TRANSCRIPTOME ANALYSIS, BIOCHEMICAL CHARACTERIZATION AND TISSUE CULTURE REGENERATION OF *LEUCAENA LEUCOCEPHALA*

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI'I AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

MOLECULAR BIOSCIENCES AND BIOENGINEERING

DECEMBER 2013

By

Dung T. Pham

Thesis Committee:

Dulal Borthakur, Chairperson

Qing X. Li

Harry Ako

Keywords: Transcriptome sequences, mimosine, tissue culture, leucaena

ACKNOWLEDGMENTS

ABSTRACT

Leucaena leucocephala (leucaena) is a leguminous tree, which is adapted well to tropical or subtropical semi-arid environments, highly resistant to both biotic and abiotic stresses, and a protein-rich forage for livestock animals. The long-term goal of this research is to identify leucaena genes for defense to plant pathogens, insects-pests and tolerance to environmental stresses, including drought and salinity. The specific objectives of this project were: (i) transcriptome analysis of genes expressing in shoots and roots of leucaena, (ii) biochemical characterization of transgenic leucaena expressing a bacterial dioxygenase-hydrolyase fusion gene (*pydA/pydB*) to reduce mimosine content of its foliage, and (iii) tissue culture regeneration and multiplication of transgenic leucaena plants. To explore the molecular basis of leucaena's drought tolerance, insect and pathogen resistance, and mimosine biosynthesis, total RNA was extracted from the shoots and roots of three-month-old plants and sent to SeqWright Inc. for transcriptome sequencing. A total of 1,022,583 and 1,165,136 scaffolds were obtained from the transcriptome sequences of shoots and roots of leucaena, respectively. The numbers of contigs obtained were 1,047,350 (in shoots) and 1,190,291 (in roots). All the transcriptome analyses of leucaena in this project were based on 199,818 (in shoots) and 112,091 (in roots) scaffolds, which showed similarities with gene sequences in the NCBI database. Among these, 35,177 and 4,745 scaffolds, that were larger than 500 bp, were used for further analysis of the shoot and root sequences. A total of 33 root sequences were identified that were absent in shoots. Several classes of potential genes for resistance to both biotic and abiotic stresses were identified in the transcriptome sequences. A total of 74 and 160 chitinase sequences were found in the root and shoot transcriptomes, respectively. As expected, a large number of disease resistance genes, encoded by NB-LRR type of genes, were found in the leucaena transcriptome. In the shoot transcriptome, the number of NB-LRR sequences (>500 bp) was 86; among these 36 NB-LRR sequences were >1.0 kb. In contrast, only 18 NB-LRR sequences (>500 bp) were identified in the root transcriptome. A large number of WRKY transcription factors, some of which may be involved in disease resistance, were also identified in the root and shoot transcriptomes; their numbers were 145 and 223 for roots and shoots, respectively. Similarly, 109 sequences encoding different members of ERD (early responses to

dehydration) family of genes were identified. These genes may be involved in drought resistance. For UV tolerance, 11 and 3 gene sequences (>500 bp) were found in the shoot and root transcriptomes, respectively. A total of 636 sequences (>500 bp) encoding different types of Ser/Thr kinases, including 29 sequences showing high similarities to receptor Ser/Thr kinase, were identified. Furthermore, 33 sequences (>500 bp) encoding histidine kinases, 22 sequences (>500 bp) encoding different types of tyrosine kinases as well as receptor tyrosine kinases, 21 sequences (>500 bp) encoding various types of PTPs, 18 MAPK sequences were identified. Thirty sequences (>100 bp), 23 from roots and 7 from shoots, encoding cysteine/mimosine synthase were also identified.

In spite of having many desirable attributes, leucaena contains a toxic non-protein amino acid mimosine, which is harmful to animals. Our laboratory is developing transgenic leucaena plants expressing a bacterial *pydA*/*pydB* fusion gene under the control of a CaMV 35S promoter. We expected that these plants should contain reduced amounts of mimosine. I had determined mimosine contents of sixteen transgenic plants. All transgenic leucaena plants showed lower mimosine content compared to the wild type leucaena. The transgenic leucaena plants contained 40.47- 95.9% less mimosine than the wild type. Among these, three plants (# 1, 3, 5) were found to contain only small amounts of mimosine (0.06-0.10% of dry weight). Currently, there is no suitable method for tissue culture micropropagation of leucaena. To multiply the transgenic leucaena plants, I improved the protocol of tissue culture. Using shoot tips and nodal segments, four plants were regenerated from the transgenic plant #3.

Identification of many genes for diseases and drought resistance from leucaena through transcriptome analysis has opened new areas of research for the future. The disease and drought resistance genes can be characterized further by isolating full-length cDNA. These genes can be used for developing disease and drought resistant varieties of other crop plants. Biochemical determination of transgenic leucaena plants has established that 3 among 16 transgenic lines contain only small amounts of mimosine. These plants will be valuable as a fodder for farm animals. The micropropagation method for leucaena developed in this research will be helpful for rapid multiplication of transgenic leucaena plants without having to wait for the plants to flower and produce seeds.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Acronym	Descriptions
3H4P	3-hydroxy-4-pyridone
ABA	Abscisic acid
BA	Benzyladenine
BIAs	Benzylisoquinoline alkaloids
BLP	Bacterial leaf pustule
CC	Coiled-coil
CSC	Cysteine synthase complex
ERD	Early responsive dehydration
ESTs	Expressed sequence tags
ETs	Expressed transcripts
HPLC	High performance liquid chromatography
LRR	Leucine-rich repeat
MAP	Mitogen activated protein
NAA	Naphthalene acetic acid
NCBI	National Center for Biotechnology Information
NB	Nucleotide-binding
PAMPS	Pathogen-associated molecular patterns
PTKs	Protein tyrosine kinases
SAM	Shoot apical meristem
SNP	Single nucleotide polymorphisms
SSH	Suppression subtractive hybridization
TSSs	Transcription start site
WRKY	zinc-figure transcription factor

Chapter 1

INTRODUCTION

Leucaena leucocephela (Lam.) de Wit (well known as leucaena) is a tropical tree-legume that is highly resistant to various biotic and abiotic stress conditions. Leucaena originated in Central America and is now found in all tropical and subtropical areas of the world (Brewbaker and Sorensson 1994; Shelton and Brewbaker 1994). It can grow and survive in poor nutrient soils and low rainfall areas where most other trees cannot grow. It is an evergreen plant that can grow in the places with high wind, low rainfall or prolonged drought. The leaves and leaflets of leucaena fold in darkness, low moisture or cool temperature (National Academy of Sciences report 1977). Leucaena is a tetraploid species with the gametic chromosome number n=52 (Pan and Brewbaker 1988). There are also diploid *Leucaena* species with 2n chromosome number of 52. *L. leucocephala* is an allotetraploid with the basic chromosome numbers of x = 26 (Harris et al. 1994; Hartman et al. 2000). Each leucaena plant produces a lot of flowers and pods and each pod contains 20-27 seeds (variety K636). Because of its evergreen nature, leucaena flowers throughout the year, although the best season of flowering is from May to August when it produces the best quality of seeds. Because of its profuse seed production capacity, leucaena can produce many seedlings, which help it to spread easily in wild habitats.

Leucaena is a perennial and multipurpose tree, which is widely used for agroforestry. Leucaena is a food resource for human in Africa and South East Asia and Central America (Shelton and Brewbaker 1994). People use young leaves and green pods as vegetables because of high protein content (15-18%) in young foliage (Soedarjo and Borthakur 1998). Herbal tea has been manufactured from leucaena leaves in Okinawa, Japan since 1986 (Tawata et al 2007). Moreover, leucaena can provide huge amount of foliage in short time; so its young foliage can be harvested as forage several times a year. In addition, because of its deep root system and its ability to grow on slopes, leucaena is an ideal plant species for preventing soil erosion and improving quality of soils (Tawata et al 2007). Leucaena forms nitrogen-fixing symbiosis with strains of *Rhizobium tropici* (Martinez et al 1991) and *Rhizobium* spp. such as *Rhizobium* sp. strain TAL 145 (George et al 1994). Therefore, leucaena can grow successfully even on relatively poor soils. Leucaena is also grown as a shade tree for growing some plant

species that do not require too much direct sunlight (Shelton and Brewbaker 1994). It is also known as a 'miracle tree' because of its worldwide success as a long-lived and highly nutritious forage tree, and its great variety of uses.

Leucaena has high level of resistance to various biotic and abiotic stresses. It has no known diseases and is highly tolerant to drought. It is also resistant to most common insect pests, although it is susceptible to psyllid (Brewbaker and Sorensson 1994) and pod borer (personal observation). Its most outstanding qualities include deep-root system and its ability to withstand drought. Even in a place like the Diamondhead crater near Waikiki, Honolulu, where the rainfall is very low and therefore trees usually cannot be grown without irrigation, leucaena grows and survives. The leucaena foliage is contains 15-18% of protein (Soedarjo and Borthakur 1998) because of which it is called the alfalfa of the tropics. However, leucaena also contains, a toxic non-protein amino acid mimosine, which is harmful to animals when fed to them in high quantities (Jones 1979).

At the University of Hawaii at Manoa, Dr. Brewbaker and his students had developed so many varieties of leucaena through selection and cross-breeding between different cultivars. Some cultivars such as K636, K18 had been grown extensity in Hawaii, Africa, Australia, India, Srilanka, Nepal and other places. They also developed several triploid varieties by crossing between diploids and tetraploids *Leucaena* spp. The triploid varieties have chromosome number 78 or 80 and are sterile and seedless. Therefore, they need to be multiplied through vegetative propagation. Recently, transgenic methods have also been used for improvement for leucaena. Jube and Borthakur (2009, 2010) developed *Agrobacterium*-mediated transformation protocol for transferring transgenes into leucaena. They have created transgenic leucaena plants expressing *pydA* or *pydB* genes from *Rhizobium* sp. strain TAL1145. Transgenic leucaena expressing the *pydA* gene contained 1.19% mimosine compared to 2.22% mimosine present in the non-transformed plants (Jube and Borthakur 2010).

In recent years, plant molecular biologists have been making efforts to develop new crop varieties with high tolerance to diseases, insects and environmental stresses. Therefore, plant biologists are looking for new genes for resistance to diseases, insect pests and environmental stresses. The miracle tree leucaena is a potential candidate as a reservoir for new

genes for resistance to diseases, insect pests, and drought stress. Earlier, Negi et al. (2011) applied interspecific subtractive hybridization method to isolate genes from leucaena that are either expressed at high levels or are specific for leucaena, with the hope that some genes for stress tolerance would be also selected. Among the 406 cDNA clones that were selected at the end of the interspecies subtractive hybridization, a few were identified as potential genes for stress tolerance (Negi et al. 2011). However, for a more complete analysis of resistance genes in leucaena, other methods such as genome sequences, transcriptome sequencing and microarray analysis will be required.

The major goal of this project is to identify genes for various biotic and abiotic stresses for leucaena through transcriptome analysis. Recently, our lab has developed several new lines of transgenic leucaena expressing a dioxygenase-hydrogenase (*pyd*A-*pyd*B) fusion gene. It is essential to characterize these plants biochemically and determine the amounts of mimosine and 3-hydroxy-4-pyridone (3H4P), which is also known to be harmful to animals. Moreover, it is required to make several copies of each transgenic plant through vegetative propagation so that physiological and agronomical characteristics can be studied. Currently, there is no suitable tissue culture method for micropropagation of leucaena in vitro. Therefore, we want to develop a tissue culture regeneration protocol for the transgenic leucaena plants.

The specific objectives of the project are:

(1) Transcriptome analysis of leucaena to identify resistance genes for biotic and abiotic stresses;

(2) Biochemical analysis of transgenic leucaena lines to determine mimosine and 3H4P contents; and

(3) Regeneration of transgenic leucaena plants by micropropagation.

References

Brewbaker J, and Sorenson C (1994). Domestication of lesser-known species of the genus Leucaena. "Tropical Trees: The potential for Domestication and Rebuilding of Forest Resources"; Institute of Terrestrial Ecology Symposium N0. 29:195-204

George M, Young J, and Borthakur D (1994). Genetic characterization of *Rhizobium* sp. strain TAL1145 that nodulates tree legumes. Canadian J. Microbiol. 40: 208-215

Harris S, Hughes R, Ingram R, and Abbott R (1994). A phylogenetic analysis of Leucaena (Leguminosae: Mimosoideae). Plant Systematics and Evolution 191:1-26

Hartman T, Jones J, Blackhall N, Power J, Cocking E, and Davey M (2000). Cytogenetics, molecular cytogenetics and genome size in *Leucaena* (Leguminosae, Mimosoideae). Cytogenetic studies of forest trees and shrubs: review, present status, and outlook on the future: Proceedings of the Second IUFRO Cytogenetics Working Party S2. 04.08 Symposium, September 6-12, 1998, Graz, Austria.

Jones R (1979). The value of *leucaena leucocephala* as a feed for ruminants in the tropics. World Animal Review 31:13-23

Jube S, and Borthakur D (2010). Transgenic *Leucaena leucocephala* expressing the *Rhizobium* gene *pyd*A encoding a meta-cleavage dioxygenase shows reduced mimosine content. Plant Physiology and Biochemistry, 48(4): 273-278

Jube S, and Borthakur D (2009). Development of an *Agrobacterium*-mediated transformation protocol for the recalcitrant tree-legume *Leucaena leucocephala* using immature zygotic embryos. Plant Cell, Tissue and Organ Culture (PCTOC): Journal of Plant Biotechnology 96:325-333

Martínez-Romero E, Segovia L, Mercante F, Franco A, Gra-ham P, and Pardo M (1991). *Rhizobium tropici*, a novel speciesnodulating *Phaseolus vulgaris L.* beans and *Leucaena sp.* trees. Int J Syst Bacteriol 41:417–426

Negi V, Pal A, Singh R, and Borthakur D (2011). Identification of species-specific genes from *Leucaena leucocephala* using interspecies suppression subtractive hybridization. Annals of Applied Biology,159(3):387–398

Pan F, and Brewbaker J (1988). Cytological studies in the genus *Leucaena* Benth. Cytologia, 53(2):393-399

Report of study (1977). Leucaena promising forage and tree crop for the tropics. National Academy of Sciences Washington D.C.

Shelton H, and Brewbaker J (1994). *Leucaena leucocephala* – the most widely used forage tree legume. CAB international:15-29

Soedarjo M, and Borthakur D (1998). Mimosine, a toxin produced by the tree-legume *Leucaena* provides a nodulation competition advantage to mimosine-degrading *Rhizobium* strains. Soil Biol. Biochem, 30:1605-1613

Tawata S, Fukuta M, Xuan T, and Deba F (2008). Total utilization of tropical plants *Leucaena leucocephala* and *Alpinia zerumbet*. J. Pestic. Sci., 33(1):40-43

Chapter 2

LITERATURE REVIEW

The transcriptome sequencing provides sequences of transcripts from a tissue, which include protein-coding mRNAs and non-coding small RNAs such as rRNA, tRNA, miRNA. In this method, mRNA is separated from the total RNA using oligo dT column by capturing the molecules through the polyadenylated tails. The mRNA pool is converted to cDNA, fragmented into sizes of 80-300 bp and then sequenced by a high throughput sequencing method such as Illumina sequencing and 454 pyrosequencing. Illumina sequencing generates sequences of 80-100 bp from each cDNA fragments. Sequencing the fragments of the cDNA produces a large number of overlapping sequences, which can be combined into contigs using computer programs. The sequences of individual contigs represent separate transcripts in the tissue. Transcriptome sequencing methods and applications have been discussed in several recent papers (Wang and Brutnell 2010, Egan et al. 2012, Kliebenstein 2012, Martin et al. 2013). Although transcriptome analysis can determine the activity of genes by measuring the amount of transcript molecules in a tissue or a group of cells, it cannot measure the spatial organization of transcript molecules within a single cell. Biologists at the University of Zurich, Switzerland have recently discovered a technique to visualize the activity of genes in single cells (Battich et al. 2013). This new technique, known as 'Image-based transcriptomics' is so efficient that, for the first time, a thousand genes can be studied in parallel in ten thousand single human cells. The technique shows that the activities of genes, and the spatial organization of the resulting transcript molecules, strongly vary between single cells. In the near future, plant biologists will also use this new method to study spatial organization of transcripts within plant cells. So far, transcriptome sequencing has been reported for many plants. One major objective of my research is transcriptome analysis of leucaena using Illumina2000 high throughput sequencing. Illumina and 454 pyrosequencing have been applied to many other plant species. Therefore, it is important to review the literature on transcriptome analysis of plants during the past ten years. Although transcriptome analyses of different plant species are done using the same sequencing methods, they often differ in the following ways: (i) objectives for which the transcriptome analysis is done; (ii) tissues from which the transcripts are obtained; and (iii) classes of transcripts that are the primary focus of analysis.

With these consideration in view, I have reviewed the transcriptome literature and organized them according to plant species.

Acacia (*A. auriculiformis* and *A. mangium*): Wong et al. (2011) identified genes for lignin biosynthesis and cell wall formation *A. auriculiformis* and *A. mangium* in young stem and inner bark tissues by using *de novo* transcriptome sequencing technology. Total 42,217 and 35,759 contigs of *A. auriculiformis* and *A. mangium,* respectively, were assembled. They found that each species had 18 lignin isoforms and five contigs homologous to R2R3-MYB proteins, which are involved in transcriptional regulation of secondary cell wall formation and lignin deposition in other plant species. They also discovered 16,648 and 9,335 high quality putative single nucleotide polymorphisms (SNPs) in transcriptomes of *A. auriculiformis* and *A. mangium,* respectively.

American Ginseng *(Panax quinquefolius L.)*: Sun et al. (2010) performed 454 sequencing with 209,747 high quality reads to generate de novo assembly 31,088 unique sequences including 16,592 contigs and 14,496 singletons. They found 21,684 (69.8%) unique sequences through BLAST search against four public sequence databases. They identified 4,097 unique sequences that were aligned to specific metabolic pathways. Additionally, they identified, 150 cytochrome P450 (CYP450) and 235 glycosyltransferase unique sequences, some of which encode enzymes related to the conversion of the ginsenoside backbone into different ginsenosides. In addition, they identified one CYP450 and four UDP-glycosyltransferases involved in ginsenoside biosynthesis through a methyl jasmonate (MeJA).

Arabidopsis *(Arabidopsis thaliana)*: Li et al. (2003) sequenced 190 polymorphic cDNA bands from leaf tissues to analyze the alignment of genomes between *Brassica oleracea* and *Arabidopsis*. They sequenced 190 polymorphic cDNA bands and compared the sequences to those of *Arabidopsis* by using FASTA and detected 169 sequences with similarity with genes in *Arabidopsis*. Transcriptome mapping based on the results of transcriptome showed that there was extensive colinearity of these genomes for chromosomal segments rather than for the whole genome. Weber et al. (2007) mapped over 15000 genes of Arabidopsis by sequencing over over 5 million expressed sequence tags (ESTs). These ESTs accounted for over 90% of transcripts that were predicted to be expressed. Recently, researchers have relied on transcriptome analysis for genome-wide studies of alternative pre-mRNA splicing in plants.

Zhou et al. (2008) used transcriptome sequence to annotate stress-inducible micro RNA in *Arabidopsis*. They found that nineteen microRNA genes of eleven microRNA families were up-regulated by cold tress. Zeller et al. (2009) used transcriptome analysis to determine the effects of salt, osmotic, cold and heat stress as well as application of the hormone abscisic acid on transcription of genes in *Arabidopsis*. They identified many stress-responsive genes, including several transcription factors as well as pseudogenes and transposons. Mining transcriptome data in search of transcription start site (TSS) Tanaka et al. (2009) detected 35,313 TSSs within 16,964 loci of *Arabidopsis*, which is more two TSSs per locus. Filichkin et al. (2010) conducted transcriptome analysis of *Arabidopsis* for genome-wide mapping of alternative splicing. They identified thousands of novel alternatively spliced transcripts and indicated that at least 42% of intron-containing genes are alternatively spliced. Gan et al. (2011) studied Genetic differences between *Arabidopsis thaliana* accessions through analysis of their transcriptomes. They reported the sequencing, assembly and annotation of the genomes of 18 natural *A. thaliana* accessions, and their transcriptomes. Sequence and expression variation was most pronounced in genes that respond to the biotic environment. Yang et al. (2011) utilized high-throughput sequencing of single-cell type of *Arabidopsis* male meiocytes to identify genes that are transcribed during meiosis. They found that more than 800 genes were preferentially expressed in meiocytes. They also found that of the 3,378 genes with the Pfam domain in *Arabidopsis*, 3,265 genes expressed in meiocytes and 18 gene families were over expressed in male meiocytes. Marquez et al. (2012) reported even greater percentage of multiexonic genes (61%) in *Arabiodopsis* that undergo alternative splicing.

Black pepper: Recently, Gordo et al. (2012) applied high-throughput sequencing to obtain black pepper root transcriptome. They found 4472 predicted proteins that showed about 52% homology with the *Arabidopsis* proteome.

Buckwheat *(Fagopyrum):* Logacheva et al. (2011) performed 454 transcriptome sequencing and de novo assembly to identify gene expression in flowers and inflorescences of two species *F. esculentum* and *F. tataricum*. They obtained 267 and 229 thousands reads for *F. esculentum* and *F. tataricum*, respectively. They found overall similarities between the two species based on the phylogenetic analysis of thirteen single-copy genes that were differently expressed.

Canola (*Brassica napus*): Trick et al. (2009) used Solexa transcriptome sequencing technology to discover Single nucleotide polymorphism (SNP) in *Brassica napus* (rapeseed). They generated approximately 40 million expressed sequence tags (ESTs) from two rapeseed cultivars. The analysis of 94000 unigenes with different read-depth stringency identified 23,330 to 41,593 putative single nucleotide polymorphisms (SNPs) between the two cultivars. Bancroft et al. (2011) used Illumina transcriptome sequencing technique to sequence oilseed rape (*Brassuca napus*). The analysis and transcript abundance demonstrated twin single nucleotide polymorphism linkage maps of *B. napus* comprising 23,037 markers. Feng et al. (2012) applied the suppression subtractive hybridization (SSH) and expressed sequence tag (EST) analysis to identify gene expression patterns during the early stages of colonization of canola roots by *Plasmodiophora brassicae*, which is a protist of the class *Phytomyxea*. They obtained 797 SSH cDNA clones, which represented 439 unigenes. Thirty-two of these genes were of *P. brassicae* origin, and remaining 407 genes were of canola origin. Liu et al. (2013) performed *de novo* transcriptome analysis of *Brassica juncea* seed coat to identify genes for flavonoid biosynthesis. They identified 69,605 unigenes from more than 116 million high-quality reads. The brown-seeded testae up-regulated 802 unigenes and down-regulated 502 unigenes in comparison with yellow-seeded ones. The dihydroflavonol reductase (DFR), leucoantho-cyanidin dioxygenase (LDOX) and anthocyanidin reductase (ANR) for late flavonoid biosynthesis were not expressed at all or expressed at a very low level in the yellow-seeded testae.

Chestnut *(Castanea dentate & Castanea mollissima)*: Barakat et al. (2009) used GS20 and 454 FLX sequencer to generate 28,890 unigenes in American chestnut (*C. dentate*) and 40,039 unigenes Chinese chestnut (*C. mollissima*) from fungal-infected and healthy stem tissues of these chestnut species. A large number of genes, which associated with resistance to biotic stimuli as well as tolerance to stresses, were identified. Stress response genes expressed more in canker tissues versus healthy stem tissues in both two chestnut varieties. They identified several candidate genes for resistance that underlie difference between American and Chinese chestnut varieties for resistance to the *Cryphonectria parasitica* fungus that causes chestnut blight.

