barrena, G., & Wilson, Z. A. (2007). *Arabidopsis* MYB26/male sterile35 regulates secondary thickening in the endothecium and is essential for anther dehiscence. *The Plant Cell*, *19*(2), 534-548.

Yang, S. W., Jang, I.-C., Henriques, R., & Chua, N.-H. (2009). far-red elongated hypocotyl1 and FHY1-like associate with the *Arabidopsis* transcription factors LAF1 and HFR1 to transmit phytochrome A signals for inhibition of hypocotyl elongation. *The Plant Cell*, *21*(5), 1341-1359.

Yang, Y., Li, R., & Qi, M. (2000). In vivo analysis of plant promoters and transcription factors by agroinfiltration of tobacco leaves. *The Plant Journal*, *22*(6), 543-551.

Yanhui, C., Xiaoyuan, Y., Kun, H., Meihua, L., Jigang, L., Zhaofeng, G., Zhiqiang, L., et al. (2006). The MYB transcription factor superfamily of *Arabidopsis*: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Molecular Biology*, *60*(1), 107-124.

Yazaki, K. (2006). ABC transporters involved in the transport of plant secondary metabolites. *FEBS Letters*, *580*(4), 1183-1191.

Ye, Z. H., Kneusel, R. E., Matern, U., & Varner, J. E. (1994). An alternative methylation pathway in lignin biosynthesis in *Zinnia*. *The Plant Cell*, *6*(10), 1427-1439.

Ying, G. G., Proost, P., Van Damme, J., Bruschi, M., Introna, M., & Golay, J. (2000). Nucleolin, a novel partner for the MYB transcription factor family that regulates their activity. *The Journal of Biological Chemistry*, *275*(6), 4152-4158.

Yu, E. Y., Kim, S. E., Kim, J. H., Ko, J. H., Cho, M. H., & Chung, I. K. (2000). Sequence-specific DNA recognition by the MYB-like domain of plant telomeric protein RTBP1. *The Journal of Biological Chemistry*, *275*(31), 24208-24214.

Zdunek-Zastocka, E. (2008). Molecular cloning, characterization and expression analysis of three aldehyde oxidase genes from *Pisum sativum* L. *Plant physiology and biochemistry*, *46*(1), 19-28.

Zhang, J. Z. (2003). Overexpression analysis of plant transcription factors. *Current Opinion in Plant Biology*, *6*(5), 430-440.

Zhang, Y. (2007). Prediction Report Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins*, *69* (8), 108-117.

Zhang, Yang. (2008a). Progress and challenges in protein structure prediction. *Current Opinion in Structural Biology*, *18*(3), 342-348.

Zhang, Yang. (2008b). I-TASSER server for protein 3D structure prediction. *BMC bioinformatics*, *9*, 40-45.

Zhong, R., & Ye, Z.-H. (2007). Regulation of cell wall biosynthesis. *Current Opinion in Plant Biology*, *10*(6), 564-572.

Zhong, R., & Ye, Z.-H. (2009). Transcriptional regulation of lignin biosynthesis. *Plant signaling behavior*, *4*(11), 1028-1034.

Zhong, R., Demura, T., & Ye, Z.-H. (2006). SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of *Arabidopsis*. *The Plant Cell*, *18*(11), 3158-3170.

Zhong, R., Lee, C., & Ye, Z.-H. (2010). Evolutionary conservation of the transcriptional network regulating secondary cell wall biosynthesis. *Trends in Plant Science*, *15*(11), 625-632.

Zhong, R., Lee, C., Zhou, J., McCarthy, R. L., & Ye, Z.-H. (2008). A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in *Arabidopsis*. *The Plant Cell*, *20*(10), 2763-2782.

Zhong, R., Ripperger, A., & Ye, Z.-H. (2000). Ectopic deposition of lignin in the pith of stems of two *Arabidopsis* mutants. *Plant Physiology*, *123*(1), 59-70.