Cucumber (*Cucumis sativus*): Gou et al. (2010) used Roche-454 for pyrosequencing to obtain the transcriptome sequences of cucumber flower buds of two isogenic lines. They obtained

353,941 high quality EST sequences, among which 188,255 were from gynoecious flowers and 165,686 from hermaphroditic flowers. A total of 81,401 unigenes, of which 28,452 were contigs and 52,949 were singletons. They identified that more than 500 alternative splicing events in 443 cucumber genes.

Grape (*Vitis vinifera*): Zenoni et al. (2010) applied transcriptome analysis of the transcriptome of grape berries during three developmental stages, post setting, pre-ripening, and ripening, and identified >6,500 genes that were expressed in a stage-specific manner. They detected 17,324 genes expressed during berry development, 6,695 of which were expressed in a stage-specific manner, suggesting differences in expression for genes in numerous functional categories and a significant transcriptional complexity.

Grass: (*Brachypodium distachyon*): Because of the ease of cultivation, genetic transformation, small genome size, and short duration, *Brachypodium* grass can be used as model plant system for developing new energy and crop plants. The International Brachypodium Initiative (2010) performed a Sanger sequencing to generate paired-end reads from different length of clones in wild grass *Brachypodium*. Comparison of the sequence and organization of the genes with those of sorghum and rice helped to explain the evolutionary history of the grass species.

Jatropha: (*Jatropha curcas*): Natarajan and Parani (2011) performed *de novo* 454 pyrosequencing to discover genes from five major tissues (roots, mature leaves, flowers, developing seeds, and embryos) of *J. curcas*. They generated 17,457 assembled transcripts (contigs) and 54,002 singletons from the sequence data. This study identified a total of 14,327 new assembled transcripts, among which 2,320 were related to major biochemical pathways including the oil biosynthesis pathway. Among the 2589 full-length transcripts identified in this study, 27 were directly involved in oil biosynthesis.

Lentil (*Lens culinaris*): Kaur et al. (2011) used Roche 454 GS-FLX Titanium technology and de novo assembly to sequence cDNA libraries of six distinct lentil genotypes. They observed that 12,639 and 7,476 unigenes matched with genes of *Medicago truncatula* and *Arabidopsis thaliana*, respectively, whereas 20,419 unigenes corresponded with genes of *Glycine max*. By screening markers of 12 cultivated lentil genotypes and one wild relative species, they identified 192 EST-SSR markers.

Maize (*Zea mays*): Poroyko et al. (2005) used serial analysis of gene expression to define the root tip transcripts of well-watered maize seedlings. The maize root transcriptome with 161,320 tags represented 14,850 genes. Comparison of the maize root transcriptome with that of *Arabiodopsis* indicated that the highly expressed transcripts differed substantially between the two species. Emrich at al. (2007) used laser-capture microdissection to isolate the maize shoot apical meristem and then used high-throughput sequencing to obtain transcriptome sequence. They found about 400 maize-specific transcripts in the meristem, demonstrating that transcriptome sequencing could be applied on specific tissue sections. Schnable et al. (2009) used Illumina sequencing method to obtain maize leaf transcriptomes from plants at different growth stages. They generated about 30 million reads from each of four developmental stages and mapped sequences to the maize genome. Vega-Arreguín et al. (2009) sequenced a cDNA library of 2 week-old *Palomero Toluqueño* maize plants. They estimated that 86,069 sequences (5.67%) did not align with public ESTs or annotated genes, putatively representing new maize transcripts. Real-time PCR of selected genes based on 74.4% of the reads in 65,493 contigs showed a correlation between corresponding levels of gene expression and the abundance of cDNA sequences in the cDNA library. Liu et al. (2012) applied transcriptome profiling to a pool of two samples generated by mixing a bulk of mutant and wild-type (WT) maize plants and demonstrated the application of RNA-seq for bulked segregant analysis (BSA) by mapping the maize mutant gene *gl3*. Using RNA-seq for this purpose does not require reference genome and differential expression profiles between the mutant and the WT are generated at no extra cost. Takacs et al. (2012) performed laser microdissection of apical domains from developing maize embryos and seedlings, and combined with transcriptome sequencing for analyses of shoot apical meristem (SAM) ontogeny. Transcriptomic profiling before and after SAM initiation indicated that organogenesis precedes stem cell maintenance in maize. He et al. (2013) generated integrated maps of transcriptomes and epigenomes of shoots and roots of two maize inbred lines and their reciprocal hybrids, and globally surveyed the epigenetic variations and their relationships with transcriptional divergence between different organs and genotypes. They observed that histone modifications were associated with transcriptomic divergence between organs and between hybrids and parents.

Mangrove apple (*Sonneratia alba*): Chen et al. (2011) used the Illumina Genome Analayzer method to obtain the transcriptome sequences of *Sonneratia alba*, a high salt resistance plant in

mangroves. More than 15 millions of 75-bp-paired-end reads were assembled into 30,628 unique sequences. They found 1266 unique genes were similar to 273 known salt responsive genes in others species.

Medicago truncatula: An adaptor-tagged normalized cDNA library from *M. truncatula* was sequenced by Cheung et al. (2006). They generated about million unique sequences through pyrosequencing and identified over ten thousand novel transcripts. Cheung et al. (2006) sequenced transcriptome of *Medicago truncatula* and obtained a total of 184,599 unique sequences contained over 400 single sequence repeats (SSRs). They found that 53,796 assemblies and singletons (29%) did not match in the existing *M. truncatula* gene index, whereas, thousands matched in a comprehensive protein database and one or more of the TIGR Plant Gene Indices.

Olive (*Olea europaea***)**: Alagna et al. (2009) applied 454 pyrosequencing technology to sequence four cDNA libraries olive fruits at different developing stages of two cultivars, one with high content of phenolics and the other lacking a specific phenolic called oleuropein. The aim of the transcriptome study was to identify genes that are expressed during fruit development. A total of 261,485 reads were obtained and 22,904 clusters were generated. The analysis provided information on variation of gene expression during fruit development between two olive cultivars with contrasting phenolic accumulation in fruits.

Palm tree: Bourgis et al. (2011) performed transcriptome sequencing and determined the metabolite content of oil palm during mesocarp development. The high oil content in oil palm was associated with much higher transcript levels for all fatty acid synthesis enzymes, specific plastid transporters, and key enzymes of plastidial carbon metabolism, including phosphofructokinase, pyruvate kinase, and pyruvate dehydrogenase.

Pea (*Pisum sativum***)**: Franssen et al. (2011) utilized *de novo* next generation of 454 sequencing to sequence cDNA libraries from different tissues (flowers, leaves, cotyledons, epi- and hypocotyl, and etiolated and light treated etiolated seedlings) of garden pea. The sequence reads were assembled into first-pass and second-pass assemblies 324,428 and 81,449 unigenes, respectively. These unigenes represented majority of transcripts in the aerial tissues.

Peanuts (*Arachis hypogaea***)** Zhang et al. (2012) analyzed the transcriptomes from three

different immature peanut seeds with different oil contents. A total of 26.1-27.2 million paired-end reads were generated from the three peanut varieties and assembled into 59,077 unigenes. Among these unigenes, only 8,252 unigenes were annotated into 42 gene ontology (GO) functional categories. A total of 18,028 unigenes were mapped to 125 biochemical pathways.

Pigeonpea (*Cajanus cajan*)*:* Dutta et al. (2011) applied 454 sequencing to cDNA from leaf, root, stem and immature seeds of two pigeonpea varieties. They identified 3,771 genic-SSR loci, excluding homopolymeric and compound repeats. From these, they developed 550 markers, which consistently amplified in eight diverse pigeonpea varieties.

Poplar (*Populus euphratica*): Poplar is a major tree important for the wood industry and is used as a model tree by plant molecular biologists because of its potential in vegetative propagation, transformation and regeneration to tree size in a relatively short time of growing. Qiu et al. (2011) studied transcriptome of poplar is well-adapted to extreme desert environments. They sequenced mRNA from living tissues of desert-grown trees and two types of callus, salt-stressed and unstressed. They found that 27% of the total genes were up- or down-regulated in response to salt stress in *P. euphratica* callus. These differentially expressed genes were mainly involved in transport, transcription, cellular communication and metabolism. In addition, they found that numerous putative genes involved in ABA (Abscisic acid) regulation and biosynthesis were also differentially regulated. Kohler et al. (2003) performed large-scale production of expressed sequence tags (ESTs) to determine sequences of cDNA libraries from roots of hybrid poplar. They identified 3021 (contributing 62%) of total 4874 unique transcripts that expressed in roots. Roots specific functions included in signaling pathways and hormone metabolism (6%); transporters and channels (4%); common metabolic pathways (5%) and energy and metabolism (8%).

Poppy *(Papaver somniferum):* Desgagné-Penix et al. (2012) studied transcriptome and metabolome of poppy cultivars in parallel-identified candidate genes involved in complex metabolic pathways. They correlated differential levels of benzylisoquinoline alkaloids (BIAs) in several cultivars with transcriptome data to pinpoint key regulatory steps of the morphine biosynthetic pathway, leading to the discovery of candidate genes implicated in BIA metabolism.

Radish (*Raphanus sativum*): Wang et al. (2012) analyzed the transcriptome of radish roots at two developmental stages and found >21,000 genes to be differentially expressed. Analysis of transcript differences between the early and late seedling developmental stages demonstrated that starch and sucrose metabolism and phenylpropanoid biosynthesis may be the dominant metabolic events during tuberous root formation.

Rice: Using transcriptome sequence analysis, Tanaka et al. (2009) determined 45,917 representative TSSs within 23,445 loci of rice. Zhang et al. (2010) performed high-throughput paired-end RNA-seq to prepare a transcriptome atlas of eight organs of cultivated rice. By analyzing alternative splicing in rice transcriptome sequences, they found that 33% of the transcripts were produced through alternative cis-splicing. They also idetified 234 putative chimeric transcripts that were produced by trans-splicing. Lu et al. (2010) applied transcriptome analysis on rice and reported that approximately 48% of rice genes show alternative splicing patterns. Mizuno et al. (2010) applied rice transcriptome sequences map to the rice genomic sequence. They identified 2,795 from shoot and 3,082 from roots that were previously unannotated in the Rice Annotation Project database. He et al. (2010) reported highly integrated maps of the epigenome, mRNA, and small RNA transcriptomes of two rice subspecies and their reciprocal hybrids. They found that gene activity was correlated with DNA methylation and both active and repressive histone modifications in transcribed regions. Differential epigenetic modifications correlated with changes in transcript levels among hybrids and parental lines. Lu et al. (2010) applied RNA-seq to globally sample transcripts of the cultivated rice *Oryza sativa indica* and *japonica* subspecies for resolving the whole-genome transcription profiles. They identified 15,708 novel transcriptional active regions, of which 51.7% have no homolog to public protein data and >63% are putative single-exon transcripts, which are highly different from protein-coding genes (<20%). They found that ~48% of rice genes show alternative splicing patterns, a percentage considerably higher than previous estimations. Using high-throughput paired-end RNA-seq, Zhang et al. (2010) analyzed the transcriptome atlas for eight organs of cultivated rice. They detected transcripts expressing at an extremely low level, as well as a substantial number of novel transcripts, exons, and untranslated regions. An analysis of alternative splicing in the rice transcriptome revealed that alternative cis-splicing occurred in approximately 33% of all rice genes. Oono et al. (2011) applied transcriptome analysis for identifying stress-inducible transcripts in rice.

Similarly, Kyndt et al. (2012) conducted transcriptome analysis of rice mature root tissue and root tips at two time points and identified 1761 root-enriched transcripts and 306 tip-enriched transcripts involved in different physiological processes. Zhai et al. (2013) used Illumina HiSeq 2000 platform to obtain the root transcriptomes of the super-hybrid rice variety Xieyou 9308 and its parents at tillering and heading stages. They analyzed ~391 million high-quality paired-end reads (100-bp in size) and aligned against the Nipponbare reference genome. They found that 92.4% annotated transcripts were represented by at least one sequence read. A total of 829 and 4186 transcripts that were differentially expressed between the hybrid and its parents were identified at tillering and heading stages, respectively

Rubber (*Hevea brasiliensis*): Chow et al. (2007) conducted transcriptome sequencing of of rubber tree using RNA from latex, which represents the rubber-producing tissue of the plant. A total 3441 unique transcripts were identified from 10,400 ESTs. Among the highly expressed ESTs, a significant proportion encoded proteins related to rubber biosynthesis, and stress- or defense responses.

Salvia miltiorrhiza (Chinese herb): Wenping et al. (2011) used Solexa deep sequencing to obtain the transcriptome sequences of *Salvia miltiorrhiza* in different stages of growing cycle and obtained 54,774 unigenes. They identified 1539 unigenes as a part of five major secondary-metabolite pathways.

Sesame (*Sesamum indicum*): Wei et al. (2011) used Illumina paired-end sequencing technology for transcriptome analysis of five tissues of sesame. Of the 86,222 assembled unigenes, 10,805 were assigned to gene ontology categories and 27,588 unigenes were clustered in orthologous groups. A number of 46,584 unigenes had significant similarities with proteins in the NCBI non-redundant database and Swiss Prot database. Total of 44,570 unigene sequences showed homologies to 15,460 genes in *Arabidopsis* based on the BLASTX analysis. In addition, 7,702 unigenes were converted into SSR makers (EST-SSR).

Solanaceae **species**: Rensink et al. (2005) analyzed ESTs and expressed transcripts (ETs) for six Solanaceae species including potato, *tomato*, tobacco, *nicotiana benthmiana*, pentunia and pepper. The cluster of 449,224 sequences of ESTs and Expressed Transcripts (ETs) were assembled into gene indices. The ESTs and ETs from different tissues presented that 55-81% of sequences had similarity at nucleotide level with sequences among the six species. However,

putative orthologs contributed 28-58% of total sequences, whereas only 16-19% of transcripts within six Solanaceae genes indices did not have matches among Solanaceae, Arabidopsis, rice or other plants.

Sorghum *(Sorghum bicolor)*: Dugas et al. (2011) applied transcriptome sequencing of sorghum and observed transcriptional activity of 28,335 unique genes from root and shoot tissues subjected to polyethylene glycol (PEG)-induced osmotic stress or exogenous ABA. Mizuno et al. (2012) performed massive parallel sequencing of mRNA to identify differentially expressed genes after sorghum plants had been infected with *Bipolaris sorghicola*, a necrotrophic fungus causing a sorghum disease called target leaf spot. They observed that pathogen infection activated the glyoxylate shunt in the TCA cycle, genes for phytoalexin synthesis and sulfur-dependent detoxification pathway.

Soybean (*Glycine max*): Komatsu et al. (2009) performed high-coverage gene expression profiling of soybean cDNA libraries from the root and hypocotyl of seedlings to identify genes that are inducible by flooding stress. They found that 97 out of 29,388 gene that were induced more than 25-fold following 12 h of flood-induced stress. Severin et al. (2010) studied trascriptomes of fourteen different types of tissues in soybean and investigated the relationship between gene structure and gene expression. They found a correlation between gene length and expression. Additionally, they observed dramatic tissue-specific gene expression of both the most highly-expressed genes and the genes specific to legumes in seed development and nodule tissues. Libault et al. (2010) performed Illumina Solexa platform to sequence cDNA libraries from 14 tissues of *Glycine max*. The annotation demonstrated 55,616 genes, among which 13,529 were *pseudogenes*. They also found that 1736 currently unannotated sequences were transcribed. The analysis of this soybean transcriptome sequences illustrated strong differential expression of genes in roots and shoots, the gene expression patterns were similar in flower and leaf. Severin et al. (2010) used the next generation Illumina sequencing to sequence cDNA libraries from fourteen different tissues (leaf, flower and pod); and two stages of pod-shell, root and nodule; and seven stages of seed development in soybean. Analysis of the transcriptome data from these tissues showed overall three clades of distinct gene expression patterns, which are represented by root, shoot and seed. They found a positive correlation between gene length and gene expression. They also observed that some genes expressed highly in seeds and some legume-specific genes expressed only in nodules. Ge et al.

(2010) performed transcriptional profiling on *Glycine soja* roots following exposure of the roots to 0.25 Hoagland's solution containing 50 mM $NaHCO_3$. The goal of this experiment was to identify genes that are induced under saline-alkaline stress. They observed that among 7088 genes, 3307 were up-regulated and 5720 were down-regulated at various time points. Most of the differentially expressed genes were involved in signal transduction, energy, transcription, secondary metabolism, transporter, disease and defense response. Guttikonda et al. (2010) used Illumina sequencing and microarray to identify cytochrome P450 monooxygenases (P450s) in soybean. A total number of 332 full-length P450 genes and 378 *pseudogenes* from the soybean genome were identified. The co-expression analysis was used to confirm that an isoflavone synthase gene, *CYP93C5*, was co-expressed with several genes related to isoflavonoid-related metabolic enzymes. Hao et al. (2011) used Solexa/Illumina sequencing and high-throughput tag-sequencing analysis of shoot and root cDNA libraries of two different soybean genotypes to identify genes associated with nitrogen-use efficiency. By comparing sequences from eight different cDNA libraries, they observed that a total of 3231 genes of 22 metabolic and signal transduction pathways were up- or down-regulated between the low-N-tolerant and low-N-sensitive varieties under N-limited condition. Kim et al. (2011) applied transcriptome analyses to investigate the plant basal defense mechanisms in resistant- and susceptible- isogenic lines of soybean that were infected with *Xanthomonas axonopodis* pv. *glycines (Xag)*. Of a total of 46367 genes that were mapped to soybean genome reference sequences, 1978 and 783 genes were found to be up- and down-regulated, respectively, in the BLP (bacterial leaf pustule)-resistant line relative to the BLP-susceptible line at 0, 6, and 12h after inoculation. Schaarschmidt et al. (2013) analyzed soybean transcriptomes to determine transcriptional changes in mycorrhizal soybean *(Glycine max)* plants and mutant lines interacting with the arbuscular mycorrhiza fungus *Rhizophagus irregularis*. They found that the colonization of roots of wild-type and a receptor kinase mutant nts1007 with *R. irregularis* resulted in a local, more than two-fold upregulation of 110 and 98 genes, respectively. Few genes were found to be downregulated in mycorrhizal wild-type or mutant nts1007.

Sweet wormwood (*Artemesia annua*): Wang et al. (2009) performed 454 pyrosequencing to characterize genes expressed in glandular trichome of the medicinal plant 'sweet wormwood'. The sets of two normalized cDNA collections from glandular trichomes yielded 406,044 ESTs that assembled into 42,678 contigs and 147,699 singletons. They also used BLAST search

against the NCBI non-redundant protein database to predict putative functions of over 28,573 unigenes. The confirmation of the expression of selected unigenes and novel transcripts in *A. annua* glandular trichomes (corresponding to enzymes for terpenoids and flavonoids biosynthesis) were done by RT-PCR analysis.

Sweetpotato (*Ipomoea batatas*): Wang et al. (2010) used Illumina paired-end sequencing technology to generate 59 million sequencing contigs that assembled *de novo* 56,516 unigenes. They applied the Kyoto Encyclopedia of Genes and Genomes Pathway database (KEGG) to map 17,598 (31.14%) unigenes in 124 KEGG pathways. Moreover, carbohydrate metabolism and biosynthesis of secondary metabolite of metabolic pathways were in 11,056 unigenes. In addition, they identified 4,114 cDNA with simple sequence repeats (SSRs), which can be used as potential molecular markers in unigenes.

Tea (*Camellia sinensis*): Shi et al. (2011) sequenced 2.59 gigabase pairs of *C. sinensis* by using high-throughput Illumina RNA-sequencing. They obtained 55,088 unigenes, among which they identified some unigenes involved in metabolic pathways related to tea quality traits, such as flavonoid, theanine and caffeine biosynthesis.

Wheat (*Triticum aestivum*): Wan et al. (2008) used Affymetrix wheat GeneChip® oligonucleotide arrays which have probes for 55,052 transcripts to transcriptome of hexaploid wheat (*Triticum aestivum, cv. Hereward*) in different time of developing caryopses. They found that number of 14,550 sequences showed significant differential regulation in the period between 6 and 42 days after anthesis. The analysis of a similar experiment on developing caryopses grown under dry and/or hot environmental treatments indicated that most environmental treatments effect on transcription depended on development, however, only a few transcripts were specifically affected.

Highlights of the plant transcriptome literature

By reviewing the above literature on transcriptome analyses of various plants, I have made the following observations:

(i) Although both Illumina and 454 pyrosequencing techniques have been used, Illumina method was used more frequently.

(ii) In all transcriptome analyses, researchers obtained thousands, and sometimes millions of reads; which were organized into hundreds and thousands of partial and full-length transcripts.

(iii) Multiple tissues were used for RNA isolation; tissues from various varieties, different tissues of the same plant, tissues of different growth stages, or tissues before and after application of treatments.

(iv) In most transcriptome analyses, researchers compared transcriptome data with that of *Arabidopsis thaliana* because it is considered a model plant and is the first plant species for which genome sequence was completed (2000). Similarly, three other major plant species, rice, maize, and soybean, for which genome sequencing has been completed, are also frequently used in comparison in transcriptome analysis.

(v) Many transcriptome analyses focused on single nucleotide polymorphism (SNPs) and simple sequence repeats (SSRs). SNPs and SSRs are used as genetic markers in plant breeding.

(vi) A few transcriptome analyses identified many peudogenes and transposons.

(vii) Many studies used plants exposed to conditional stresses such as heat, cold, and salinity. The goals of these studies were to identify stress-inducible genes. In addition, many transcriptome studies aimed at identification of disease resistance genes. For example, in canola, transcripts that are produced in response to colonization by *Phytomyxea* were identified. Similarly, in chestnut, genes induced by *Cryphonectria parasitica* were determined.

(viii) Alternative splicing was studied in several plants, including *Arabidopsis* and rice. Alternative splicing is a pre-RNA processing mechanism by which several different mRNA can be produced from one kind of pre-mRNA.

(ix) In medical plants, such as ginseng, black pepper and sweet wormwood, the main goals of transcriptome analysis were to identify genes for flavonoids, terpenoids and alkaloids biosynthesis, which were increased under certain conditions.

(x) In woody plants like poplar, the focus of transcriptome analysis was primarily on genes in the lignin biosynthesis pathway.

(xi) In oilseed plants, such as canola, palm tree, jatropha and olive, researchers identified genes in biochemical pathways for fatty acids and flavonoid biosynthesis. For example, in olives a group of researchers isolated genes for oleuropein biosynthesis.

A list of important genes that were identified by transcriptome analyses in some of the plant species are listed in Table 2.1.

The primary goal of present investigation of leucaena is to apply transcriptome analysis for identification of genes for resistance to various biotic and abiotic stresses. Many biotic and abiotic stresses originate from soils and they directly affect root systems. For example, pathogenic bacteria in soils as well as environmental conditions such as salinity, drought, heavy metals and other nutritional stresses can adversely affect the root system. We hypothesize that we will be able to identify some genes for resistance to biotic and abiotic stresses through comparison transcriptome sequences of leucaena root and shoot. Above literature survey on transcriptome analysis of various plants strongly supports that this approach should be successful.

Table 2.1: Gene identification in various plants through transcriptome analysis

Plants	Important genes identified from transcriptome sequences
Acacia	Genes for lignin biosynthesis and cell wall formation
Ginseng	Ginsenoside biosynthesis through a methyl jasmonate
Arabidopsis	Genes for tolerance to salt, osmotic, cold and heat stresses as well as hormone abscisic acid; stress-responsive genes including several transcription factors, pseudogens and transposons; genes for resistance to biotic stresses during meiosis.
Canola	Gene for flavonoid biosynthesis
Chestnut	Genes for resistance to biotic stimuli and stresses; genes resistant to chestnut blight disease caused by the *Cryphonectria parasitica* fungus.
Mangrove apple	Genes for high salt tolerance.
Grape	Different gene expression during three development stages (post setting, pre-ripening and ripening)
Jatropha	Genes involving major biochemical pathways and oil biosynthesis pathway
Olive	Gene expression during fruit development; biosynthesis of phenolics
Poplar	Genes involving ABA regulation and biosynthesis; signaling pathways and hormone metabolism in roots.
Poppy	Expression of benzylisoquinoline and morphine biosynthesis pathway genes.
Radish	Genes for starch, sucrose metabolism and phenylpropanoid biosynthesis during tuberous root formation.
Rice	Genes for alternative splicing patterns; genes with DNA methylation and active and repressive histone modifications in transcribed regions.
Rubber	Genes for rubber biosynthesis and stress- and defense responses.
Sorghum	Genes for phytoalexin biosynthesis and sulfur-dependent detoxification pathway.
Soybean	Pseudogenes, genes for saline-alkaline stress and genes for resistance to mycorrhiza *Rhizophagus irregularis.*
Sweet wormwood	Genes correspond to enzyme for terpenoids and flavonoids biosysnthesis.
Tea	Genes for flavonoids, theanine and caffeine biosynthesis.

References

Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796-815 (14 December 2000) | doi:10.1038/35048692

Alagna F, D'Agostino N, Torchia L, Servili M, Rao R, Perrotta G, et al. (2009). Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. BMC Genomics,10:399

Bancroft I, Morgan C, Fraser F, Higgins J, Trick M, et al. (2011). Dissecting the genome of the polyploid crop oilseed rape by transcriptome sequencing. Nature Biotechnology, 29:762–766

Barakat A, DiLoreto D, Zhang Z, Carlson J, et al. (2009). Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection. BMC Plant Biology, 9:51

Battich N, Stoeger T, and Pelkmans L (2013). Image-based transcriptomics in thousands of single human cells at single-molecule resolution. Nature Methods. doi:10.1038/nmeth.2657

Bourgis F, Kilaru A, Cao, Ngando-Ebongue G, Drira N, Ohlrogge J, et al. (2011). *Comparative transcriptome and metabolite analysis of oil palm and date* palm mesocarp *that differ dramatically in carbon partitioning.* Proc. Natl. Acad. Sci. U.S.A. 108:12527–12532

Brunner A, Busov V, and Strauss S. (2004). Poplar genome sequence: functional genomics in an ecologically dominant plant species. Trends in Plant Science, 9(1): 49-56

Chen S, Zhou R, Shi S, et al. (2011). Transcriptome sequencing of a highly salt tolerant mangrove species *Sonneratia alba* using *Illumina* platform. Marine Genomics 4:129–136

Cheung F, Haas B, Goldberg S, et al. (2006). Sequencing *Medicago truncatula* expressed sequenced tags using 454 Life Sciences technology. BMC Genomics, 7:272

Cheung F, Haas J, Goldberg S, May G, Xiao Y, and Town C (2006). Sequencing *Medicago truncatula* expressed sequenced tags using 454 Life Sciences technology. BMC Genomics, 7:272

Chow K, Wan K, Isa M, Bahari A, Tan S, Harikrishna K, and Yeang H (2007). Insights into rubber biosynthesis from transcriptome analysis of *Hevea brasiliensis* latex. Journal of Experimental Botany, 58(10):2429–2440

Desgagné-Penix I, Farrow S, Cram D, Nowak J, and Facchini P (2012). *Integration of deep transcript and targeted metabolite profiles for eight cultivars of* opium poppy. Plant Mol. Biol, 79:295–313

Dugas D, Monaco M, Olsen A, Klein R, Kumari S, Ware D, et al. (2011). Functional annotation of the transcriptome of *Sorghum bicolor* in response to osmotic stress and abscisic acid. BMC Genomics, 12: 514

Dutta S, Kumawat G, Singh B, Gupta D, Singh N, et al. (2011). Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan (L.) Millspaugh*]. BMC Plant Biology, 11:17

Egan A, Schlueter J, and Spooner D (2012). Applications of next-generation sequencing in Plant Biology. American J of Botany 99(2): 175–185

Emrich S, Barbazuk W, Li L, et al. (2007) Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Res, 17:69–73

Feng J, Hwang S, and Strelkov W (2012). Analysis of expressed sequence tags derived from a compatible *Plasmodiophora brassicae*–canola interaction. Can. J. Plant Pathol., 34(4): 562–574

Filichkin S, Priest H, Givan S, She R, Bryant D, Fox S, et al. (2010). Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. Genome Res., 20:45–58

Franssen S, Shrestha R, Bräutigam A, Bornberg-Bauer E, and Weber A (2011). Comprehensive transcriptome analysis of the highly complex *Pisum sativum* genome using next generation sequencing. BMC Genomics, 12:227

Gan X, Stegle O, Behr J, et al. (2011). Multiple reference genomes and transcriptomes for *Arabidopsis thaliana.* Nature 477:419–423

Ge Y, Li Y, Zhu Y, Bai X, De-Kang L, Guo D, Cai H, et al. (2010). Global transcriptome profiling of wild soybean (*Glycine soja*) roots under NaHCO3 treatment. BMC Plant Biology, 10:153

Gordo S, Pinheiro D, Edith C, Moreira E, et al. (2012). High-throughput sequencing of black pepper root transcriptome. BCM Plant Biology, 12:168

Guo S, Zheng Y, Fei Z, et al. (2010). Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types. BMC Genomics, 11:384

Guttikonda S, Trupti J, Bisht N, Chen H, Yu O, et al. (2010). Whole genome co-expression analysis of soybean cytochrome P450 genes identifies nodulation-specific P450 monooxygenases. BMC Plant Biology, 10:243

Hao Q, Zhou X, Sha A, Wang C, Zhou R, and Chen S (2011). Identification of genes associated with nitrogen-use efficiency by genome-wide transcriptional analysis of two soybean genotypes. BMC Genomics, 12:525

He G, Zhu X, Elling A, Chen L, Wang X, Guo L, Liang M, He H, Zhang H, Chen F, et al (2010). Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. Plant Cell, 22:17–33

He G, Chen B, Wang X, et al. (2013). Conservation and divergence of transcriptomic and epigenomic variation in maize hybrids. Genome Biology, 14**:**R57

Kaur S, Cogan N, Forster J, et al. (2011). Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery. BMC Genomics, 12:265

Kim K, Kang Y, Kim D, Yoon M, Moon J, Kim M, et al. (2011). *RNA-seq analysis of a* soybean *near-isogenic line carrying bacterial leaf pustule-resistant and -susceptible alleles.* DNA Res., 18:483–497

Kliebenstein D (2012). Exploring the shallow end; estimating information content in transcriptomic studies. *Front. Plant Sci*., 3:213

Kohler A, Delaruelle C, Martin D, Encelot N, and Martin F (2003). The poplar root transcriptome: analysis of 7000 expressed sequence tags. FEBS Letters, 542:37-41

Komatsu S, Yamamoto R, Nanjo Y, Mikami Y, Yunokawa H, and Sakata K (2009). A comprehensive analysis of the soybean genes and proteins expressed under flooding stress using transcriptome and proteome techniques. Journal of Proteome Research, 8:4766–4778

Kyndt T, Denil S, Haegeman A, Trooskens G, De Meyer T, Van Criekinge W, Gheysen G (2012). Transcriptome analysis of rice mature root tissue and root tips in early development by massive parallel sequencing. J Exp Bot., 63:2141-2157

Li G, Gao M, Yang B, and Quiros C (2003). Gene for gene alignment between the *Brassica* and *Arabidopsis* genomes by direct transcriptome mapping. Theor Appl Genet, 107:168–180

Libault M, Farmer A, Joshi T, Takahashi K, Langley R, Franklin L, Stacey G, et al. (2010). An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. The Plant Journal, 63: 86–99

Liu S, Yeh C, Tang H, Nettleton D, and Schnable P (2012). *Gene mapping via bulked segregant RNA-seq (BSR-Seq).* PLoS ONE, 7: e36406

Liu X, Lu Y, Yuan Y, Liu S, Guan C, Chen S, and Liu Z (2013). *De Novo* transcriptome of *Brassica juncea* seed coat and identification of genes for the biosynthesis of flavonoids. PLoS ONE 8(8): e71110

Logacheva M, Kasianov A, and Penin A (2011). De novo sequencing and characterization of floral transcriptome in two species of buckwheat (Fagopyrum). BMC Genomics, 12:30

Lu T, Lu G, Fan D, Zhu C, Li W, Zhao Q, Feng Q, Zhao Y, Guo Y, Huang X, et al. (2010). Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. Genome Res, 20:1238–1249

Marquez Y, Brown J, Simpson C, Barta A, and Kalyna M (2012). *Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis.* Genome Res. 22 1184–1195

Martin L, Fei Z, Giovannoni J, and Rose J (2013). Catalyzing plant science research with RNA-seq. *Front. Plant Sci.*, 4:66

Mizuno H, Kawahara Y, Sakai H, Kanamori H, Wakimoto H, Yamagata H, Oono Y, Wu J, Ikawa H, Itoh T, Matsumoto T (2010). Massive parallel sequencing of mRNA in identification of unannotated salinity stress-inducible transcripts in rice (*Oryza sativa L.*). BMC Genomics, 11: 683

Mizuno H, Kawahigashi H, Kawahara Y, Kanamori H, Ogata J, Minami H, et al. (2012). Global transcriptome analysis reveals distinct expression among duplicated genes during sorghum-Bipolaris sorghicola interaction. BMC Plant Biol.,12:121

Natarajan P, and Parani M (2011). *De novo* assembly and transcriptome analysis of five major tissues of *Jatropha curcas L.* using GS-FLX titanium platform of 454 pyrosequencing. BMC Genomics, 12:191

Oono Y, Kawahara Y, Kanamori H, Mizuno H, Yamagata H, Yamamoto M, Hosokawa S, Ikawa H, Akahane I, Zhu Z, et al. (2011). mRNA-Seq reveals a comprehensive transcriptome profile of rice under phosphate stress. Rice, 4:50–65

Poroyko V, Hejlek L, Spollen W, Springer G, Nguyen H, Bohnert H, et al. (2005). The maize

root transcriptome by serial analysis of gene expression. Plant Physiology, 133:1700–1710

Qiu Q, Ma T, Hu Q, Liu B, Wu Y, Zhou H, Wang Q, Wang J, and Liu J (2011). Genome-scale transcriptome analysis of the desert poplar, *Populus euphratica*. Tree Physiol. 31:452-61

Rensink W, Lee Y, Buell C, et al. (2005). Comparative analyses of six *solanaceous* transcriptomes reveal a high degree of sequence conservation and species-specific transcripts. BMC Genomics, 6:124

Schaarschmidt S, Gresshoff P, and Hause B (2013). Analyzing the soybean transcriptome during autoregulation of mycorrhization identifies the transcription factors GmNF-YA1a/b as positive regulators of *arbuscular mycorrhization*. Genome Biology, 14:R62

Schnable PS, Ware D, Fulton RS, et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. Science, 326:1112–5

Severin A, Woody J, Bolon Y, Joseph B, Diers B, Shoemaker R, et al. (2010). RNA-Seq Atlas of *Glycine max*: A guide to the soybean transcriptome. BMC Plant Biology, 10:160

Severin A, Woody J, Bolon Y, et al. (2010). RNA-Seq Atlas of *Glycine max*: A guide to the soybean transcriptome. BMC Plant Biology, 10:160

Shi C, Yang H, Wei C, Wan X, et al. (2011). Deep sequencing of the Camellia sinensis transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. BMC Genomics, 12:131

Sun C, Li Y, Wu Q, Chen S, et al. (2010). De novo sequencing and analysis of the American ginseng root transcriptome using a GS FLX *Titanium* platform to discover putative genes involved in ginsenoside biosynthesis. BMC Genomics, 11:262

Takacs E, Li J, Du C, Ponnala L, Janick-Buckner D, Yu J, et al. (2012). *Ontogeny of the maize shoot apical meristem.* Plant Cell, 24:3219–3234

Tanaka T, Koyanagi K, and Itoh T (2009). Highly diversified molecular evolution of downstream transcription start sites in rice and *Arabidopsis.* Plant Physiol., 149:1316–1324

The International Brachypodium Initiative (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*, 463(11)

Trick M, Long Y, Meng J, and Bancroft I (2009). Single nucleotide polymorphism (SNP) discovery in the polyploid Brassica napus using Solexa transcriptome sequencing. Plant Biotechnology Journal, 7:334–346

Vega-Arreguín J, Ibarra-Laclette E, Herrera-Estrella A, et al. (2009). Deep sampling of the *Palomero* maize transcriptome by a high throughput strategy of pyrosequencing. BMC Genomics, 10:299

Wan Y, Poole R, Huttly A, Mitchell R, et al. (2008). Transcriptome analysis of grain development in *hexaploid wheat*. BMC Genomics, 9:121

Wang L, Li P, and Brutnell T (2010). Exploring plant transcriptomes using ultra high-throughput sequencing. Briefings in Functional Genomics, 9(2):118-128

Wang S, Wang X, He Q, Liu X, Xu W, Li L, et al. (2012). Transcriptome analysis of the roots at early and late seedling stages using Illumina paired-end sequencing and development of

EST-SSR markers in radish. Plant Cell Rep., 31:1437–1447

Wang W, Wang Y, Zhang Q, Qi Y, and Guo D (2009). Global characterization of Artemisia annua *glandular trichome* transcriptome using 454 pyrosequencing. BMC Genomics, 10:465

Wang Z, Fang B, Yujun Li, et al. (2010). *De novo* assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (*Ipomoea batatas*). BMC Genomics, 11:726

Weber A, Weber K, Carr K, et al. (2007). Sampling the *Arabidopsis* transcriptome with massively parallel pyrosequencing. PlantPhysiol, 144:32–42

Wei W, Qi X, Wang H, Zhang Y, Hua W, Li D, Haixia L, and Zhang X (2011). Characterization of the sesame (*Sesamum indicum L.*) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. BMC Genomics, 12:451

Wenping H, Yuan Z, Jie S, Lijun Z, and Zhezhi W (2011). De novo transcriptome sequencing in *Salvia miltiorrhiza* to identify genes involved in the biosynthesis of active ingredients. Genomics, 98:272–279

Wong M, Cannon C, and Wickneswari R (2011). Identification of lignin genes and regulatory sequences involved in secondary cell wall formation in *Acacia auriculiformis* and *Acacia mangium* via *de novo* transcriptome sequencing. BMC Genomics, 12:342

Yang H, Lu P, Wang Y, and Ma H (2011). The transcriptome landscape of *Arabidopsis* male meiocytes from high-throughput sequencing: the complexity and evolution of the meiotic process. The Plant Journal, 65:503–516

Zeller G, Henz S, Widmer C, et al. (2009). Stress-induced changes in *Arabidopsis thaliana* transcriptome analyzed using whole-genome tiling arrays. Plant J., 58:1068-1082

Zenoni S, Ferrarini A, Giacomelli E, Xumerle L, Fasoli M, Malerba G, et al. (2010). Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq. Plant Physiol., 152:1787–1795

Zhai R, Feng Y, Wang H, et al. (2013). Transcriptome analysis of rice root heterosis by RNA-Seq. BMC Genomics, 14:19

Zhang G, Guo G, Hu X, Wang J, et al. (2010). Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. Genome Res., 20: 646-654

Zhang G, Guo G, Hu X, Zhang Y, Li Q, Li R, Zhuang R, Lu, He Z, Fang X, et al (2010). Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. Genome Res., 20:646–654

Zhang J, Liang S, Duan J, Wang J, Chen S, Li Y, et al. (2012). *De novo* assembly and characterisation of the transcriptome during seed development, and generation of genic-SSR markers in Peanut (*Arachis hypogaea L.*) BMC Genomics, 13:90

Zhou X, Wang G, Sutoh K, Zhu J, and Zhang W (2008). Identification of cold-inducible microRNAs in plants by transcriptome analysis. Biochimica et Biophysica Acta, 1779: 780–788

Chapter 3

TRANSCRIPTOME ANALYSIS OF LEUCAENA

3.1. Introduction

In recent years, plant biologists have applied transcriptome sequence analysis as a tool to identify functional genes, especially for resistance to biotic and abiotic stresses in different plants. For example, Rensink et al. (2005) used expressed sequence tags (EST) to determine the effects of abiotic stresses such as heat-, cold-, salt-, and drought treatments on gene expression in potato leaves and roots. They found 5240 unique sequences that expressed in response to abiotic stress. Rodriguez et al. (2010) characterized *Craterostigma plantagineum* transcriptome from leaves at four stages of dehydration and rehyderation by using deep sequencing technologies. A total of 182 Mb reads were assembled into 29,400 contigs. The *C. plantagineum* transcripts were more similar to *Vitis vinifera*, castor bean (*Ricinus communis*) and poplar (*Populus trichocarpa*) reaching to sixty-seven per cent than to any other species. They found that some dehydration-responsive transcripts, which accumulate in desiccation-sensitive plants in response to dehydration are expressed constitutively in desiccation-tolerant *C. plantagineum.*

Considering that leucaena is highly tolerant to drought, insect-pests, diseases and other environmental stresses, we are interested in identifying and isolating leucaena genes for tolerance to various biotic and abiotic stresses. Many genes for disease resistance in plants have been characterized by sequencing in recent years. Higher plant species have two main branches of immune systems (i) one branch uses transmembrane pattern recognition receptors (PRRs) that respond to pathogen-associated molecular patterns (PAMPS) (Jones and Dangl, 2006); (ii) second branch acts mostly inside the cell using NB-LRR-proteins encoded by R-genes. Proteins encoded by R-genes are called NB-LRR because they contain a nucleotide-binding (NB) domain and a leucine-rich repeat (LRR) structural motif. Plant pathogens infect plants by injecting various effector molecules through type III and type IV secretion systems (Hauser 2009, Fronzes et al. 2009). NB-LRR proteins recognize the effectors introduced by pathogens and induce defense response to infection in plants. The effector molecules injected by pathogens target certain cellular components of the host for inducing a disease response.

The NB-LRR resistance proteins (R-proteins) recognize pathogen's effectors by monitoring their interactions with targets. The recognition of effectors by NB-LRR results in effector-triggered immunity, which is usually a hypersensitive reaction resulting in localized cell death at the site of infection. Considering that leucaena has no known plant diseases and is a tetraploid species, it is expected that it will have many NB-LRR resistance genes. It is our goal to identify those NB-LRR genes through transcriptome analysis.

In *Arabidopsis*, R genes encode 125 NB-LRR proteins. Those NB-LRR resistance genes are subdivided based on N-terminal structure. Some of them contain coiled-coil (cc) motif called cc-NB-LRR (Dangl and Jones, 2001). Another group has a domain with homology to the *Drosophila* Tol receptor and the mammalian IL-1 receptor and is known as Tir-NB-LRR. Some resistance genes contain a WRKY domain in addition to NB-LRR. WRKY proteins are zinc-figure transcription factors that are activated during plant defense responses. In addition to this domain, some R-genes carry a protein kinase domain. Some resistance genes, such as the Pto gene of *Arabidopsis*, tobacco and tomato, (He et al. 2006, Martine et al. 1994, Thilmony et al. 1995) encode a cytoplasmic Ser/Thr kinase, which may be membrane-associated through fatty acid myristoylation at the N-terminus. The rice gene *Xa*21, which confers resistance to bacterial blight caused by *Xanthomonas orizae* pv *orizae* carries an LRR motif and a Ser/Thr kinase-like domain (Song et al. 1995)

Some pathogenic effectors target MEK or MAPKKK kinases, thereby inhibit MAP kinase pathway enzymes that are necessary for hosts defense. The MAP kinase pathway is involved in resistance to infection caused by *Phytophthora* spp. in tobacco (Yang et al. 2001). We consider it important to identify the MAP kinase cascade genes in leucaena.

Plant chitinases are enzymes that hydrolyze the β-1,4-glycoside bond of N-acetylglucosamine polymer chitin. Chitinase enzymes are up-regulated by various biotic (fungi, bacteria, viruses, viroids, fungal cell wall components and oligosaccharides) and abiotic (ethylene, salicylic acid, salt solution, ozone and UV light) stresses (Punija and Zhang 1993, Kasprzewska 2003). Therefore, chitinases play a crucial role in plant response to pathogens and stresses. Previous studies indicated that chitinases were found in different organs such as flowers, seeds and tubers (Karasuda et al. 2003). There are seven classes of chitinases known to be present in plants. They are classified according to primary structure, substrate specificity,

and catalytic mechanism. Genes for chitinases generally express constitutively in higher plants, but they may be up-regulatd following exposure to a number of fungal, bacterial, and viral pathogens (Conrads-Strauch et al. 1990; Daugrois et al. 1990; Gerhardt et al. 1997; Joosten and de Wit 1989; Vasanthaiah et al. 2010). Chitinases of classes I through IV have been shown to be upregulated in response to pathogen infection (Hamel et al. 1997; Kasprzewska 2003). In our laboratory, previously, Rushanaedy et al. (2012) showed that four classes of chitinases were induced in *Acacia koa* in response to infection by *Fusarium oxysporum*. In addition, chitinases helps in forming symbiotic interaction of legumes with nitrogen-fixing rhizobia or mycorrhizal fungi by localized suppression of the defense reaction of the plants. Leucaena is well-known for its high resistance to biotic and abiotic stresses; therefore, we expect that leucaena contains many genes for chitinases, whose expression levels may be higher in roots than in shoots. From the analysis of root and shoot transcriptome of leucaena, we expect to identify many genes for chitinase.

In recent years, molecular aspects of drought tolerance have received special attention from plant biologists (Shinozaki and Yamaguchi-Shinozaki 2007). Researchers have identified many genes in different plants that were induced by drought stress. To date, a total of 16 early responsive dehydration (ERD) genes have been isolated from *Arabidopsis*. These genes encode proteins that include ATP-dependent protease, heat-sock proteins (HSP), S-adenosyl-methionine-dependent methyltransferases, membrane proteins, proline dehydrogenase, sugar transporter, senescence-related genes, glutathione-S-transferase, group II LEA (Late Embryogenesis Abundant) protein, chloroplast and jasmonic acid biosynthesis proteins, and ubiquitin extension protein (Alves and Fietto, 2013). The ERD gene family has been collectively characterized as genes that are rapidly induced by dehydration (Kiyosue et al. 1994) ERD15 from *Arabidopsis* has been functionally characterized as a common regulator of the abscisic acid (ABA) response and salicylic acid (SA)-dependent defense pathway (Kariola et al. 2006). ERD16 encodes a ubiquitination extension protein (Kiyosue et al. 1994). In soybean, eight genes ERD1, ERD2, ERD3, ERD9, ERD10, ERD12, ERD15 and ERD16 were isolated in response to stress. Some of ERD genes in soybean have orthologs, for examples, ERD5, which encodes a precursor of a proline dehydrogenase, has five orthologous genes in soybean. Similarly, ERD7, which encodes a protein related to senescence and dehydration has nine orthologous genes. ERD7 plays a central role in response to drought and osmotic stress

and it is related with drought-induced leaf senescence in plants (Munné-Bosch and Alegre 2004). Considering that leucaena is much more tolerant to drought than *Arabidopsis* and soybean, we expect that leucaena has many more ERD genes.