Zhou, H., Pandit, S. B., Lee, S. Y., Borreguero, J., Chen, H., Wroblewska, L., & Skolnick, J. (2007). Analysis of TASSER-based CASP7 protein structure prediction results. *Proteins*, *69* (8), 90-97.

Zhou, J., Lee, C., Zhong, R., & Ye, Z.-H. (2009). MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in *Arabidopsis*. *The Plant Cell*, *21*(1), 248-266.

Zhu, J., Verslues, P. E., Zheng, X., Lee, B.-H., Zhan, X., Manabe, Y., Sokolchik, I., et al. (2005). HOS10 encodes an R2R3-type MYB transcription factor essential for cold acclimation in plants. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(28), 9966-9971.

Zimmermann, I. M., Heim, M. A., Weisshaar, Bernd, & Uhrig, J. F. (2004). Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. *The Plant Journal*, *40*(1), 22-34.

Zuckerkandl, E., & Pauling, L. (1965). Evolutionary divergence and convergence in proteins. *Analysis*, *97*, 97-166.

PUBLICATIONS

Books/ Research Papers published/communicated

- Bashir M. Khan, Shuban K. Rawal, Manish Arha, Sushim K. Gupta, Sameer Srivastava, Noor M. Shaik, Arun K. Yadav, Pallavi S. Kulkarni, Abhilash O. U., Santosh Kumar, **Sumita Omer** *et al*. (2011) Genetic Engineering of Phenylpropanoid Pathway in *Leucaena leucocephala* (Book chapter accepted for publication *In* genetic Engineering, Intech –Open Access Publishers (http://www.intechweb.org/booksprocess/allchapters/)
- **Sumita Omer**, Santosh Kumar and B. M. Khan. (2011) Isolation and characterization of a novel R2R3 type MYB transcription factor gene from *Leucaena leucocephala* (manuscript under preparation)
- **Sumita Omer** and B. M. Khan. (2011) Comparative *in silico* studies on two novel R2R3 MYB genes from *Leucaena leucocephala* (manuscript under preparation)
- Santosh Kumar, **Sumita Omer**, Krunal Patel and B. M. Khan. (2011) Molecular studies on Cinnamate 4-hydroxylase from *Leucaena leucocephala*: A pulp yielding leguminous tree (manuscript under preparation for Tree Genetics and Genomes)
- Santosh Kumar, **Sumita Omer**, Shruti Chitransh and B. M. Khan. (2011) Downregulation of *Cinnamate 4-Hydroxylase* in transgenic tobacco alters the transcript abundance of other phenylpropanoid pathway gene(s) (manuscript under preparation)

Posters presented

- **Sumita Omer**, Santosh Kumar and B.M. Khan. Studies on Transcriptional Regulation of Lignin Biosynthesis in *Leucaena Leucocephala*. Poster presented in National Symposium and XXXII Annual Meet of Plant Tissue Culture Association (India) held from 4th to 6th February, 2011 at Bikaner, India
- **Sumita Omer**, Santosh Kumar and B.M. Khan. Studies on Regulation of Lignin Biosynthesis pathway gene(s) in *Leucaena Leucocephala*. Poster presented in the area of Biological Sciences in the poster presentation held on 24-25th February,

2010 as part of the National Science Day celebration, 2010 at National Chemical Laboratory, Pune, India

Best poster prizes

- Santosh Kumar, **Sumita Omer** and B.M. Khan. Molecular Studies on Cinnamate 4-Hydroxylase (C4H) and its Down-regulation studies in *Leucaena leucocephala* and tobacco. Best poster prize in the area of Biological Sciences in the poster presentation held on 24-25th February, 2011 as part of the National Science Day and International Year of Chemistry celebrations 2011 at National Chemical Laboratory, Pune, India
- Santosh Kumar, **Sumita Omer** and B.M. Khan. Isolation, Cloning and Expression of three Isoforms of Cinnamate 4-Hydroxylase (*C4H*) and their Down-regulation studies in *Leucaena leucocephala*. Best poster (second prize) award in National Symposium and XXXII Annual Meet of Plant Tissue Culture Association (India) held from 4th to 6th February, 2011 at Bikaner, India