In tropical environmental conditions like Hawaii, plants have to withstand high level of UV radiation, which is harmful to plants. Important targets of UV in plant cells are DNA, lipids and proteins and also vital processes such as photosynthesis. It had been reported that mitochondria and chloroplasts produce reactive oxygen species (ROS) in response to UV, making plants vulnerable to cell death. UV tolerant plants produce UV sunscreen pigments in response to UV radiation. The UV sunscreen pigments include flavonoid, DNA repair enzymes and proteins involved in mitigating oxidative stress (Favory et al. 2009). In plants, UV-B radiation evokes diverse phenotypic responses, including hypocotyl growth inhibition, cotyledon expansion, phototropic curvature and induction of UV-B-protecting pigmentation. Some genes involved in UV-B light perception, signaling and stress pathway are COP1, HY5, HYH, BBX24 and RUP2 (Nawkar et al. 2013). Considering that leucaena grows well in sunny hot tropical environment, it is expected it is naturally highly tolerant to UV radiation. We expected that leucaena has multiple mechanisms for UV tolerance and therefore, we are interested in isolating these genes from leucarna transcriptome sequencing.

3.2. Materials and Methods

(a) Growing leucaena seedlings for RNA isolation:

Healthy mature seeds of leucaena cultivar K636 were germinated and grown in Leonard's jar assemblies containing a mixture of vermiculite and perlite in the upper chamber and Hoagland's plant nutrient solution in the lower chamber (Hoagland and Arnon, 1950). The plants were maintained in a growth room with a 16 h light/10 h dark photoperiod at 25 ˚C. Three-months-old leucaena plants were 10-12 cm in height and had profuse roots extending the entire upper chamber of the Leonard jar. These plants were carefully washed with clean water to remove all vermiculite particles from the roots.

(b) Total RNA isolation:

The washed leucaena plants were frozen in liquid nitrogen and were then dissected into the shoot and root sections. Approximately 100 mg of shoot and root tissues were separately

used to isolate total RNA. The samples were pulverized in a mortar with a pestle to obtain powered plant tissues, which were then used for RNA isolation using a Qiagen RNeasy Plant mini kit. The procedure was repeated multiple times with additional plant tissue samples until the desired amount of RNA was obtained. To remove any traces of contaminating DNA in the samples, the samples were pooled and treated with Turbo DNase according to the manufacturer's instruction (2 μl per 100 μl of total RNA solution). The quantity of RNA in the shoot and root samples and their A260/A280 ratios were determined by using a Nanodrop spectrophotometer.   The A260/A280 ratios for the shoot and root samples were 2.14 and 2.09, respectively.

(c) Illumina sequencing of trancriptome:

Total amount of shoot RNA (68 μg) and total root RNA (61 μg) were sent to SeqWright, Inc (Houston, TX) for transcriptome sequencing using the Illumina HiSeq2000 sequencer. The HiSeq2000 is capable of generating up to 200 gigabases per run with 100 bp read length.   They purified mRNA from total RNA and converted it into cDNA before sequencing as stated in the following steps:

LIBRARY PREPARATION

   i.   DNA Fragmentation
   ii.   End Repair
   iii. Ligate Adapters
   iv. Size Selection
   v.   Load Cluster Station

CLUSTER GENERATION

   i.   Bind DNA to surface of flow cell channels
   ii.   Primer Extension
   iii. Bridge Amplification
   iv. Linearization
   v.   Blocking
   vi. Primer Hybridization

SEQUENCING

     i.   Unload Cluster Station

    ii.   Prep Instrument and Begin Sequencing

   iii.   Data Analysis

(d) Bioinformatics analysis:

Blast analysis: The homology of 3.5 million contigs was determined using the BLASTX algorithm to compare sequence to the NCBI non-redundant protein sequences (nr) database on server of The Zhao Bioinformatics Laboratory (the Noble Foundation, Ardmore, OK).

ClustalW analysis: the multiple alignment was made using clustalW (http://www.genome.jp/tools/clustalw/) to identify the relationship of ERD gene in leucaena and different plant species.

Statistical analyses: Excel tool was used to analyze other statistic traits in leucaena transcriptome sequences.

3.3. Results

(a) The leucaena transcriptome at a glance:

The leucaena transcriptome data revealed 2237,641 contigs and 2187,719 Scaffolds (Table 3.1). A contig is a contiguous length of DNA including a set of overlapping DNA segments. Each contig represents the sequence of a part or an entire transcript. A scaffold is an ordered set of contigs, which are linked by sequences that were derived from the paired-end information of contigs. Thus, scaffolds always consist of contigs separated by gaps. Some scaffolds can be made of only one contig; in that case, the scaffold is same as the contig.

Table 3.1: Sequences at a glance of leucaena

	In leucaena roots	In leucaena shoots
Total number of contigs	1,190,291	1,047,350
Total number of scaffolds	1,165,136	1,022,583
BLASTX databases		
Total number of scaffolds done by BLASTX	112,091	199,818
0 – 499 bp	90,416	128,161

500 – 599 bp	1,633	8,722
600 – 699 bp	976	4,217
700 – 799 bp	636	3,490
800 – 899 bp	421	3,146
900 – 999 bp	290	14,012
>1.0 kb	789	1,590

The number of root scaffolds was 97.9% of total number of contigs. Similarly, the number of shoot scaffolds was 97.6% of total number of contigs. The number of contigs as well as scaffolds in root transcriptome sequence was slightly higher than in the shoot transcriptome. Although the total number of scaffolds and contigs in root sequences was slightly higher than in the shoot, however, the number of root contigs or scaffold >500 bp was less in the transcriptome for roots than for shoots. All of the following analyses in this chapter are based on BLAST search results of total number scaffold sequences. BLASTX search was performed against the nr database on server of Bioinformatics Laboratory, The Nobel Foundation, Ardmore, OK.

(b) Identification of root-specific transcripts:

Among 1.1 million scaffolds of leucaena root transcriptome sequences, 3687 sequences, which were more than 500 bp in length, were individually selected to make microarray chips. The longest sequence selected for microarray was 5,178 bp.
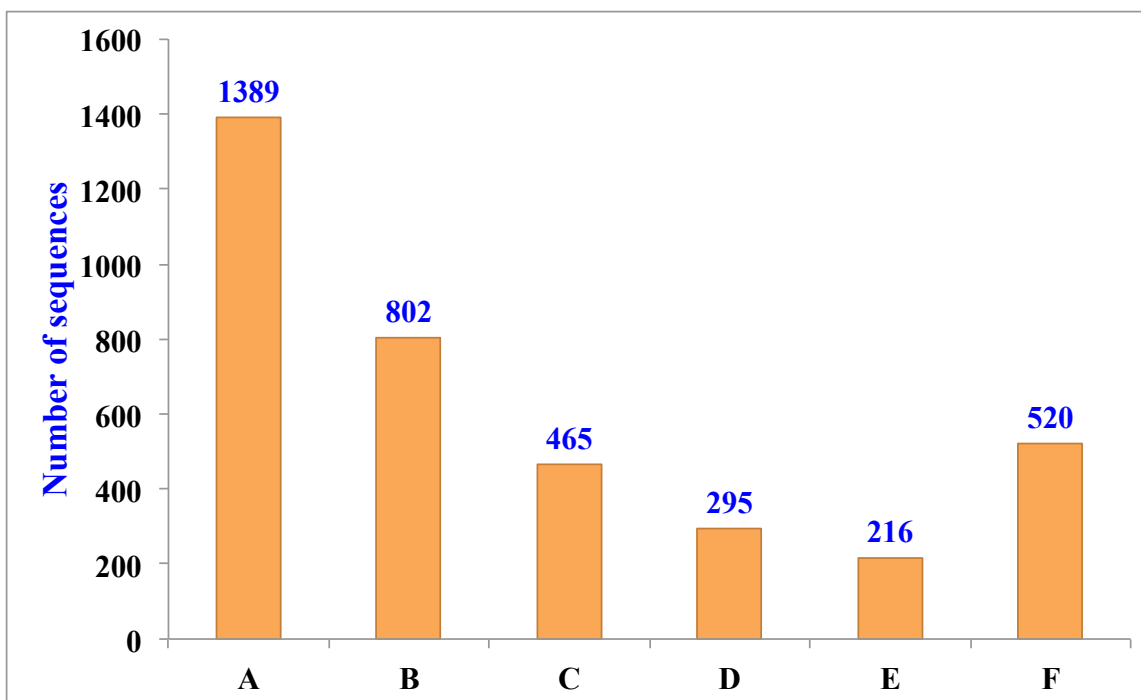
Figure 3.1: The number of sequences that were selected for microarray from the root transcriptome sequences. (A) 500-599 bp, (B) 600-699 bp, (C) 700 – 799 bp, (D) 800 – 899 bp, (E) 900 – 999, (F) > 1000 bp.

(c) Root transcriptome sequences that are absent in the shoot trancriptome:

Through BLAST analysis, we found 33 root transcriptome sequences that were absent in the shoot trancriptome and were larger than 500 bp. Twelve of them showed significant homologies with known proteins in the NCBI non-redundant database when group BLAST analysis was done by Dr. Patrick Zhao and Dr. Xinbin Dai of Noble Foundation, OK (Table 3.2a). Among the remaining 21 sequences that exhibited no homology or insignificant homology to known proteins, in BLASTx analysis, were grouped as hypothetical proteins (Hyps). Although these sequences showed some insignificant homologies, their functional role cannot be assigned based on homology to proteins in the database with known functions (Table 3.2b).

Table 3.2 (a): >500 bp root transcriptome sequences that are absent in the shoot trancriptome and show homology with known proteins in the databse.

Query name	Query length	Query name	Hsp expect	Remarks
991799	520	Lactosylceramide 4-α-galactosyltransferase	1E-09	Lactosylceramide 4-alpha-galactosyltransferase of *Arabidopsis*
1806553	1002	Succinoaminoimidazolec arboximide ribonucleotide synthetase	e-116	Ribonucleotide synthetase of long bean
489168	599	Anthocyanin 5-aromatic acyltransferase-like	2.00E-49	Anthocyanin 5-aromatic acyltransferase-like in soybean
24395	979	Bark storage protein A	e-102	Bark storage A protein of grape.
624	523	Linoleate 9S-lipoxygenase 5, chloroplastic-like	9E-77	Lipoxygenase of grape and cacao
981413	822	Vicilin-like antimicrobial peptides	e-100	Vicilin-like-antimicrobial peptides of soybean.
1539993	888	GAMYB-binding protein	0.00004	Regulator of gibberellin-responsive genes in different plants
1475153	539	Insulin metalloproteinase	0	Insignificant homology
999761	661	Pelota homolog	0.00001	Pelota homolog of grape
974403	530	S-locus F-box brothers	4.00E-12	F-box containing protein in cacao
493205	566	Sucrose-6-phosphate hydrolase	0	Insignificant homology
1806606	2651	Uracil transporter-like protein	0	Allantoin permease of chickpea

Query 1806606: This sequence is 2651 bp long; nearly half of this sequence 1342 showed 82% identity with allantoin permease of chickpea. The same sequence also showed 81% identify with uracil transporter protein in *Medicago truncatula*. Allatoin is a product of urea pathway of

34

nitrogen transport from root to shoot in some tropical legume. Therefore, this gene may be involved in nitrogen transport from root to shoot in leucaena. First 1300-bp segment of this sequence, which no homology to allantoin permease, was used again BLASTX analysis. A 224 bp segment (from 659 to 879) showed up to 84% identify with cysteine/histidine-rich domain family protein in *Theobroma cacao*. However, the direction of sequence was opposite from sequence of allantoin permase like protein. Therefore, it appears likely that this sequence might have been created by a miss-combination of two unrelated contigs.

Query 974403: The sequence has a conserved domain belonging to the F-box superfamily. A 404 bp segment of the sequence showed 34% similarity with F-box protein of cacao.

Query 999761: A 120-bp region of the sequence showed 65% identify and 80% similarity with 'protein pelota homolog' of grape (*vitris vinfera*), which may be involved in cell control and cell division.

Query 981413: The sequence has a conserved homology domain belonging to cupin 1,2 domain families. It showed 67% identify and 79% similarity with predicted protein of *Poplus* (poplar) with unknown function. It also showed 64% identify and 80% similarity with the vicilin-like-antimicrobial peptide of soybean.

Query name 624: The 523 bp entry sequence has 79% identify and 89% similarly with PLA/TLH2 of cacao. The same homology was with linolate -9S-lipoxygenase 5 of grape. This enzyme family of iron-containing enzymes that catalyze the dioxygenation of polyunsaturated fatty acids, involving in growth and development as well as pest resistance and response to wounding.

Query name 24395: This sequence has a conserved domain belonging to PNP/ UP1 super family. It showed 62% identify and 80% similarly with 'bark storage A protein' of grape. It also showed 59% identify and 79% similarly with 'phospholyase protein' in cacao.

Query 489168: It has a conserved domain belonging to the PSL1 super family. It showed 52% identify and 69% similarly with anthocyanin 5-aromatic acyltransferase-like protein in soybean.

Query 180653: A 700-bp region of the sequence showed 82% identify and 79% similarly with succinoaminoimidazolecarboximide ribonucleotide synthetase of long bean, which is involved in purine biosynthesis.

Query 991799: A short segment (from 9 to 134) of this sequence showed 67% identify and 85% similarly with lactosylceramide 4-alpha-galactosyltransferase of *Ricinus communis*. The same sequence also showed 60% identity and 83% similarity with alpha, 1,4-glycosyltransferase-like protein in *Arabidopsis*.

Query 1539993: The sequence showed low identity and similarity with GAMYB proteins. GAMYB may be a regulator of gibberellin-responsive genes in leucaena.

Surprisingly, queries 493205 and 1475153, which showed some significant homologies with known protein in group BLAST analysis, did not show significant homologies with any protein when I did the analysis with each of them individually.

Table 3.2 (b): >500 bp root transcriptome sequences that are absent in the shoot trancriptome and show no significant homology with known proteins in the database.

Query name	Query length	Query name	Hsp expect	Remarks
1804543	503	Brichos domain-containing protein 5	4.1	Hyps
984890	668	Cysteine string protein	6.4	Hyps
1503097	799	Delangin	0.66	Hyps
1004388	502	Deoxyguanosinetriphosphate triphosphohydrolase	4.1	Hyps
1581451	762	Erythrocyte membrane-associated antigen	0.78	Hyps
487575	679	ESX conserved componant EccC2	6.7	Hyps
490943	699	Foldase protein PrsA	1.9	Hyps
493239	555	Inositol-pentakisphosphate 2-kinase	0.78	Hyps
1492679	569	Lymphocyte cytosolic	3.2	Hyps
1808928	1040	Malaria antigen	3	Hyps
7873	677	Obtusifoliol-14-demethylase	1	Hyps
1571651	515	Phage/plasmid primase, P4 family, C-terminal domain	7	Hyps
1803348	1322	Pneumococcal surface protein A	0.67	Hyps

1552689	700	Proline-specific permease put4	9.3	Hyps
1498862	649	Regulator Ustilago maydis	0.41	Hyps
979360	645	SMC domain protein	6	Hyps
1552323	724	Thiol-disulfide isomerase-like thioredoxin	7.7	Hyps
1509002	829	Uracil permease	7.7	Hyps
2586	667	YD repeat-containing protein	1.7	Hyps
1462324	542	Acid phosphatase stationary-phase	3.6	Hyps
1543587	523	ALK tyrosine kinase receptor	9.3	Hyps
1540840	507	Anti-SigV factor	2.4	Hyps
983032	514	Bacterial Ig-like domain	0.37	Hyps

(d) Identification of chitinase genes:

Considering that chitinase plays an important role in providing resistance to various biotic and abitoc stresses in plants, we looked for chitinase-related genes from the leucaena transcriptome. There were 74 and 160 chitinase sequences in root and shoot transcriptome, respectively. Among these 234 sequences, 124 were larger than 100 bp, which we selected for further analysis. The longest sequence was query 1673697 (1955 bp) that has 91% identity with chitinase class Ib from *Acacia koa*, which was previously characterized in our lab (Figure 3.2) (Rushnaedy et al. 2012). Beside class I, many class II and III were observed in leucaena transcriptome.

Score	Expect	Method	Identities	Positives	Gaps	Frame
524 bits(1350)	1e-179	Compositional matrix adjust.	242/266(91%)	255/266(95%)	0/266(0%)	-1

```
Query 1097  WECKGWSKYCCNLTITDYFQTYQFENLFSKRNTPVAHAVGFWDYHSFITAAALFEPLGFG  918
            WECKGWS YCCNLTITDYFQ YQFENLFSKRN+PVAHAVGFWDYHSFITAAA++EPLGFG
Sbjct   44  WECKGWSIYCCNLTITDYFQPYQFENLFSKRNSPVAHAVGFWDYHSFITAAAVYEPLGFG  103

Query  917  TTGNKTMQMMEIAAFLGHVGSKTSCGYGVATGGPLAWGLCYNHEMSPSQSYCDNYYKYIY  738
            TTGNKTMQM EIAAFL HVGSKTSCGYGVATGGP AWGLCY+HEMSPSQSYCD+Y+KY Y
Sbjct  104  TTGNKTMQMKEIAAFLAHVGSKTSCGYGVATGGPFAWGLCYSHEMSPSQSYCDDYFKYTY  163

Query  737  PCAPGAQYYGRGALPIFWNYNYGAAGEALKVDLLSHPEYVEQNATLAFQAAIWRWMTPIN  558
            PCAPGA YYGRGALPIFWNYNYGAAGEALKVDLLSHPEYVEQNATLAFQAAIWRWMTPI
Sbjct  164  PCAPGADYYGRGALPIFWNYNYGAAGEALKVDLLSHPEYVEQNATLAFQAAIWRWMTPIK  223

Query  557  KKQPSAHDAFVGNWKPTRNDTLENRVPGFGASMNILYGDGVCGKGDVDSMNNIISHYLYY  378
            KKQPSAHDAFVG+WKPT+NDT+ NR+PGFG +MNILYGDGVCG+GDVDSMNNI+SHYLYY
Sbjct  224  KKQPSAHDAFVGSWKPTKNDTIANRLPGFGTTMNILYGDGVCGQGDVDSMNNIVSHYLYY  283

Query  377  LDLLGVGREGAGPHELLTCAEQVPFN  300
            LDLLGVGRE AGPHELLTCAEQVPFN
Sbjct  284  LDLLGVGREDAGPHELLTCAEQVPFN  309
```

Figure 3. 2: BLASTX analysis showing 91% identities of deduced amino acid sequence of query 1673697 with the *Acacia koa* chitinase class 1b
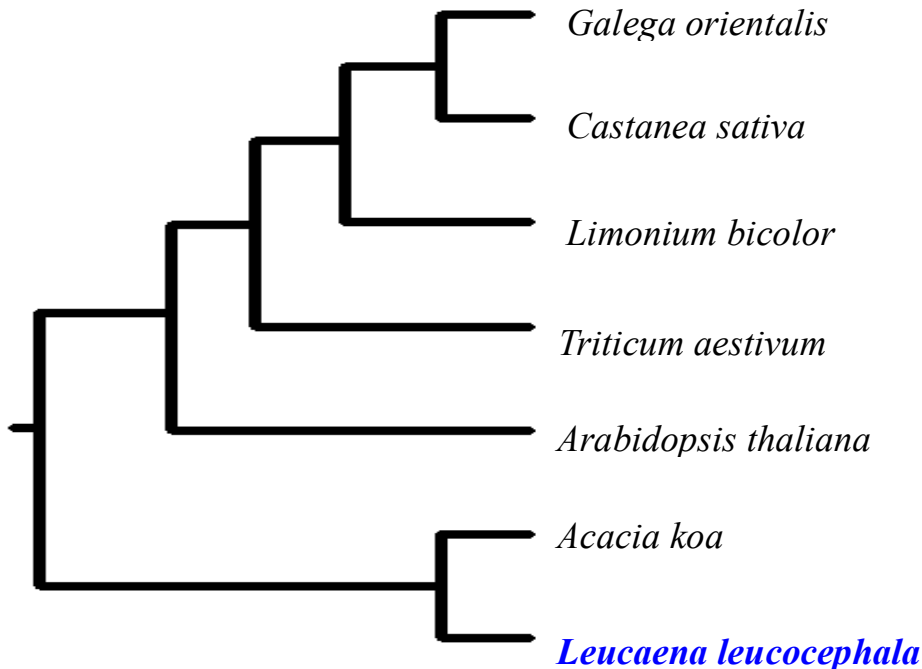
Figure 3.3: Phylogenetic tree showing relationship of chitinase class Ib in *Leucaena leucecophala* and other plant species. The deduced amino acid sequence (321 amino acid) of

query 1673697 was used with homologous sequences from other plant species for phylogenetic tree construction.
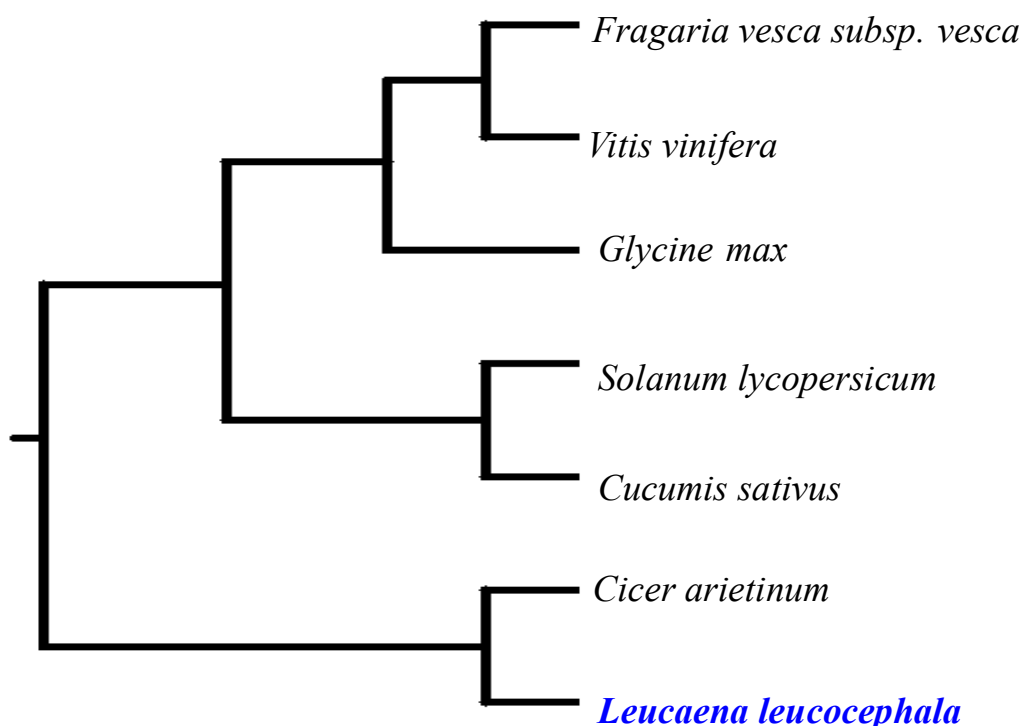
Figure 3.4: Phylogenetic tree showing relationship of acidic mammalian chitinase-like protein in *Leucaena leucecophala* and other plant species. The deduced amino acid sequence (339 residues) of query 1632644 was used with homologous sequences from other plant species for phylogenetic tree construction

The long chitinase sequences were observed mostly in the shoot transcriptome. Twelve chitianse-related sequences in shoots were longer than 500 bp. (Figure 3.4) Among these 12 sequences, 5 showed homology with acidic mammalian chitinase-like protein, previously identified in soybean (query ID 1655904, 1650971) and chickpea (query ID 1632644, 1645828, 1654177). Query 1661959 (1017 bp) showed homology 51% identity with chitinase homologue in *Sesbania rostrata.* Query 482347 (575 bp) was not significant. Query 472878 (1100 bp) showed homology with chitinase class I, previously identified from *Leucaena leucocephala* (Kaomek et al. 2003), while query 1673697 (1955 bp) and query 1652009 (885 bp) showed homologies with chitinase class Ib and class III, respectively from *Acaia koa* (Rushanaedy et al. 2012). The same chitinase also expressed in root. Query 971835 (1154 bp) showed the homology with *lecaena leucocephela.* Class II chitinase was observed only in root

while class III chitinase was found in both shoot and root transciptome sequences. Class IV chitinases were not observed in the leucaena transcriptome. It is known that the class IV chitinases are similar to the class I chitinases, except that the class IV chitinases have four large deletions (Ancillo et al. 1999; Collinge et al. 1993; Shakhbazau and Kartel 2008).

Table 3.3: Number of queries (>100 bp) encoding chitinase-related proteins in the leucaena transcriptome.

	Query description	Number of Query
In shoot	chitinase homologue	1
In root	chitinase 10-like	1
	chitinase 3-like	1
	chitinase class II	1
	chitinase domain-containing protein 1-like isoform X2	1
	chitinase-related agglutinin	3
In both shoot and root	acidic mammalian chitinase-like	7
	chitinase 2-like	3
	chitinase 2-like isoform 2	2
	chitinase class I	42
	chitinase class Ib	7
	chitinase class III	5
	chitinase domain-containing protein 1-like	3
	chitinase KBchit5-3-1	22
	chitinase-like protein 1-like	3
	chitinase-like protein 1-like isoform 1	4
	chitinase-like protein 2-like	16
Total		124
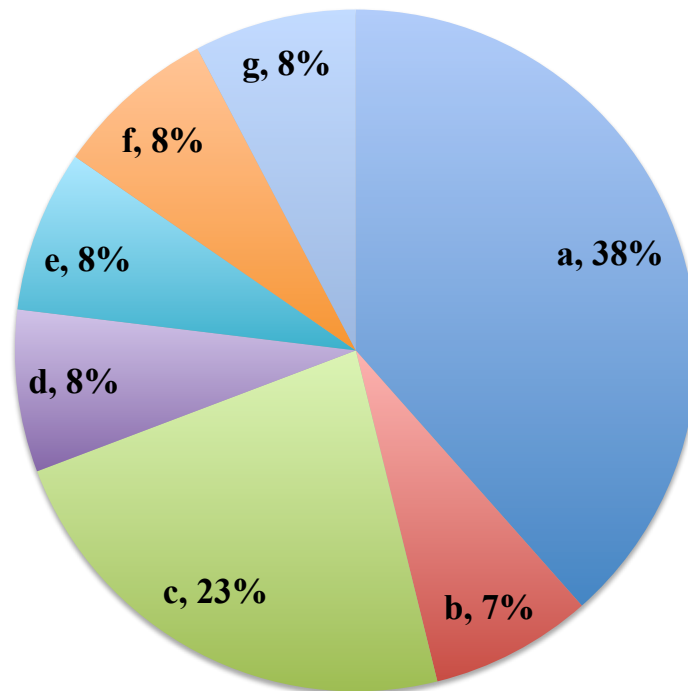
Figure 3.5: The percentage of different kinases in 12 leucaena transcriptome sequences >500 bp (a) acidic mammalian chitinase-like, (b) class Ib chitinase, (c) chitinase I, (d) chitinase homologue, (e) class III chitinase, (f) chitinase-like protein 2-like, (g) chitinase 2-like

Five queries showed high homology with acidic mammalian chitinase-like protein. They also showed 80% identity and 65% similarity with chitinase V of tobacco. Interestingly, they showed 65% identity and 79% similarity with chitinase-related agglutinin in black locust (*Robinia pseudoacacia*). Agglutinins are also known as lectins, which are carbohydrate-binding proteins present in plants. The chitinase V-like agglutinin is present in the bark of black locust; thus, it is likely that leucaena also has chitinase V-like lectin in the bark. The RobpsCRA protein or the chitinase V-like agglutinin from black locust can cause agglutination of animal blood cells (Van Dames et al. 2007). Therefore, it is likely that bark extracts of leucaena may also cause agglutination in animals due to presence of this chitinase V-like protein. Plant bark lectins may represent a class of entomotoxic proteins providing resistance to phytophagous insects (Vandenborre et al. 2011). Recently, Yamaji et al. (2012) showed that plant lectins provide both broad and specific immunity to virus infection in tobacco. Therefore, the chitinase V-like protein may be one of the defense arsenals of leucaena against virus, pathogens and insects. Phylogenetic analysis of deduced amino acid of query 1673697 encoding chitinase class Ib of other plant species (Figure 3.3) showed that leucaena chitinase

class Ib was very similar, which was 91% homology with *Acacia koa*. Mostly, distingue from chitinase class Ib of other plants including *Arabidopsis*. In addition, another phylogenetic analysis of the leucaena chitinase V-like sequence with homologous sequences from other plant species (Figure 3.4) showed that the deduced amino acid sequence (339 residues) of query 1632644 was associated with *Cicer arietinum* and this distance from other plants including legume species G*lycine max.*

(e) *Identification of NB-LRR genes:*

Leucaena leucocephala is well known as a legume that highly resistant to pathogens. From extensive plant literature on disease resistance, it is apparent that effector-triggered immunity or ETI in plants is governed by NB-LRR proteins (Eitas and Dangl 2010). Therefore, we are interested in identifying genes encoding NB-LRR in leucaena. In the leucaena transcsriptome, we found 231 sequences longer than 100 bp that showed homology with NB-LRR genes (Table 3.4). Among the NB-LRR genes there are several subgroups, including, CC-NBS-LRR, NBS-LRR, TIR-NBS-LRR, LB-ARC and NB-ARC LRR etc. In the leucaena trancriptome there were 54, 78, and 97 sequences, which showed high homologies with CC-NBS-LRR, NBS-LRR and TIR-NBS-LRR genes, respectively. In addition, we also found one sequence encoding BED-finger-nbs resistance protein and a sequence for resistance gene analog NBS9. BED finger-nbs resistance protein has a zinc-finger domain. Another sequence has homology with NB-ARC-LRR, which contains an *Apaf*I domain. The NB-ARC interacts with a part of LRR, keeping the protein in close conformation in the absent of effectors produced by pathogens. In addition to the NB-ARC-LRR, there was one sequence that showed homology with TIR-NB-ARC-LRR, which contains a TIR (Tol/intermediating receptor) domain. The analysis of transcriptome sequence in leucaena showed that the NB-LRR genes expressed predominantly in shoots and the length of these sequences was longer than those in the root transcriptome. Shoot transcriptome has 86 sequences that are longer than 500 bp, and among these, 36 queries are larger than 1.0 kb (Table 3.5a). The root transcriptome has only 18 NB-LRR sequences that are longer than 500 bp (Table 3.5b). Query 1670725 was the longest sequence (3868 bp) encoding TIR-NBS-LRR type disease resistance protein in the leucaena transcriptome. It showed 66% identity and 71% similarity with TIR-NBS-LRR from *Glycine max.* (Figure 3.6).

Table 3.4: Number of sequences encoding NB-LRR genes in the leucaena transcriptome.

Query group name	Number of sequences
BED finger-nbs resistance protein	1
Cc-nbs resistance protein	6
CC-NBS-LRR disease resistance protein	2
CC-NBS-LRR protein	1
Cc-nbs-lrr resistance protein	32
cc-nbs-lrr resistance protein	13
NB-ARC domain containing protein	1
NB-ARC LRR protein	1
NB-LRR resistance-like protein RC68	1
NB-LRR resistance-like protein RGC22	1
NB-LRR disease resistance protein	1
NB-LRR disease resistance protein Rps1-k-2	5
NBS resistance protein	3
NBS resistance protein-like protein	1
NBS-containing resistance-like protein	21
NBS-LRR	1
NBS-LRR resistance protein	7
NBS-LRR disease resistance protein precursor	1
NBS-LRR disease-resistance protein scn3r1	1
NBS-LRR disease resistance protein	8
NBS-LRR resistance protein RGH2	2
NBS-LRR resistance protein-like protein	2
NBS-LRR resistance-like protein 1O	2
NBS-LRR resistance-like protein 4G	5
NBS-LRR resistance-like protein 4T	2
Nbs-lrr resistance protein	11
nbs-lrr resistance protein	1
TIR NB-ARC LRR protein	1
TIR-NBS-LRR	4

TIR-NBS-LRR disease resistance protein	60
TIR-NBS-LRR RCT1 resistance protein	11
Tir-nbs-lrr resistance protein	14
tir-nbs-lrr resistance protein	7
Resistance gene analog NBS9	1
Total	231

Table 3.5 (a): NB-LRR sequences (>1.0 kb) in the leucaena shoot transcriptome

Query_name	Query description	Length (bp)
1670725	TIR-NBS-LRR type disease resistance protein	3868
1673806	TIR-NBS-LRR type disease resistance protein	3450
1657587	NBS-LRR resistance-like protein 4G	3081
1670645	NBS/LRR resistance protein-like protein	2981
1661306	NBS/LRR resistance protein-like protein	2876
1655482	Cc-nbs-lrr resistance protein	2506
1647077	TIR-NBS-LRR-TIR type disease resistance protein	2344
564515	NBS resistance protein	2325
1653038	NBS resistance protein-like protein	2175
1646263	Cc-nbs-lrr resistance protein	2130
1656810	NBS-LRR resistance-like protein 4G	2084
1676289	Cc-nbs-lrr resistance protein	2003
1667480	cc-nbs-lrr resistance protein	1952
1658426	TIR-NBS-LRR-TIR type disease resistance protein	1908
1663927	TIR-NBS-LRR RCT1-like resistance protein	1891
1655483	Cc-nbs-lrr resistance protein	1813
1674124	TIR-NBS-LRR-TIR type disease resistance protein	1694
1656292	TIR-NBS-LRR RCT1-like resistance protein	1622
1663415	TIR-NBS-LRR RCT1-like resistance protein	1605
1636745	Tir-nbs-lrr resistance protein	1583
1675232	Nbs-lrr resistance protein	1563

1671802	Cc-nbs-lrr resistance protein	1521
1673248	Cc-nbs-lrr resistance protein	1428
1675808	Cc-nbs-lrr resistance protein	1395
1652656	Cc-nbs-lrr resistance protein	1392
1652657	Cc-nbs-lrr resistance protein	1354
1656174	Cc-nbs-lrr resistance protein	1354
1670930	Cc-nbs-lrr resistance protein	1315
1636390	Cc-nbs-lrr resistance protein	1103
1649203	Cc-nbs-lrr resistance protein	1103
1664932	Cc-nbs-lrr resistance protein	1078
1643703	Cc-nbs-lrr resistance protein	1076
1666958	Cc-nbs-lrr resistance protein	1070
1667524	Cc-nbs-lrr resistance protein	1067
1653044	Cc-nbs-lrr resistance protein	1054
1666545	Cc-nbs-lrr resistance protein	1010

Table 3.5 (b): NB-LRR sequences (>500 bp) in the leucaena root transcriptome

Query_name	Query description	Length (bp)
1805102	Tir-nbs-lrr resistance protein	1619
1553454	TIR-NBS-LRR-TIR type disease resistance protein	1403
986992	NBS-LRR resistance-like protein 4G	1253
1808471	NBS/LRR resistance protein-like protein	945
1468634	TIR-NBS-LRR RCT1-like resistance protein	880
1806996	Cc-nbs-lrr resistance protein	855
1803355	Cc-nbs-lrr resistance protein	808
1804433	tir-nbs-lrr resistance protein	806
1804277	Cc-nbs-lrr resistance protein	805
485770	aminotransferase	631
1578104	tir-nbs-lrr resistance protein	611
1808845	hypothetical protein CKR_0758	602
1806243	Cc-nbs-lrr resistance protein	586

1808534	TIR-NBS-LRR type disease resistance protein	563
1512352	Tir-nbs-lrr resistance protein	561
1494276	cc-nbs-lrr resistance protein	528
1576969	TIR-NBS-LRR type disease resistance protein	523
1806535	NBS-LRR resistance-like protein 4G	519

```
Query  3399  NVINAYAPQVGTEAHLKEKFWEDLEGLIHSIPFTEKIFIGGDCNGHVGKEAGQYAWAHSG  3220
             +VI+AYAPQVG++   K  FWEDLE L+  IP  +KIF+GGD NGHVG+E    Y    H G
Sbjct  597   HVISAYAPQVGSDEQHKISFWEDLESLVQGIPLGDKIFLGGDLNGHVGREVTGYGSIHGG  656

Query  3219  FSFGEINNEGQSIFEFSLAYNFKIVNTCFKKREEHLITYKSRVHKSQIDFFLVRSHDRRL  3040
              FG IN EG++I +FS  ++  I NTCFKKR+EHLITYKS +  SQIDFFL+R  DR+
Sbjct  657   HGFGVINAEGKTILDFSSTFDLLIANTCFKKRDEHLITYKSGMTSSQIDFFLLRRVDRKF  716

Query  3039  CTNCKVIPGDRVTTQHRLMVLDMHIKCRKKKCRHVSNTIVKWWQLKGEKRETFKKTMLNE  2860
             C NCK+IPG+ +TTQHR++VLD  ++ + +K  H  N    +WW++KGE++ +F + +   E
Sbjct  717   CINCKIIPGESLTTQHRVLVLDFRVEQKLRKRHHTKNPRTRWWRMKGEEQRSFLRRVGEE  776

Query  2859  GVWEEQENANIMWKEMAEKVRTIVKAILGESKGFGRRDKKTWWWNEDVQEKVKNKRECFK  2680
               W+   +A  MW+EMAE +R   K   GESKG G RDK++WWWN  +QEK+K KRECFK
Sbjct  777   AKWDGNGSAEEMWREMAEVIRRTAKESFGESKGIGPRDKESWWWNASIQEKIKIKRECFK  836

Query  2679  AIHLC--NTEN*EKYRLAKKETKKAVSAARFKAFEEFYKELGTKSGERKIYKIARDRERKS  2503
                LC N +N EKY+ AKKETK AVS AR +A+E  Y+ L TK GE+ IY+IA+ RER++
Sbjct  837   EWSLCRNVDNWEKYKAAKKETKVAVSEARTRAYEGLYQSLDTKEGEKGIYRIAKSRERRT  896

Query  2502  RDLDQVRYIKDEEGKVLVADSDIK*WWETYFYKLFKDEREGSSYELEDLTREVEPNSAFY  2323
             RDLDQV+ IKD++ +VL  +  I    W++YFY+LF + ++            TRE + N  +Y
Sbjct  897   RDLDQVKCIKDKDREVLAQEEKINERWKSYFYELFNEGQKTLPSLGRLCTREEDQNFNYY  956

Query  2322  RRIRVGKVKEALKKIENSKFIGPDGIPIEVWKCLGEVGVVCLAKLFNVILSSKKMPDDWR  2143
             RRIR  +VKEALK+++N + +GPD IPIEVWK LG  G+  L KLF   IL SKKMPD+WR
Sbjct  957   RRIRDFEVKEALKQMKNGRAVGPDNIPIEVWKGLGGKGINWLTKLFYEILRSKKMPDEWR  1016

Query  2142  KSTLVPIYKNKG  2107
             KSTLVPIY+ KG
Sbjct  1017  KSTLVPIYRIKG  1028
```

Figure 3.6: BLASTX analysis showing 66% identities of query 1670725 with the *Glycine max* TIR-NBS-LRR type disease resistance protein

(f) *WRKY transcription factors:*

WRKY transcription factors are one of the largest families of transcriptional regulators, which may be involved resistance to various biotic and abiotic stresses (Figure 3.7)
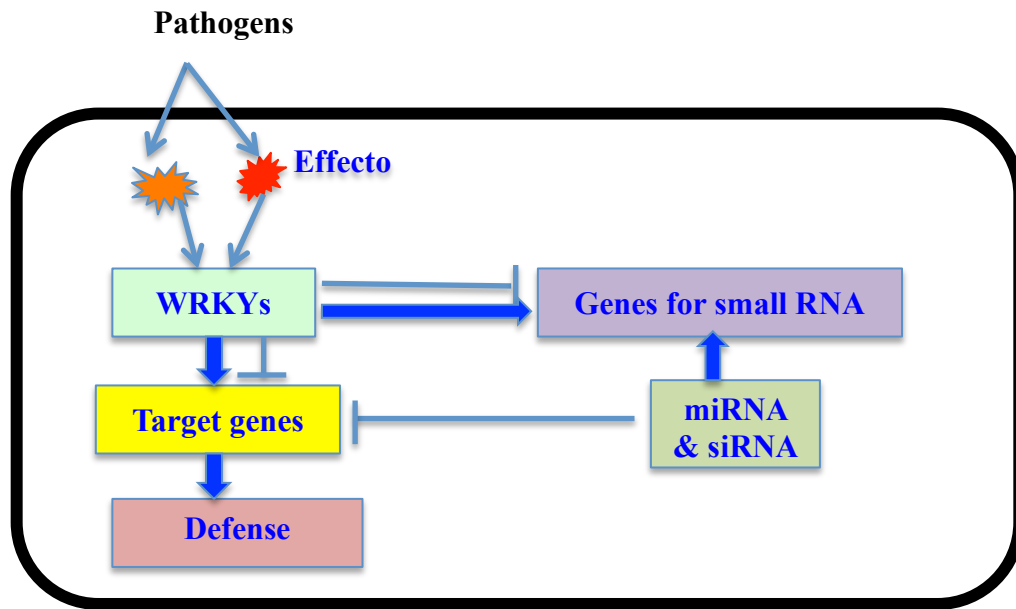
Figure 3.7: The role of WRKY in inducing defense response in plants (based on Pandey and Somssich 2009).

The number of WRKY genes identified in other recently sequenced plant genomes is 74 in *Arabidopsis*, 109 in rice, 66 in papaya (*Carica papaya*), 104 in poplar (*Populus spp.*), and 68 in sorghum (*Sorghum bicolor*). In leucaena, we found 145 sequences encoding WRKY transcriptions factors expressing in roots and 223 in shoots (Table 3.6). Thus the number of WRKY in leucaena appears to be more than in other plants. These transcription factors may be involved in reprograming of defense response in plants (Pandey and Somssich 2009). From Table 3.6, it can be seen that some WRKY transcription factors appear to be specific for expression in shoots or roots; some of them expressed in shoots while others expressed in roots. The WRKY transcription factors that expressed in shoots but not in roots include WRKY DNA-binding protein 49, WRKY transcription factor 13, 19, 35, 39, 51, 57, 7, WRKY-like drought-induced protein, WRKY1, WRKY32 protein, WRKY43 and WRKY54. The WRKY transcription factors 31, 49, 56, WRKY2, WRKY53 and WRKY6 were identified only in roots. WRKY transcription factors 2 and 20 expressed in both shoots and roots.

Among 367 sequences encoding WRKY family in leucaena, 30 queries were longer than 1.0 kb (Table 3.7) and the longest sequence 1669410 (3845 bp) showed 53 % identity and 70% similarity with a WRKY from *Arachis hypogaea* (Figure 3.8). Only two sequences larger than 1.0 kb were found in the root transcriptome. Query 762 (1850 bp) encoding a WRKY transcription factor 2-like protein showed homology with a WRKY from *Glycine max*, while query 986527 (1000 bp) encoding a WRKY transcription factor 20-like protein showed homology with a WRKY from *Solanum lycopersicum* (Figures. 3.9a. and 3.9b).

Table 3.6: WRKY transcription factors in leucaena transcriptome

WRKY	In root	In shoot
WRKY DNA-binding domain superfamily protein	3	2
WRKY DNA-binding protein 49	0	1
WRKY transcription factor	12	27
WRKY transcription factor 1	2	4
WRKY transcription factor 11	9	4
WRKY transcription factor 12	2	4
WRKY transcription factor 13	0	1
WRKY transcription factor 14	1	1
WRKY transcription factor 15	1	1
WRKY transcription factor 17	5	5
WRKY transcription factor 19	0	1
WRKY transcription factor 2	6	7
WRKY transcription factor 20	17	24
WRKY transcription factor 21	2	5
WRKY transcription factor 22	5	9
WRKY transcription factor 23	2	5
WRKY transcription factor 28	3	4
WRKY transcription factor 3	6	7
WRKY transcription factor 31	1	0
WRKY transcription factor 32	14	14
WRKY transcription factor 33	4	10
WRKY transcription factor 35	0	2
WRKY transcription factor 39	0	4
WRKY transcription factor 4	4	5
WRKY transcription factor 40	1	6
WRKY transcription factor 41	2	8
WRKY transcription factor 42	2	2
WRKY transcription factor 44	4	8
WRKY transcription factor 47	2	3
WRKY transcription factor 48	2	5
WRKY transcription factor 49	1	0
WRKY transcription factor 51	0	1

WRKY transcription factor 56	1	0
WRKY transcription factor 57	0	3
WRKY transcription factor 6	11	10
WRKY transcription factor 65	5	4
WRKY transcription factor 69	1	1
WRKY transcription factor 7	0	3
WRKY transcription factor 70	2	2
WRKY transcription factor 72	2	2
WRKY transcription factor 75	1	1
WRKY transcription factor 9	4	6
WRKY-A1244	1	1
WRKY-like drought-induced protein	0	2
WRKY1	0	1
WRKY2	1	0
WRKY32 protein	0	2
WRKY43	0	1
WRKY53	1	
WRKY54	0	1
WRKY6	1	0
WRKY78	1	1
WRKY86	1	2

Table 3.7: WRKY transcription factors sequences (>1 kb) in leucaena transcriptome

Query name	Query description	Length (bp)
1669410	WRKY transcription factor 2-like	3845
1662388	WRKY transcription factor 3-like	2975
1674485	WRKY transcription factor 20-like	2113
1669652	WRKY transcription factor 40-like isoform 1	2086
1660961	WRKY transcription factor 17-like	2009
1663500	WRKY transcription factor 1-like	1975
1672286	WRKY transcription factor	1769
1672518	WRKY transcription factor 33	1711
1669544	WRKY transcription factor 33	1681
1661694	WRKY transcription factor 70-like	1578
1665277	WRKY transcription factor	1563
1670406	WRKY transcription factor	1526
1656679	WRKY transcription factor 4	1489

1660763	WRKY transcription factor 17-like	1481
1672186	WRKY transcription factor 13	1454
1665160	WRKY transcription factor 48-like	1452
1634826	WRKY transcription factor 17-like	1424
1653533	WRKY transcription factor 6-like	1410
1659096	WRKY transcription factor 44	1378
1676074	WRKY transcription factor 7	1317
1666471	WRKY transcription factor 65-like	1229
1666276	WRKY domain class transcription factor	1213
1675618	WRKY transcription factor	1165
1655360	WRKY transcription factor 6	1162
1665660	WRKY transcription factor 41-like	1129
1672240	WRKY transcription factor 3-like	1069
1675210	WRKY transcription factor 32	1062
1671114	WRKY transcription factor 48-like	1049
1669866	WRKY transcription factor 41-like	1044
1661090	WRKY transcription factor 44-like isoform X1	1016

```
Query   2606   DTRSAYFTIPPGLSPTTLLESPVFLSNSPAQPSPTTGKFPFISNGNFQGSELNSGAPEKT   2427
               D +S Y TIPPGLSPTTLL+SPVFL+NS AQPSPTTGKF F++NG  + SEL+S APEK
Sbjct   125    DIQSPYLTIPPGLSPTTLLDSPVFLANSLAQPSPTTGKFLFMANGIMRNSELSSDAPEKC   184

Query   2426   KDNNFGDIYASSFAFKTTTTDLGSFYNGASRKM-NLTTLPEQYLTA-EVSA---------   2280
               KDN F DIY SSFAFK  T D GSFY+GA RKM N TTLP+Q L    EVSA
Sbjct   185    KDNGFDDIYTSSFAFKRAT-DSGSFYHGAGRKMINPTTLPQQSLPGIEVSAQSENSFQSQ   243

Query   2279   ------------------PDMTDPPTQNDS---------RA-----------SVEEPAD   2217
                                 D  + P Q D+         RA            +EE  D
Sbjct   244    SVDAVKAQTENKSGFRLQADFAESPPQKDNGIKMFSADQRAFDVVGGGNEHSTPIEEQVD   303

Query   2216   EEGGQRGNEDSQAAGVGGTLSEDGYNWRKYGQKQVKGSEYPRSYYKCTHTNCTVKKKVER   2037
               E G QRGN DS A+GVGG  SEDGYNWRKYGQKQVKGSEYPRSYYKCTH NC VKKKVER
Sbjct   304    E-GDQRGNGDSMASGVGGAPSEDGYNWRKYGQKQVKGSEYPRSYYKCTHPNCQVKKKVER   362

Query   2036   SHEGHITEIIYKGNHNHPKPPTNRRSAIGSVNPLGDMQADVSESTEPHGGGDGELGWAST   1857
               SHEGHITEIIYKG HNHPKPP NRRS IG VN   DMQ D  E   EPH GGDG+LGWA+
Sbjct   363    SHEGHITEIIYKGTHNHPKPPPNRRSGIGLVNLHTDMQVDHPEHVEPHNGGDGDLGWANV   422

Query   1856   KRGNIARNGDWKHENLELTSSAASVGPEFGNQS-TNLQGQNGTQFESEEAVDASSTFSN-   1683
               ++GNIA    WKH+NLE  SS ASVGPE+ NQ   NLQ QNGT F+S EAVDASSTFSN
Sbjct   423    QKGNIAGAASWKHDNLEAASS-ASVGPEYCNQQPPNLQTQNGTHFDSGEAVDASSTFSNE   481

Query   1682   DDDDDRATHGSVSVGYDGEGDESESKRRKLESYPNTAELSGATRAIREPRVVVQTTSEVD   1503
               +D+DD+ THGSVS+GYDGEGDESESKRRKLESY   AELSGATRAIREPRVVVQTTSEVD
Sbjct   482    EDEDDQGTHGSVSLGYDGEGDESESKRRKLESY---AELSGATRAIREPRVVVQTTSEVD   538

Query   1502   ILDDGYRWRKYGQKVVKGNPNPRSYYKCTNVGCTVRKHVERASHDLKSVITTYEGKHYHD   1323
               ILDDGYRWRKYGQKVVKGNPNPRSYYKCTN GCTVRKHVERASHDLKSVITTYEGKH HD
Sbjct   539    ILDDGYRWRKYGQKVVKGNPNPRSYYKCTNAGCTVRKHVERASHDLKSVITTYEGKHNHD   598

Query   1322   VPaarnsnqgqnaaassaatgqasSVIQSHRAEASQVHNSIGRLDRPVLGLGTFNFPGSGP   1143
               VPAAR S+  NA AS+A  GQAS    HR E S+VHN IGRL+RP   LG+FN
Sbjct   599    VPAARASSHVNANASNAVPGQASLQTHVHRPEPSEVHNGIGRLERP--SLGSFNL-----   651

Query   1142   GPGRPQLGPSPGFSFGIGMNQSGFPNLATMAALGPAHAKLPVMPIHPFLP----------   993
                PGR QLGPS GFSF  GMNQS    NL  M+ LG A AKLPVMP+H FL
Sbjct   652    -PGRQQLGPSHGFSF--GMNQSMLSNL-VMSGLGHAQAKLPVMPVHSFLAAHQQQQQHQQ   707

Query   992    ----NQRPPN-MGFMLPKGEANLEPIPDR-GLNMPAGSSVYQDIMSRMPLGPHM   849
                   QR  N +GFMLPKGE N+E IP+R GLN+  GSSVYQ+IMSRMPLGPHM
Sbjct   708    QQNQQQRAANDLGFMLPKGEPNVEAIPERGGLNLSNGSSVYQEIMSRMPLGPHM   761
```

Figure 3.8: BLASTX analysis showing 53 % identities of query 1669410 with the *Arachis hypogaea* WRKY transcription factor 2-like

```
Query   999   PPGLSPSSFLESPVLLSNVKVSSRFLISGFS   907
              PPGLSPSSFLESPVLLSN+K         FS
Sbjct   61    PPGLSPSSFLESPVLLSNIKAEPSPTTGSFS   91
```

Figure 3.9 (a): BLASTX analysis showing 68% identities of query 986527 (1000 bp) WRKY transcription factor 20-like with WRKY from *Solanum lycopersicum*

```
Query  1545  KGGISDCRPLVPYQAPLDFSIPAEFPKVHKMKR-EVHSYDDVRIMQDAIYNANNLEMQMH  1369
             KGG +   L   Q PLDFS  A+F K H +K  EV++Y+D++++ D I NANN+EM M
Sbjct  160   KGGNRESHLLAQVQPPLDFSFRADFSKGHSVKNSEVNAYNDMKMVNDVILNANNVEMPMS  219

Query  1368  RSEEVADKGFLPKKA--NKDTGQHSLVEENKRETSYLMGMVRTSEDGYHWRKYGQKQVKG  1195
              SEEV+D+ LPK       +D G     E  ++E S+  G VRTSEDGY+WRKYGQKQVKG
Sbjct  220   GSEEVSDESALPKNTINGEDFGGQPASEGEQKEASHTTGAVRTSEDGYNWRKYGQKQVKG  279

Query  1194  SEYPRSYYKCTHPNCQVKKKVERSLDGQITEIIYKGTHNHGKPQPSRRPSLGSALSTDEM  1015
             SEYPRSYYKCT P CQVKKKVERS DGQITEIIYKG HNH +P P  R    S+LSTDE+
Sbjct  280   SEYPRSYYKCTQPKCQVKKKVERSHDGQITEIIYKGAHNHAQPHPGHR---ASSLSTDEV  336

Query  1014  LDAGEGGGTFIKADGG--WRNVHSGVKDIKQNLDWKADNQERTSSSSVVTELSDLVSTNK  841
              D  G  T  K +GG  WRN+ +G+++ KQ+ DWKAD QERT ++S VTELSD +STN
Sbjct  337   SDMA-GDSTLAKIEGGYVWRNIQTGLRETKQSFDWKADGQERTPTTSAVTELSDPISTNN  395

Query  840   GKPISMFESEDTRELSSTLASKDDDEDVATQGSLSLEGDANDEEPDSKRRKKESCLIEAN  661
              K + M ESEDT ELSSTLAS D DED   Q +S E +A ++E DSKRRKKES  +E N
Sbjct  396   AKSLCMLESEDTPELSSTLASHGDEDGTAQALVSAEDEAENDELDSKRRKKESYAVEPN  455

Query  660   FA-SRAVREPRIVVQIESEVDILDDGYRWRKYGQKVVKGNPNPRSYYKCTSAGCLVRKHV  484
             +RAVREPR+VVQIES+VDILDDGYRWRKYGQKVVKGNPNPRSYYKCTSAGC+VRKHV
Sbjct  456   LPPTRAVREPRVVVQIESDVDILDDGYRWRKYGQKVVKGNPNPRSYYKCTSAGCMVRKHV  515

Query  483   ERASHNLKYVFTTYDGKHNHEVPTARSNNQMSSTG-----------ACALPKPESQVQN  340
             ERAS NLKYV TTY+GKHNHEVPTAR+NNQ++S+         A  LPKPE+
Sbjct  516   ERASQNLKYVLTTYEGKHNHEVPTARTNNQVNSSDGGLPPNGANGQVALTLPKPETHQTL  575

Query  339   TAPHFDRKPEFS------SFMGGFGDAMKFGSTSI----YP---KTLPYGSYD-------  220
              HFDRKPEFS       S +G F + MKFG +++    YP     T+PYGSY
Sbjct  576   FGHHFDRKPEFSNEFLRASLVGSFSNDMKFGPSTLCQMKYPSLNNTMPYGSYGLNHEHCT  635

Query  219   --------TVFPDFPISLPLNLPSSAGFNLNCVTPM----------------GFLRPKQ  115
                     ++FPDFP+ LPLNLPSS G N NCV PM                GFLRPKQ
Sbjct  636   APQAGSIASMFPDFPMPLPLNLPSS-GLNFNCVKPMNRVQSFLSGQQVKDIDAGFLRPKQ  694

Query  114   EQKDASI  94
             EQKD ++
Sbjct  695   EQKDDTM  701
```

Figure 3.9 (b): BLASTX analysis showing 57% identities of query 762 (1850 bp) WRKY transcription factor 2-like with WRKY from *Glycine max*

(g) Early responsive to dehydration (ERD):

The analysis of leucaena transcriptome showed that 109 sequences encoding different members of ERD family. Some of them expressed only in root and some others expressed only in shoot. Among 71 ERD-related genes in shoot transcriptome, nineteen sequences were larger than 500 bp (Table 3.8) and the longest ERD query 1666699 (2054 bp) showed 81% identity with dehydration responsive element-binding protein from *Sophora davidii* (Figure 3.10). Surprisingly, all sequences encoding ERD in root were smaller than 500 bp and the longest one (query 999375) was only 492 bp.

Table 3.8: ERD sequences (>500 bp) in the leucaena transcriptome.

Query name	Description	Length (bp)
1666699	Dehydration responsive element-binding protein	2054
1664244	Dehydration-responsive protein RD22	2032
1666289	Dehydration-responsive element-binding protein 3-like	1574
1669243	Dehydration-responsive protein RD22	1498
1658698	Dehydration-responsive protein RD22-like	1471
1670926	Dehydration responsive element binding protein	1294
1662585	Dehydration-responsive element binding protein	1234
1658240	Protein dehydration-induced 19-like	1171
1662544	Protein dehydration-induced 19-like	1169
1674880	Protein dehydration-induced 19-like	925
1665761	Protein dehydration-induced 19 homolog 3-like	901
1675900	Dehydration responsive element-binding protein 3	880
1658446	Protein dehydration-induced 19 homolog 3-like isoform 1	875
1640014	Dehydration-responsive element-binding protein 3-like	730
1667829	Dehydration-responsive protein RD22-like	709
1676249	Dehydration-responsive element binding protein	686
1663912	Protein dehydration-induced 19 homolog 5-like	663
1675288	Protein dehydration-induced 19 homolog 3-like	661
1662485	Protein dehydration-induced 19 homolog 3-like isoform 1	639

Figure 3.10: BLASTX analysis showing 81% identities of query 1666699 (2054 bp) dehydration responsive element-binding protein from *Sophora davidii*

```
Query 726   LEKAHAKGDGSKSLAEKLARWKEYNTQIDSCNDADKPIRKVPAkgskkgcmkgkgGPENS  547
            L KA +KGDGSKSLA+ LA+WKEYN Q+DS NDADKPIRKVPAKGS+KGCMKGKGGPENS
Sbjct 2     LAKAQSKGDGSKSLAKILAKWKEYNAQLDSSNDADKPIRKVPAKGSRKGCMKGKGGPENS  61

Query 546   RCNYRGVRQRTWGKWVAEIREPNRGSRLWLGTFPTAIGAALAYDEAARAMYGSFARLnfp  367
            RCNYRGVRQRTWGKWVAEIREPNRG+RLWLGTFPTAIGAALAYDEAARAMYG+ ARLNFP
Sbjct 62    RCNYRGVRQRTWGKWVAEIREPNRGNRLWLGTFPTAIGAALAYDEAARAMYGTCARLNFP  121

Query 366   nvsvsnfsnesSRDSLVTTQSDSA----VPTESLMSPTNSG  256
            N+SVS+FS ESS+DSL  QS S+     TES++SP NSG
Sbjct 122   NISVSSFSEESSKDSLGANQSGSSETVLANTESMVSPINSG  162
```
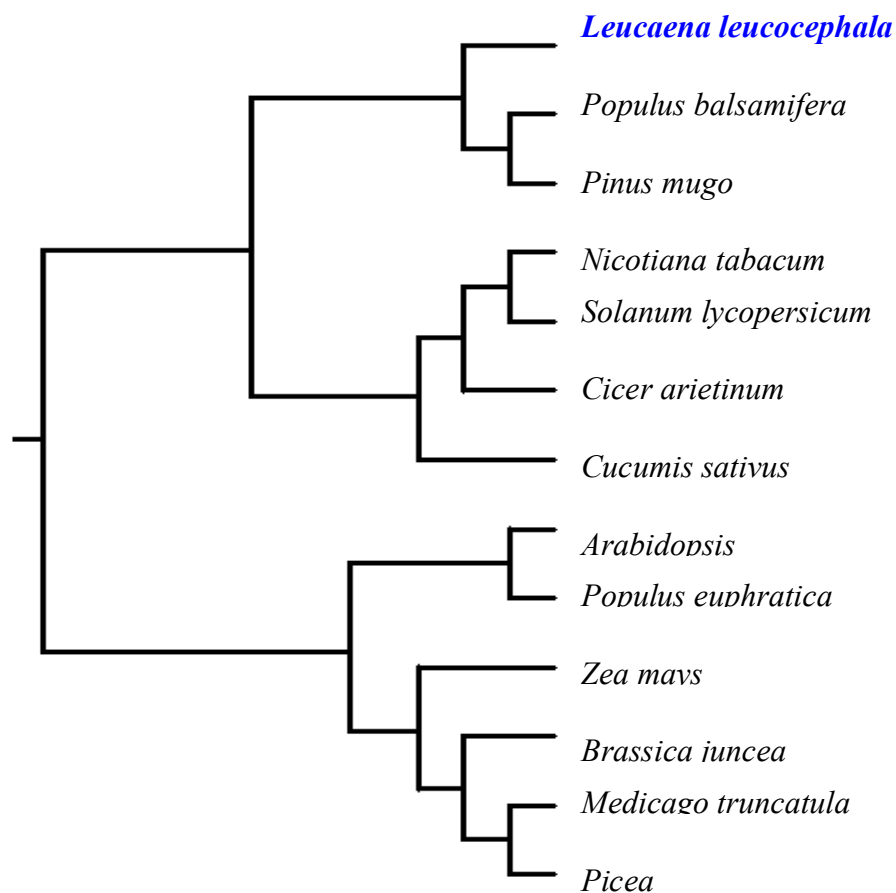
Figure 3.11 : Phylogenetic tree of ERD gene, query 1666699 with other ERD from various plant species

(h) Identification of UV resistance-related genes

A total 96 sequences encoding different UV-related genes were found in the leucaena transcriptome. Among these, 13 and 4 queries, which were longer than 500 bp in the shoot and root transcriptomes, respectively (Table 3.9). All longer sequences in root encoded DNA repair helicase UVH3 and UVH6. BLASTX analysis of the longest query 1809443 (4355 bp) of root transcriptome showed 63% identity and 73% similarity with DNA repair protein UVH3-like from *Glycine max*. The longest query 1659895 (1902 bp) in shoot showed 74% identity and 83% similarity with UV radiation resistance-associated gene protein-like from *Cicer arietinum* (Figure 3.12).

Table 3.9: UV radiation resistance-related sequences (>500 bp) in the leucaena transcriptome

Query name	Description	Length (bp)	Organism
1809443	DNA repair protein UVH3-like	4355	Root
1659895	UV radiation resistance-associated gene protein	1902	Shoot
1805786	DNA repair helicase UVH6-like isoform X1	1299	Root
1666396	DNA repair helicase UVH6-like	1034	Shoot
1803623	DNA repair helicase UVH6-like	1031	Root
1674260	DNA repair helicase UVH6-like isoform X1	933	Shoot
1663183	Ultraviolet-B receptor UVR8-like	808	Shoot
1653855	DNA repair helicase UVH6-like	799	Shoot
1656143	Helicase, belonging to UvrD family	794	Shoot
136407	DNA repair helicase UVH6-like isoform X1	790	Shoot
1644832	DNA repair helicase UVH6-like	726	Shoot
1659245	DNA damage-binding protein 1a	645	Shoot
1665775	UV radiation resistance-associated gene protein-like	612	Shoot
482871	UvrD/REP helicase	606	Shoot

```
Query  740   NVSMHSRIAKEDSTKKKEQLGVEVESLLVAGTALSVSRRRLQESNRLLFEEKGYVQLSNL   561
             N+S+ SR+AKED  K++EQL   V+SLLVAG  LSV+ R LQESNRLL EE GYV+L NL
Sbjct   88   NMSIRSRLAKEDVNKQEEQLSGAVQSLLVAGGTLSVTSRNLQESNRLLSEENGYVRLRNL   147

Query  560   QKMLRLRQQYMVTQISMLYPVKISVGPAQEQELEAYPAGSLAGNSAELKPVNQGSLTILG   381
             QKMLR+RQQYM +QISMLYPVK+ VGPAQEQELEAYP GS AG   ELKPVNQGSL I G
Sbjct  148   QKMLRMRQQYMTSQISMLYPVKLLVGPAQEQELEAYPLGSPAGTPPELKPVNQGSLMIQG   207

Query  380   LHLTMLSFKKMSFFTD   333
             LHL+M SF+KMSFFTD
Sbjct  208   LHLSMQSFRKMSFFTD   223
```

Figure 3.12: BLASTX analysis showing 74% identities of the shoot query 1659895 (1902 bp) with UV radiation resistance-associated gene protein-like from *Cicer arietinum*

(i) Identification of genes encoding kinases and phosphatases

In addition to WRKY transcription factors discussed above, there are a number of kinases and phosphatases that are involved in regulation of various physiological functions in plants such as resistance to biotic and abiotic stresses. These include serine/ threonine kinases, histidine kinases, tyrosine kinases, MAP kinases, and tyrosine phosphatases. Moreover, leucaena contains large quantities of mimosine, which may also provide resistance to insect-pests and diseases. Therefore, we are interested in genes for mimosine metabolism in leucaena. One gene for a C-N lyase encoding a mimosinase and another gene for *O*-acetylserine (thiol) lyase involved in mimosine biosynthesis in leucaena have been isolated in our lab. *O*-acetylserine (thiol) lyase is also a cysteine synthase required for cysteine biosynthesis. There may be additional copies of mimosinase and cysteine synthase genes, which we wanted to identify in the leucaena transcriptome sequences.

Serine/ threonine kinases

Some serine/ threonine kinases are receptor proteins that interacted with other proteins to affect processes like disease resistance and development regulations (Goring and Walker, 2004). Xu and Deng (2010) identified defense-related gene Ser/ Thr kinase gene in wild crop "chestnut rose". Similarly, Cao et al. (2011) found that serine/ threonine kinase confering by gene *Stpk*-V, which resistant to powdery mildrew caused by *Blumeria graminis f. sp. tritici* (Bgt) in wheat.

Total of 636 sequences (>500 bp) encoded different types of Ser/ Thr kinases (Table 3.10). Among these, 338 are Ser/ Thr protein kinases, which may be located in cytosol and function as a part of regulatory pathways related to signal transductions. Leucine-rich repeat Ser/ Thr protein kinases were 183 sequences. Likely NB-LRR, these genes have LRR domain and may function to formation of complex with other receptor proteins (Afzal et al. 2008). We found 29 sequences homology to receptor Ser/ Thr kinase, which may have membrane spanding receptor domain and Ser/ Thr kinase domain in cytosol.

Table 3.10: Serine/ threonine kinase sequences (>500 bp) in the leucarna transcriptome

Query description	Sequences
ATP binding/protein serine/threonine kinase	1
Calcium/calmodulin-dependent serine/threonine-protein kinase 1-like	2
CBL-interacting serine/threonine-protein kinase	29
Checkpoint serine/threonine-protein kinase BUB1-like	2
Extracellular serine/threonine protein kinase Fam20C	1
G-type lectin S-receptor-like serine/threonine-protein kinase	42
Inactive leucine-rich repeat receptor-like serine/threonine-protein kinase	2
Inactive serine/threonine-protein kinase scy1-like isoform X1	1
Leucine-rich repeat (LRR) receptor-like serine/threonine-protein kinase	183
Microtubule-associated serine/threonine-protein kinase	2
Receptor serine/threonine kinase	29
Receptor-like cytosolic serine/threonine-protein kinase	4
Serine/threonine protein kinase	338
Total	636

Histidine kinase

Plants have a unique regulatory mechanism involving a two-component histidine kinase system, which is absent in animals. A typical histidine kinase in plants contains one conserved histidine kinase domain and a response regulator. In *Arabidopsis*, eleven histidine kinases have been identified. Histidine kinases are known to mediate responses to hormone ethylene and cytokinin in *Arabidopsis* (Bleecker and Kendle 2000, Higuchi et al. 2004). Therefore, histidine kinases are generally involved in stress-related signaling pathways. One of the histidine kinases in *Arabidopsis*, AHK5, has been shown to be involved in defense against infection by *Pseudomonas syringae* pv DC3000 and the fungal pathogen *Botrytis cinerea* (Pham et al. 2012). The AHK5 gene is also involved in modulating responses to salt stress; expression of AHK5 makes the plant sensitive to high salt (Pham et al. 2012).

In leucaena, we found 33 sequences (>500 bp) that encoded histidine kinases (Table 3.11). However, most of the longer histidine kinase sequences were in the shoot transcriptome.

The longest sequence, query 1673656 (3978 bp), showed 71% identity and 78% similarity, and 68% identity and 78% similarity with histidine kinase 5-like protein from *Glycine max* and *Cicer arietinum,* respectively. Therefore, leucaena histidine kinase 5-like protein is close to that of *Glycine max* and *Cicer arietinum* (Figure 3.13).

Table 3.11: Histidine kinase sequences (>500 bp) in the leucaena transcriptome

Query name	Description	Length (bp)	Shoot/ Root
1596066	Signal transduction histidine kinase	807	Root
1489929	Histidine kinase	641	Root
1809134	Histidine kinase 5-like	604	Root
1540804	Sensor histidine kinase	529	Root
1673656	Histidine kinase 5-like	3978	Shoot
1661659	Histidine kinase 3-like	2073	Shoot
1658939	Histidine kinase 2	1987	Shoot
1650520	Histidine kinase 1-like	1735	Shoot
1666493	Histidine kinase 3-like	1293	Shoot
1675474	Histidine kinase 2-like isoform X2	1231	Shoot
1663492	Histidine kinase 4-like	1212	Shoot
1659357	Histidine kinase	1203	Shoot
1674734	Histidine kinase cytokinin receptor	1182	Shoot
1646154	Multi-sensor signal transduction histidine kinase	1065	Shoot
1673194	DNA-binding response regulator/sensor histidine kinase	987	Shoot
1662605	Histidine kinase 3-like	929	Shoot
1640137	Histidine kinase 2-like isoform X2	755	Shoot
1669123	Histidine kinase 4-like	750	Shoot
1655172	Two-component sensor histidine kinase bacteria	696	Shoot
1666520	Histidine kinase 3-like	662	Shoot
1632269	Hybrid signal transduction histidine kinase M-like	636	Shoot
12319	Signal transduction histidine kinase	625	Shoot
47243	Receptor histidine kinase	603	Shoot
138168	Histidine kinase of the competence regulon, ComD	597	Shoot

1640138	Histidine kinase 2-like isoform X2	592	Shoot
1665319	Histidine kinase 4-like	590	Shoot
480696	Multi-sensor signal transduction histidine kinase	573	Shoot
1636601	Two-component sensor histidine kinase bacteria	566	Shoot
1632177	Histidine kinase	547	Shoot
531181	Sensor histidine kinase	515	Shoot
544175	Sensor histidine kinase	514	Shoot
556929	Histidine kinase 1-like isoform X2	512	Shoot
1632515	Histidine kinase osmosensor protein	502	Shoot
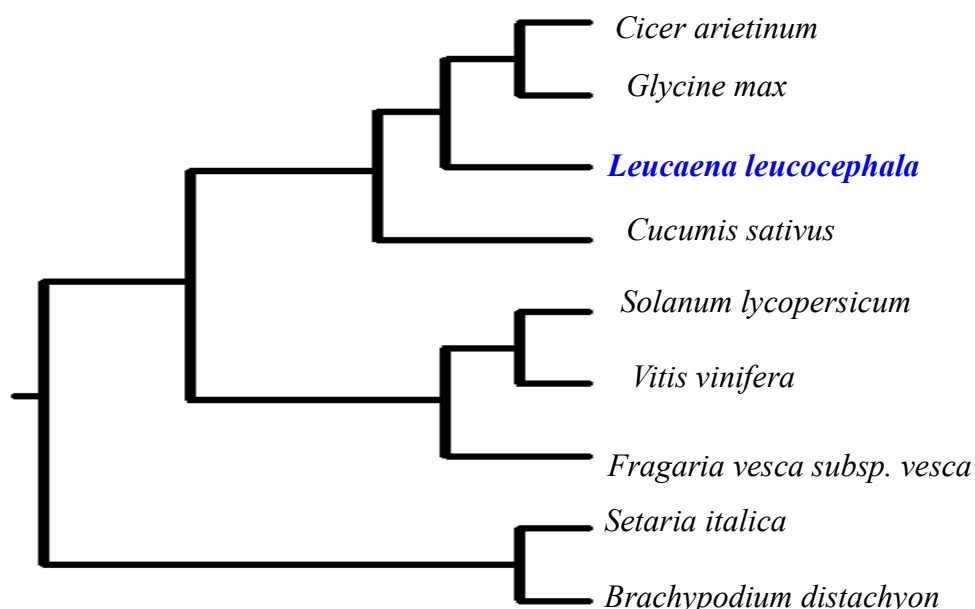
Figure 3.13: Phylogenetic tree showing the relation of leucaena histidine kinase 5-like protein with those of other plant species.

Tyrosine kinases

Tyrosine kinases catalyze the phosphorylation of tyrosine residues in proteins. Tyrosine kinase activity in the nucleus involves cell-cycle control and properties of transcription factors. Tyrosine kinase may be involved in cellular growth and reproduction. In contrast to Ser/Thr phosphorylation, Tyr phosphorylation is less common. Tyr phosphorylated proteins have been identified in higher plants, and the roles of Tyr kinases in some physiological responses has been shown. Protein tyrosine kinases (PTKs) in plants can be divided into two groups based on

their specificity: Tyr-specific PTKs and dual-specificity PTKs (DsPTKs). Protein Tyr kinases (PTKs) phosphorylate only Tyr, while DsPTKs phosphorylate Ser/Thr and Tyr. Several PTKs from *Arabidopsis* and other species have been characterized (Ghelis et al. 2008). In plants, in addition to PTKs and dsPTKs, there are receptor protein kinases (RPK), which are located on the plasma membrane with cytoplasmic kinase domains. Compared to the animal system, the role of protein Tyr phosphorylation is less documented for plants. Tyr phosphorylated proteins have been detected in carrot, *Mimosa pudica*, and *Arabidopsis* (Kameyama et al. 2000; Huang et al., 2003; Sugiyama et al., 2008). Tyr phosphorylation has been shown to modulate embryogenesis in coconut, and Tyr phosphorylation of proteins in rice is affected by copper (Islas-Flores et al. 1998; Hung et al. 2007). Tyr phosphorylation has been implicated in disease-resistance signaling in plants, although the evidence is indirect. For instance, the plant pathogen *Pseudomonas syringae* injects the virulence factor HopPtoD2, which is a protein tyrosine phosphatase, into its host cytoplasm. By removing the phosphates from the host proteins, the pathogen suppresses hypersensitive reaction in the plant (Bretz et al. 2003, Espinosa et al. 2003, Underwood et al. 2007)

A total 22 sequences (>500 bp) encoding different types of tyrosine kinases as well as receptor tyrosine kinases were identified in the leucaena transcriptome (Table 3.12). Among these, only two sequences were found in root and they were shorter than those in shoot. The longest sequence from shoot (query 1671721, 2403 bp) showed 90% identity and 95% similarity with PTI1-like tyrosine-protein kinase At3g15890-like protein from *Cicer arietinum* (Figure3.14)

Table 3.12: Tyrosine kinases sequences (>500 bp) in the leucaena transcriptome

Query name	Description	Length (bp)
1458725	PTI1-like tyrosine-protein kinase At3g15890-like	550
1543587	ALK tyrosine kinase receptor	523
1671721	PTI1-like tyrosine-protein kinase At3g15890-like	2403
1669511	Dual specificity tyrosine-phosphorylation-regulated kinase 4-like	1733
1663612	Dual specificity tyrosine-phosphorylation-regulated kinase 4-like	1238
1666255	PTI1-like tyrosine-protein kinase 3-like isoform X1	1222
1658710	LRR tyrosine-protein kinase At2g41820-like	1165

1655816	C-type lectin receptor-like tyrosine-protein kinase At1g52310-like	1057
1655815	C-type lectin receptor-like tyrosine-protein kinase At1g52310-like	1032
1666063	Hepatocyte growth factor-regulated tyrosine kinase substrate	951
1674263	LRR tyrosine-protein kinase At2g41820-like	951
1665275	PTI1-like tyrosine-protein kinase 3-like isoform X1	804
1654418	C-type lectin receptor-like tyrosine-protein kinase At1g52310-like	730
1675528	LRR tyrosine-protein kinase At2g41820-like	702
1651320	LRR tyrosine-protein kinase At2g41820-like	686
1631798	LRR tyrosine-protein kinase At2g41820-like	672
1644616	LRR tyrosine-protein kinase At2g41820-like	665
1647434	Dual specificity tyrosine-phosphorylation-regulated kinase 4-like	599
1652394	Hepatocyte growth factor-regulated tyrosine kinase substrate-like isoform X1	590
135371	Tyrosine-protein kinase CSK-like	572
1665181	Tyrosine-sulfated glycopeptide receptor 1-like precursor	557
1657837	Dual specificity tyrosine-phosphorylation-regulated kinase 2-like	542

Note: LRR: leucine-rich repeat receptor-like/ ALK: anaplastic lymphoma kinase

```
Query  277   MAFCPIFCFGKGSDRKARGKAQPTWRVFSLKELHSATNNFNYDNKLGEGGFGSVYWGQLW   456
             MAFCPIFC G GSDRK RGK QP WRVFSLKELHSATNNFNYDNKLGEGGFGSVYWGQLW
Sbjct  1     MAFCPIFCCGNGSDRKGRGKKQPPWRVFSLKELHSATNNFNYDNKLGEGGFGSVYWGQLW   60

Query  457   DGSQIAVKRLKVWSNKADMEFAAEVEILARVRHKNLLSLRGYCAEGQERLIVYDYMPNls   636
             DGSQIAVKRLKVWSNKADMEFA EVEILARVRHKNLLSLRGYCAEGQERLIVYDYMPNLS
Sbjct  61    DGSQIAVKRLKVWSNKADMEFAVEVEILARVRHKNLLSLRGYCAEGQERLIVYDYMPNLS   120

Query  637   llshlhGQHSAECLLNWSRRMSIAIGSAEGVAYLHQATPHIIHRDIKASNVLLDADFGA   816
             LLSHLHGQHSAE LL+W+RR++IAIGSAEG+ YLH+QATPHIIHRD+KASNVLLD++F A
Sbjct  121   LLSHLHGQHSAESLLDWNRRINIAIGSAEGIVYLHNQATPHIIHRDVKASNVLLDSEFQA   180

Query  817   QVADFGFAKLIPEGVTHVTT   876
             +VADFGFAKLIP+G THVTT
Sbjct  181   RVADFGFAKLIPDGATHVTT   200
```

Figure 3.14: BLASTX showing 90% identity of the leucaena PTI1-like tyrosine-protein kinase At3g15890-like protein with that from *Cicer arietinum*

Tyrosine phosphatases

Phosphorylation and dephosphorylation of a protein are two important stages, which act as "on-and-off" switch in the cell regulation (Luan 2003). Tyrosine phosphatase are involved

in removing the phosphate group from phosphorylated tyrosine residues. Several protein tyrosine phosphatases (PTPs) were characterized from *Arabidopsis* and other species (Xu et al. 1998, Gupta et al. 1998, Fordham-Skelton et al. 1999). The first protein tyrosine phosphatase (At PTP1) was identified in *Arabidopsis* (Xu et al. 1998). The first dual-specificity phosphatase (AtDsPTP1), which removes phosphates from both tyrosine and serine/threonine, was also identified from *Arabidopsis* (Gupta et al. 1998). *Arabidopsis thaliana* has a few protein Tyr-specific phosphatases (PTPs) and 22 dual-specificity phosphatases (DSPs) (Rayapureddi et al. 2005, Kerk et al. 2008). Tyrosine phosphatase activity is involved in the regulation of stomatal movement, a highly regulated process pivotal for plant survival (MacRobbie 2002).

In plants, a large group of receptor-like kinases (RLK) has been identified. RLKs are similar in structural organization to receptor tyrosine kinases (RTK) of animals. They are transmembrane proteins with an extracellular domain and a cytosolic tyrosine kinase domain. RLKs are members of the TKL (tyrosine kinase-like) proteins with a putative tyrosine kinase domain. Some TKLs in plants have been shown to have tyrosine kinase activity. The brassinosteroid receptor, BRI1 and its co-receptor, BAK1 have been shown to autophosphorylate tyrosine. The phosphorylation, in maize, was resistant to alkali treatment. Chemical, immunological and enzymatic data indicated the presence of tyrosine kinase activity and also phosphotyrosine residues in proteins of maize seedlings (Triozanek et al. 1995).

In leucaena, we found 21 sequences (>500 bp) encoding various types of protein tyrosine phosphatases (Table 3.13). Among these, two tyrosine phosphatase 3-like protein, one tyrosine phosphatase IBR5-like protein and one protein tyrosine phosphatase were found in the root transcriptome. The longest sequence (query 1668898, 1839 bp) from the shoot sequences showed 83% identity and 87% similarity with protein-tyrosine phosphatase mitochondrial 1-like protein from *Glycine max* (Figure 3.15)

Table 3.13: Tyrosine phosphatases sequences (> 500 bp) in the leucaena transcriptome

Query name	Descritpion	Length (bp)
1805304	Tyrosine-protein phosphatase 3-like	1640
1809372	Tyrosine-protein phosphatase 3-like	1200
1807762	Protein-tyrosine-phosphatase IBR5-like	682

1001214	Protein tyrosine phosphatase	553
1668898	Protein-tyrosine phosphatase mitochondrial 1-like	1839
1660614	Protein-tyrosine phosphatase mitochondrial	1819
1672784	Tyrosine phosphatase	1614
1665405	Protein-tyrosine-phosphatase IBR5-like	1555
1674473	Protein-tyrosine phosphatase mitochondrial 1-like	1155
1666542	Probable tyrosine-protein phosphatase At1g05000-like	1145
1666467	Tyrosine-protein phosphatase At1g05000-like	1021
1673627	Protein-tyrosine-phosphatase IBR5-like	967
1660445	Protein-tyrosine phosphatase mitochondrial 1-like protein-like	960
1648262	Tyrosine-protein phosphatase 3-like	881
1662954	Tyrosine-protein phosphatase At1g05000-like	867
1667712	Protein-tyrosine-phosphatase MKP1-like	779
1666209	Tyrosine-protein phosphatase At1g05000-like	729
1638318	Tyrosine-protein phosphatase 3-like	625
1646911	Protein-tyrosine phosphatase mitochondrial 1-like isoform 1	559
1633608	Tyrosine-protein phosphatase non-receptor type 20-like	529
108190	Tyrosine-protein phosphatase non-receptor type 21	525

```
Query  1839   VPRLKKLGVGGVITLNEPYETLVPSSLYHAHEIDHLVIPTRDYLFAPSLEDISLAVQFIH   1660
              VP LKKLGVGGVITLNEPYETLVPSSLYHAH IDHLVIPTRDYLFAPS  DI+ AVQFIH
Sbjct  79     VPHLKKLGVGGVITLNEPYETLVPSSLYHAHGIDHLVIPTRDYLFAPSFVDINRAVQFIH   138

Query  1659   QNACCGKTTYVHCKAGRGRSTTIVLCYLVEYKHMTPVTALDYVRSRRPRVLLAPSQWKAV   1480
              QNA CGKTTYVHCKAGRGRSTTIVLCY+VEYKHMTP  AL+YVRSRRPRVLLAPSQWKAV
Sbjct  139    QNATCGKTTYVHCKAGRGRSTTIVLCYMVEYKHMTPAAALEYVRSRRPRVLLAPSQWKAV   198

Query  1479   QNYSKCRPSPSNARptptptPYSPSRDAVLITQADLEGYHNACDAGMQFAIVTKVPKAKP   1300
              QNY+K RPSP        PYSPS DAVLIT+ADLEGYH+ CDAGM+ AIV K+PK KP
Sbjct  199    QNYNKRRPSP---------LPYSPSGDAVLITKADLEGYHSTCDAGMELAIVPKMPKTKP   249

Query  1299   MIARLSCLFASLKVSGSSVPAIRRLPISEARAC   1201
              MIARLSCLFASLKVSGSSVP  RRLP+SE+RAC
Sbjct  250    MIARLSCLFASLKVSGSSVPMTRRLPVSESRAC   282
```

Figure 3.15: BLASTX analysis showing 83% identity of protein-tyrosine phosphatase mitochondrial 1-like protein in leucaena with that of *Glycine max*.

MAP kinase

Mitogen-activated protein kinases (MAPKs) are signal transduction proteins and are found in all eukaryotes analyzed to date. They are typically involved in transducing extracellular signals. Different types of MAPKs were identified in various plant species that play an important role in plant signaling of a variety of abiotic and biotic stresses. The *Arabidopsis* genome contains about 110 genes coding for putative MAPK pathway components: 20 MAPKs, 10 MAPKKs (MAPK kinase) and more than 80 MAPKKKs (MAPK kinase kinase) (The *Arabidopsis* Genome Initiative 2000). Another *Arabidopsis* MAPK defined was Mpk 4, which is activated by multiple abiotic stresses, including cold, low humidity, hyperosmolarity, touch and wounding (Innes 2001). MPK4 functions as a regulator of pathogen defense responses, which is required for both repression of salicylic acid (SA)-dependent resistance and for activation of jasmonate (JA)-dependent defense gene expression (Andreasson et al. 2005). The primary role of MPK4 might be in regulating adaptation to environmental stress (Ichimura et al. 2000). In tobacco, the SIPK protein, which is orthologous to the *Arabidopsis* MAP kinase, MPK6, has many known activators, including avirulent pathogens (Zhang et al. 1998). Moreover, SIPK is directly responsible for induction of cell death during a pathogen-induced hypersensitive response (HR) (Inner 2001). Therefore, MAPKs serves as important factors in pathogen resistance as well as stress tolerance in plants.

From the BLASTX analysis we found 16 sequences encoding MAPK proteins in the shoot transcriptome and 2 MAPK sequences in the root transcriptome (Table 3.14). Among these, seven encoded MAPKKK 1 proteins. The longest sequence 1676167 (1918 bp) showed 84% identity and 93% similarity with big map kinase (bmk) from *Ricinus communis*. The big MAPK in leucaena is phylogenetically closest to those of *Ipomoea batatas* and *Pisum sativum* and farthest from that of *Nitotiana tabacum* (Figure 3.16)

Table 3.14: MAP kinases sequences (>500 bp) in the leucaena transcriptome

Query name	Description	Length (bp)
1805812	Big map kinase/bmk	955
1806898	MAP kinase kinase kinase mkh1-like	774

1676167	Big map kinase/bmk	1918
1672248	Map3k delta-1 protein kinase	1903
1669083	MAP kinase kinase kinase mkh1-like	1601
1669713	MAPepsilon protein kinase	1409
1674193	MAP kinase kinase kinase mkh1-like	1402
1675796	Map3k delta-1 protein kinase	1152
1670612	Big map kinase/bmk	1114
1668793	Big map kinase/bmk	1042
1651431	Big map kinase/bmk	1018
1659919	MAP kinase kinase kinase mkh1-like	923
1655949	MAP kinase kinase kinase mkh1-like	902
1652460	Map3k delta-1 protein kinase	897
1635045	MAP kinase kinase kinase mkh1-like	824
1637035	MAP kinase homolog	742
1634058	Big map kinase/bmk	626
1653633	MAP kinase kinase kinase mkh1-like	584
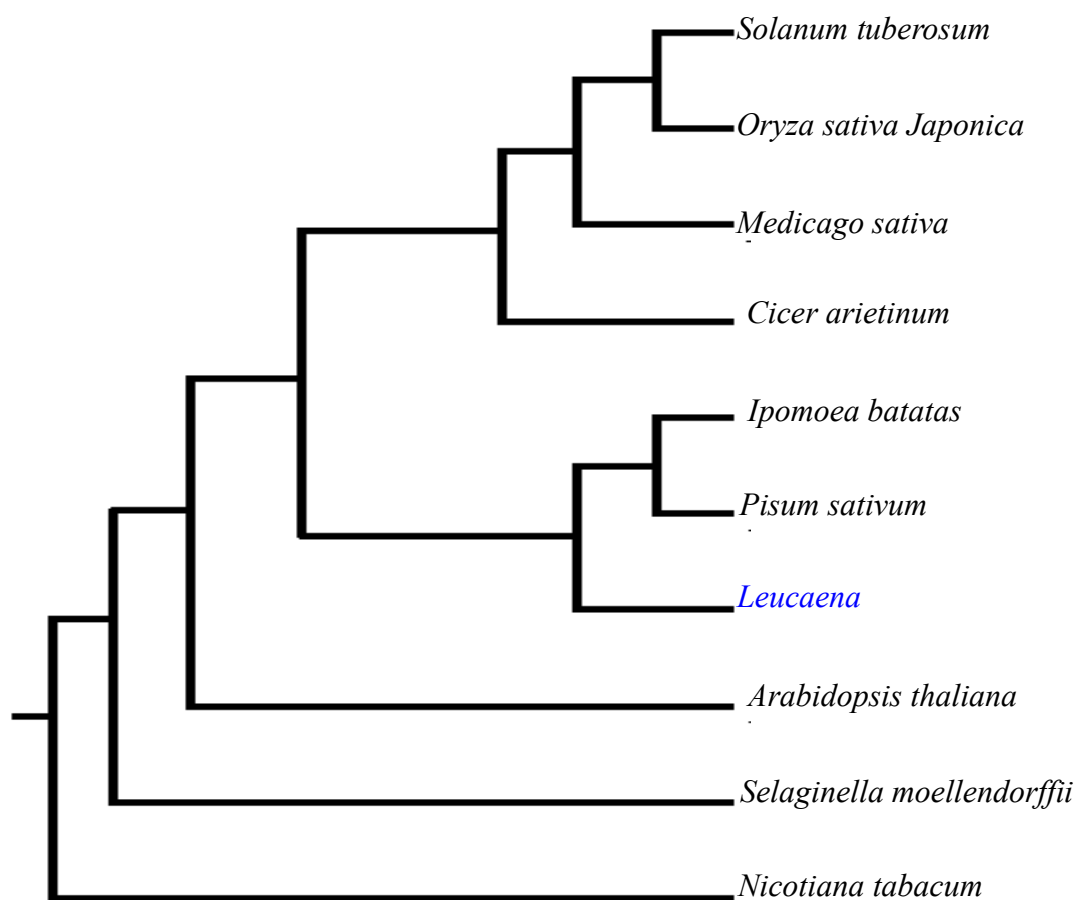
Figure 3.16: Phylogenetic tree showing the distance of big MAPK in leucaena with those of other plant species.

(j) Cysteine/mimosine synthases

Cysteine synthesis in plants represents the final step of assimilatory sulfate reduction and the almost exclusive entry reaction of reduced sulfur into metabolisms in plants. It is accomplished by the sequential reaction of two enzymes, serine acetyltransferase (SAT) and O-acetylserine (thiol) lyase (OAS-TL). Together they form the hetero-oligomeric cysteine synthase complex (CSC). The dual function of the CSC as a sensor and as part of a regulatory circuit controls cellular sulfur homeostasis. Regulation of cysteine synthesis in plant cells by the cysteine synthase complex (Wirtz and Hell 2006). Several cysteine synthases have been shown to function as a mimosine synthase, which catalyzes the synthesis of mimosine from 3-hydroxy-4-pyridone and acetyl serine (Ikegami et al. 1990, Ikegami and Murakoshi 1994; Saito 1997). In the current research, we are interested in cysteine synthesis because of its role in the

biosynthesis of mimosine, which is an important secondary metabolite produced by leucaena. Cysteine synthases were identified in different plant species. In *Arabidopsis,* 5 and 9 genes encoding SAT-like and OAS-TL-like proteins, respectively, accomplish cysteine synthesis.

We searched sequences for cysteine synthase in the leucaena transcriptome. Seven cysteine synthases were identified in shoot (Table 3.15). Surprisingly, when we searched for sequences longer than 500 bp, we did not find out any cysteine synthases in root transcriptome. However, we found 23 sequences that were longer than 100 bp in the root transcriptome. The longest sequence from the shoot transcriptome (query 1675987, 1911 bp) showed 52% identity and 53% similarity with cysteine synthase-like protein from *Fragaria vesca subsp. Vesca.* Phylogenetic analysis indicated that this cysteine synthase in leucaena is genetically divergent from those in other plant species (Figure 3.17).

Table 3.15: Cystein synthase sequences (> 500 bp) in the leucaena transcriptome

Query name	Description	Length (bp)
1675987	Cysteine synthase-like	1911
1670699	Cysteine synthase 2-like	1654
1666107	Cysteine synthase-like	1068
1650677	Cysteine synthase	1037
1637485	Cysteine synthase-like	763
1646843	Cysteine synthase, chloroplastic-like	667
1636325	Cysteine synthase-like	594
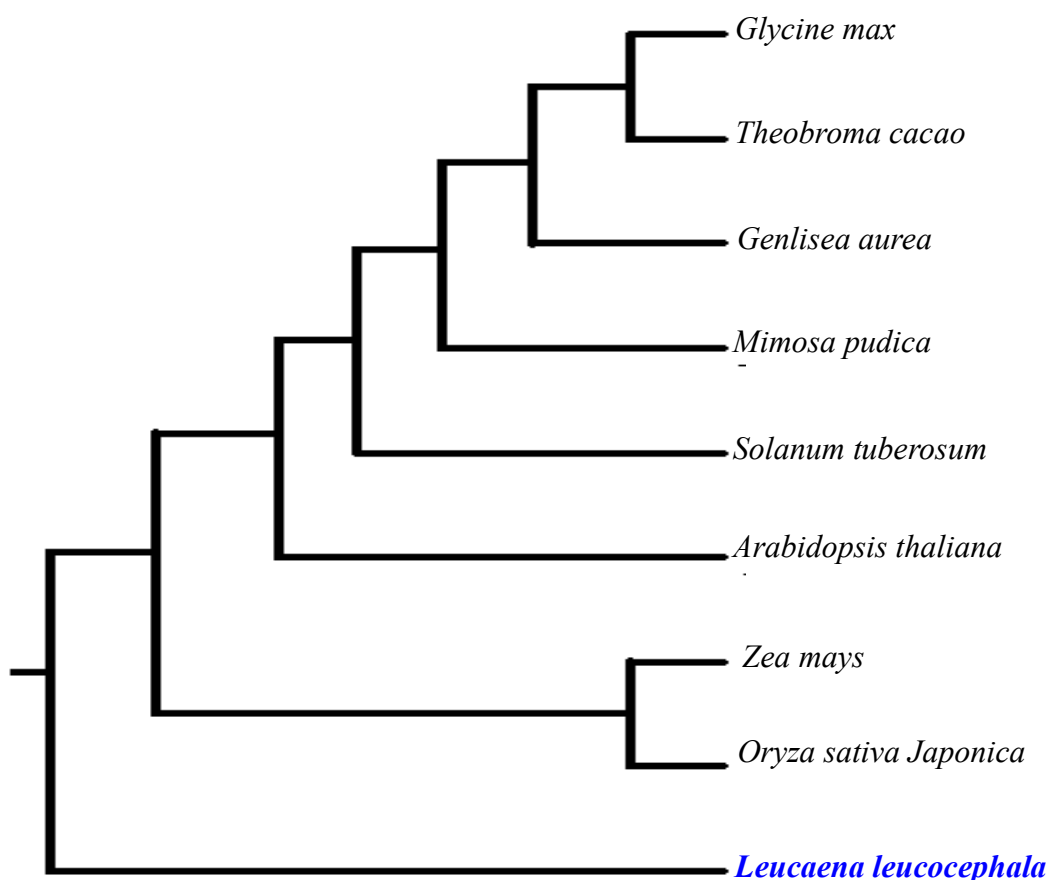
Figure 3.17: Phylogenetic tree of cysteine synthase in leucaena and different plant species

Mimosinase

Mimosinase is an leucaena enzyme that degrades mimosine into 3-hydroxy-4-pyridone, pyruvate and ammonia (Vishal Negi 2012). Mimosine degradation by mimosinase may be a biochemical mechanism of leucaena to withstand certain physiological stresses such as drought (Negi and Borthakur, unpublished). Seven sequences (>500 bp) encoding mimosinase were identified in leucaena (Table 3.16). Fukuta et al. (2007) isolated one sequence with 443 amino acid encoding mimosinase. The analysis of BLASTX leucaena transcriptome, we found that the longest query 1663223 (1843 bp) in shoot showed 90% identity and 94% similarity with mimosinase from *Leucaena leucocephla* that identified by Fukuta et al. (2007) (Figure 3.18 a). While, the only query 1806653 (>500 bp) in root showed 59% identity and 86% similarity with mimosinase from *Mimosa pudica* (Figure 3.18 b). The presence of several mimosinase-like

sequence in the shoot transcriptome suggests that leucaena may have more than one isoforms of mimosinase.

Table 3.16: Mimosinase sequences (>500 bp) in the leucaena transcriptome

Query name	Description	Length (bp)
1806653	Mimosinase	1055
1663223	Mimosinase	1843
1667384	Mimosinase	1335
1662801	Mimosinase	1034
1666921	Mimosinase	1019
1655633	Mimosinase	572

```
Query  363    TFLNPLVSSVTVNPHPKITNGKGFRVNCLIRTQQTVVKTEAKENAAVLTQEKRVEKEPSV    542
              TFLNPLVSSV VNP PKIT+GKGFRVNCLIRTQQTV+KT+ KENAAVLT  KRVEKEPSV
Sbjct  6      TFLNPLVSSVAVNPQPKITSGKGFRVNCLIRTQQTVIKTDTKENAAVLTPGKRVEKEPSV    65

Query  543    STVLANYHADWDPFEATSTPIYQSATFRMKNATEYNDYYYSRVANPTTSTLEKILAEIEH    722
              STVLANYHADWDPFEATSTPIYQSATFRMKNATEYN+YYYSRVANPTTSTLEKI+AEIE+
Sbjct  66     STVLANYHADWDPFEATSTPIYQSATFRMKNATEYNEYYYSRVANPTTSTLEKIIAEIEN    125

Query  723    AEYVTCFTSGMSALTAVCELVSPGDEILTVEDIYGGSYSFIENLMVRKAGITVKRVDTSQ    902
              AEYVTCFTSGMSALTAVCELV+PGDEILTVEDIYGGSY FIENLMVRKAGITVKRVDTS
Sbjct  126    AEYVTCFTSGMSALTAVCELVNPGDEILTVEDIYGGSYGFIENLMVRKAGITVKRVDTSN    185

Query  903    IEKVKAAMTSKTKLVWLESPSNPQLKISDIREIARIAHAYGAIVFIDNCIMSPLLSNPLD    1082
              IE VKAAMT+KTKLVWLESPSNPQLKISDIREIARIAHAYGAIVFIDNCIMSPLLS+PL+
Sbjct  186    IENVKAAMTNKTKLVWLESPSNPQLKISDIREIARIAHAYGAIVFIDNCIMSPLLSHPLE    245

Query  1083   LGADIVMHSATKFIAGNSSCMAGCLATNNKELADLLLSYKNATGCGLSPQDAWICLEGIK    1262
              LGADIVMHSATKFIAGNSSCMAG LATNNKELAD LLSY++ATGCGLSPQDAWICLEGIK
Sbjct  246    LGADIVMHSATKFIAGNSSCMAGSLATNNKELADKLLSYRSATGCGLSPQDAWICLEGIK    305

Query  1263   TLPLRIEEKQKNAQTVANYLANHSKVTKVNYPGLSDNPGYELHKSQSKGPGSVISFETGS    1442
              TLPLR+EEKQKNA TVANYL N+ K+TKVNYPGL DNPGY+LHKSQSKGPGSV+S ETGS
Sbjct  306    TLPLRVEEKQKNALTVANYLDNNPKITKVNYPGLPDNPGYDLHKSQSKGPGSVMSCETGS    365

Query  1443   LPLSKQIVEDTKYFSKIVGFGGVGSAICLPWYTSHKAIPEAEKISMGIAKDLVRMSVGIE    1622
              LPLSKQIVEDTK+FSKIVGFGGVGSAICLPWYTSHKAIPE EKI MGI KDL+R+SVGIE
Sbjct  366    LPLSKQIVEDTKFFSKIVGFGGVGSAICLPWYTSHKAIPEPEKIRMGITKDLIRISVGIE    425

Query  1623   DVEDLIHDLKNVMSTSPF    1676
              DV+DLI DL N MST  F
Sbjct  426    DVQDLIQDLDNAMSTPTF    443
```

Figure 3.18 (a): BLASTX showing 90% identity of mimosinase in leucaena with that in *Leucaena leucocephala* identified by Fukuta et al. 2007

```
Query  192   AKVENAKYGYCFTSGMSALTAISELVKPG   278
             A+++NAK+ YCF+SGM+AL ++ ELV PG
Sbjct  119   AQLDNAKFAYCFSSGMTALNSVCELVSPG   147
```

Figure 3.18 (b): BLASTX showing 59% identity of mimosinase in leucaena with that in *Mimosa pudica*
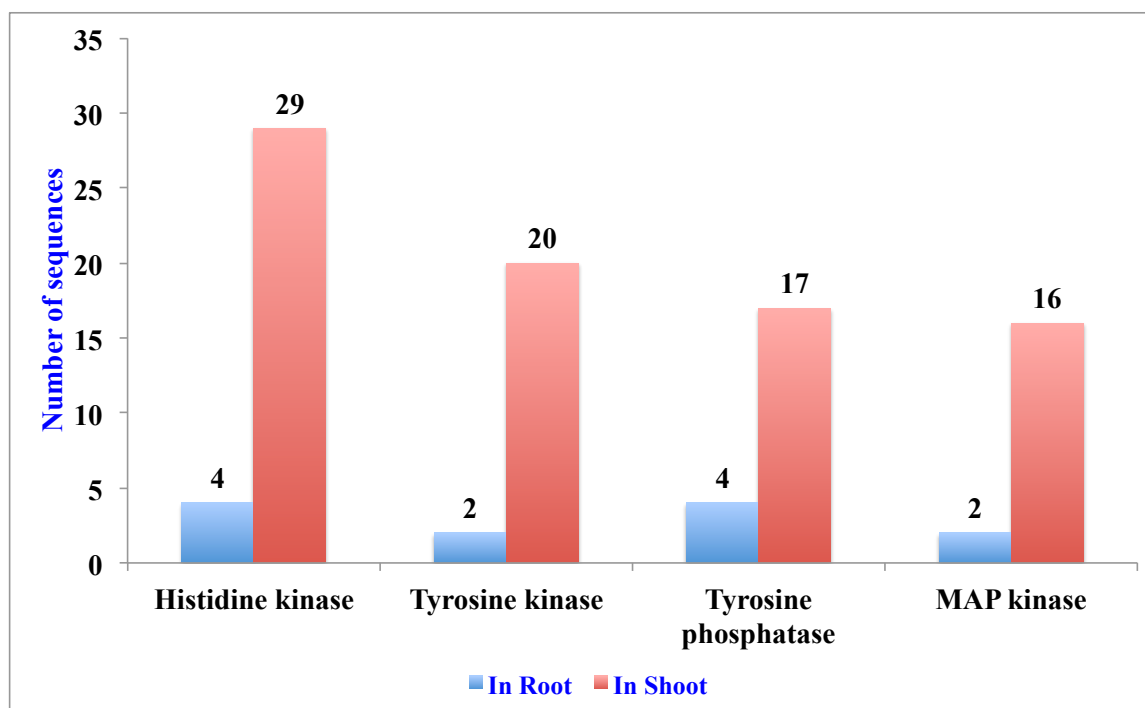

Figure 3.19: Kinases and phosphatases in the leucaena transcriptome

Overall, various member of kinases and phosphatases (>500 bp) were found in the leucaena transcriptome. However, Ser/ Thr kinases were the biggest group kinase with 62 and 574 sequences in root and shoot, respectively. The number of sequences encodes histidine kinases, tyrosine kinases, tyrosine phosphatases, MAP kinases and cysteine synthases in shoot were always higher than those in root. Surprisingly, we did not find any sequences (>500 bp) encoding cysteine synthases in root transcriptome, whereas, 7 sequences were found in shoot.

3.4. Discussion

Transcription analysis of leucaena was considered important for a number of reasons: (i) genome sequencing has not been done for leucaena; (ii) Leucaena represents an important forage legume as well as a tropical tree for agroforestry; (ii) Leucaena is a tetraploid species with a large number of chromosome (2n=104); so it should contain genes encoding several

isoforms of important proteins; and (iv) leucaena is highly resistant to various biotic and abiotic stresses; we are interested in isolating those genes. From the transcriptome analysis, we expected to identify many genes, which were transcribed in shoots and roots of leucaena. From our data, although we observed nearly equal number of contigs for both shoots and roots, the length of transcriptome sequences for roots was shorter than those for shoots. On the other hand, many sequences in the root transcriptome were quite long; for examples, query 1805591 (5,178 bp), query 1809735 (4,793 bp) and query 1809424 (3,948 bp), which encoded hypothetical proteins. Homologs of these hypothetical proteins have been found in other plants also. This suggests that the short sequence length of the root transcriptome was not due to poor quality of RNA isolated from roots.

By comparing sequences in the shoot and root transcriptomes, we identified 33 sequences that were absent in the shoot transcriptome. Notable among these root sequences were 'bark storage protein A' (query 24395, 979 bp), lypoxygenase (query 624, 523 bp) and vicilin-like antimicrobial peptides (query 981413, 982 bp). Previous studies in poplar found that the bark storage protein was associated with seasonal nitrogen storage within vesicles of inner bark parenchyma and xylem ray cells. The degradation of bark storage protein and nitrogen remobilization were observed to be related to sink demands from active bud growth (Coleman et al. 1994). It is likely that leucaena bark storage may also function in the same way in balancing between nitrogen storage and current demand for nitrogen for growth. Lipoxygenase, which catalyzes the dioxygenation of polyunsaturated fatty acids are found widely in plants. Lipoxygenase initiates the synthesis of a group of acyclic or cyclic compounds collectively called oxylipins, which are products of fatty acid oxidation, with diverse functions in the cell (Melan et al. 1993). Lipoxygenase catalyzes peroxidation of membrane lipids, which may lead to structural changes in the membrane (Brass 1999). Lipoxygenase may initiate the synthesis of signal molecules in plants (Brash 1999). One product of lipoxygenase reaction may be jasmonic acid, which is a signal molecule that regulates a wide range of processes in plants for growth and development. Jasmonic acid has also a role in establishing symbiosis of legumes with nitrogen-fixing rhizobia and phosphate-solulbilizing arbuscular mycorrhizal fungi (Hause and SchaarSmith 2009). Moreover, lipoxygenase gene expression is induced by different stresses such as wounding, drought and pathogen attacks (Porta et al. 1999, Melan et al. 1993). It is likely that lipoxygenase expressed

in leucaena roots may be involved in signal productions. Vicilin-like antimicrobial peptide is a globulin found in seeds of legumes. Vicilin-like antimicrobial peptide was previously identified in *Macadamia integrifolia* kenels (Marcus et al. 1999). The peptide, containing 45 amino acid residues, is obtained from a precursor polypeptide with 666 residues. The peptide has antimicrobial properties that inhibited various plant fungi in vitro. It is interesting that leucaena root transcriptome contains sequence encoding a vicilin-like antimicrobial peptide, which may increase leucaena's potential for biotic stresses resistance.

Chitinases are known to provide a general and non-specific resistance against fungal pathogens. Considering that leucaena has no known fungal pathogens, it is not surprising that many chitinases express even in uninfected plants grown in the growth room. Different plants are known to express four different classes of chitinases. Among these, class IV chitinases are very similar to class I chitinases. In the leucaena transcriptome, we found chitinases of class I, II, and III, but not class IV. One of the leucaena chitinases identified in this study had identical sequences with a class I chitinase, previously isolated and characterized by Kaomek et al. (2008).

Considering that leucaena is highly resistant to bacterial and fungal pathogens, we were interested in identification of resistance genes from leucaena. Unlike chitinases, proteins encoded by resistance genes in plants provide specific resistance through recognition of pathogenic effectors. A general characteristic of resistance gene is the presence of nucleotide-binding (NB) and leucine-rich repeat (LRR) domains. We have identified 231 sequences (>100 bp) in leucaena transcriptome that showed homology with NB-LRR proteins. Compared to *Arabidopsis*, which has 149 genes encoding for NB-LRR protein, leucaena has higher number of NB-LRR genes. The presence of a high number and many types of NB-LRR in the transcriptome leucaena is consistent with the general observation that leucaena is naturally resistant to diseases. Many NBS-LRR-encoding genes are constitutively expressed at low levels in healthy, unchallenged tissue, although some show tissue-specific expression (McHale et al. 2006). However, the transcription of some NB-LRR genes is induced in the presence of a pathogen (Yoshimura et al. 1998, Navarro et al. 2004). Our plants were not exposed to any pathogens; therefore our transcriptome may not represent all NB-LRR genes in leucaena. The presence of a high number and many types of NB-LRR in the transcriptome leucaena is consistent with the general observation that leucaena is naturally resistant to diseases. Isolation

of full-length cDNA and further characterization of these genes will provide a valuable resource for resistance genes for plant improvement programs in the future.

The next category of resistance genes identified in the leucaena transcriptome is WRKY. WRKY proteins recognize pathogen-associated molecular patterns (PAMPs), which induce innate immune responses to protect the host plant from the pathogen. We found 368 WRKY sequences in the leucaena transcriptome. Therefore, leucaena may contain a large number of genes for WRKY, and this may be one reason why leucaena is naturally resistant to pathogens.

Among the fast-growing tropical tree legumes, there are no other plants as drought resistant as leucaena. Under drought conditions, leucaena can remain alive for a long time. We are interested in isolating genes related to drought tolerance from leucaena. Previously, several ERD genes for drought tolerance had been identified from *Arabidopsis* and soybean (Alves and Fietto 2013). ERD genes are induced by IBA, salisalic acid, injuries and pathogens. In our experiment, we grew leucaena under normal growth conditions without exerting any stress conditions; we do not know why these genes are expressed at normal conditions. It is likely that some ERD genes are constitutively expressed. Therefore, isolation of full-length cDNA and further characterization of these genes may provide clues for genetic basis of drought tolerance in leucaena.

Recently, a number of genes for UV resistance, such as UVR8, UVH6 and UVH3, related UV resistance in plants have been isolated. UVR-8 is a UV-B photoreceptor that induces expression of other genes related to UV protection (Rizzini et al. 2011). UV-B is very energetic, and high UV-B irradiance induces the formation of reactive oxygen species, which can damage DNA, proteins, photosynthetic organelles and adversely affect growth and development (Jenkins 2009). UVR8 regulates the accumulation of phenolics and other metabolites under natural sunlight. The accumulation of phenolic compounds provides protection in plants to UV radiation (Morales et al. 2010). UVR8 is a key regulator of gene expression in *Arabidopsis* leaves under solar UV (Morales 2013). Among 17 UV-related genes (>500 bp) identified in the leucaena transcriptome, one sequence encoded UVR-8 like protein. Another UV-related gene, UVH-6 in *Arabidopsis*, has been shown to be a transcription regulator (Hall et al. 2009). UVH-6 mutant of *Arabidopsis* was sensitive to cold; so UVH-6 is also involved in cold-tolerance. UVH-6-encoded protein facilitates repair of DNA damaged

due to UV radiation (De Boer and Hoeijmakers 2000). High UV radiation damages plant DNA, which may be repaired by activities of UVH-6 in association with other proteins. In the leucaena transcriptome, we also found a sequence encoding UVH-3. UVH-3 encodes a nucleotide excision repair protein. UVH-3 mutant of *Arabidopsis* was sensitive to H_2O_2 and showed premature senescence due to the failure to repair oxygen damage in plant tissues (Liu et al. 2001). We found 7 sequences encoding UVH6-like proteins in the transcriptome of leucaena plants, which were grown under fluorescent lights in the growth room at 22–25 ˚C. These plants were not exposed to UV light; therefore, it appears that some UV-related genes are constitutively expressed in leucaena.

In leucaena, we found 636 sequences (>500 bp) encoding different kinds of Ser/Thr kinases. Among these, 183 sequences encoded LRR-receptor Ser/Thr kinases. One example of LRR receptor-like Ser/Thr kinase is BAK-1 in *Arabidopsis*. Proteins encoded by BAK-1 suppressed another gene, BRI-1, which is involved in brassinosteroids (BRs) signaling. BRI-1 mutant of *Arabidopsis* is sensitive to application of exogenous BR. BAK-1 mutant, on other hands, showed reduced sensitivity to BR (Li et al. 2002). LRR receptor-like Ser/Thr kinase isolated from leucaena may be involved in BR signaling. LRR receptor-like Ser/Thr kinases contain an external receptor domain and a cytoplasmic Ser/Thr kinase domain. With a LRR external ligand-binding domain and a cytoplasmic Ser/Thr kinase domain, these proteins are involved in sensing environmental chemical signals and transducing the signal to plant cytoplasm. In addition to the LRR-Ser/Thr kinases, there were 338 additional Ser/Thr protein kinase sequences in the leucaena transcriptome. From the BLASTX analysis, the role of these Ser/Thr kinases in the cell could not be determined.

We were also interested in identifying histidine kinases. They are the most unique kinases in plants. Generally, histidine kinases are a part of a two-component system in which one domain of the protein functions as the receptor and a cytoplasmic domain serves as the response regulator. We found a total of 33 sequences encoding histidine kinases in the leucaena transcriptome. Isolation of full-length cDNA and further characterization of genes will be necessary for complete elucidation of functions of histidine kinases in leucaena.

We also identified 22 sequences (>500 bp) encoding different types of tyrosine kinases. Some tyrosine kianses serve as receptor kinases. Few others may function as dual-specific

kinases that phosphorylate both Tyr and Ser/Thr. The functional roles of the tyrosine kinases in leucaena could not be determined from the BLASTX analysis. It is likely that these kinases are involved in signal transduction and gene regulation. We also found 21 sequences encoding tyrosine phosphatases. These phosphatases may function in concert with tyrosine kinases in gene regulation. Another class of regulatory gene sequences identified in this study is MAPKs, which are signal transduction proteins. They function down-stream from LRR tyrosine receptor kinases or NB-LRR resistance gene proteins. The MAPKs are a part of the signal transduction pathways related to various biotic and abiotic stresses in plants. Further characterization of the genes with full-length cDNA will be necessary to fully understand the role of these genes in signal transduction.

One primary goal of research in our lab is to understand mimosine metabolism pathways in leucaena. Leucaena naturally produces at least 3% mimosine in the foliage. Understanding of mimosine metabolism in leucaena is necessary: (i) to understand the role of mimosine in leucaena physiology, (ii) to develop mimosine-free leucaena by disrupting mimosine biosynthesis genes, (iii) to identify genes encoding enzymes for mimosine degradation from leucaena, so that such genes can be used for developing transgenic leucaena with reduced mimiosine content. Cysteine synthases are known to be involved in the last step of mimosine biosynthesis from 3-hyddroxy-4-pyridone. Plant contains several cysteine synthases. We expected that one or more cysteine synthases in leucaena could be specific for biosynthesis of mimosine. We have isolated 23 sequences encoding cysteine synthase-like proteins. Recently, Vishal Negi from our lab, isolated and characterized one full-length cDNA encoding mimosinase. From the transcriptome analysis, we found that there are additional isoforms of mimosinase in leucaena.

References

Afzal A, Wood A, and Lightfoot D (2008). Plant receptor-like Serine/Threonine kinases: roles in signaling and plant defense. MPMI, 21(5):507–517

Alves M and Fietto L (2013). Functional diversity of early responsive to dehydration (ERD) genes in Soybean. "A Comprehensive survey of International soybean research - genetics, physiology, agronomy and nitrogen relationships" (Book, chapter 22). Edt James E. Board. ISBN 978-953-51-0876-4. DOI: 10.5772/50993

Ancillo G, Witte B, Schmelzer E, et al. (1999). A distinct member of the basic (class I) chitinase gene family in potato is specifically expressed in epidermal cells. Plant Molecular Biology, 39:1137-1151

Andreasson E, Jenkins T, Brodersen P, Mundy J, et al. (2005). The MAP kinase substrate MKS1 is a regulator of plant defense responses. The EMBO Jol., 24: 2579–2589

Bleecker B and Kende H (2000). Ethylene: A gaseous signal molecule in plants. Annu. Rev. Cell Dev. Biol.,16:1-18

Brash R (1999). Lipoxygenases: occurrence, functions, catalysis, and acquisition of substrate. J Biol Chem., 274(34):23679-82.

Bretz R et al. (2003). A translocated protein tyrosine phosphatase of *Pseudomonas syringae pv.* tomato DC3000 modulates plant defence response to infection. Mol. Microbiol. 49:389–400

Coleman C, Baíiados M, and Chen T (1994). Poplar Bark storage protein and a related wound-induced gene are differentially induced by nitrogen. Plant Physiol., 106:211-215

Collinge D, Kragh K, Mikkelsen I, Nieler K, Rasmussen U, and Vad K (1993). Plant chitinases. Plant J., 3:31-40

Conrads-Strauch J, Dow J, Milligan D, Parra R, and Daniels M (1990). Induction of hydrolytic enzymes in *Brassica campestris* in response to pathovars of *Xanthomonas campestris.* Plant Physiol, 93:238-243

Dangl J and Jones J (2001). Plant pathogens and integrated defence responses to infection. Nature, 411:826 – 833

Daugrois J, Lafitte C, Barthe J, et al. (1990). Induction of β-1,3-glucanase and chitinase activity in compatible and incompatible interactions between *Colletotrichum lindemuthianum* and bean cultivars. J Phytopathol, 130:225-234

De Boer J and Hoeijmakers H (2000). Nucleotide excision repair and human syndromes. Carcinogenesis, 21:453–460

Eitas T, and Dang J (2010). NB-LRR proteins: pairs, pieces, perception, partners, and pathways. Current Opinion in Plant Biology, 13:472–477

Espinosa A, et al. (2003). The *Pseudomonas syringae* type III-secreted protein HopPtoD2 possesses protein tyrosine phosphatase activity and suppresses programmed cell death in plants. Mol. Microbiol., 49:377–387

Favory J, Stec A, Gruber H, Rizzini L, Oakeley E, et al. (2009). Interaction of COP1 and UVR8 regulates UVB-induced photomorphogenesis and stress acclimation in *Arabidopsis.* EMBO J., 8:591–601

Fordham-Skelton P, Skipsey M, Evans M, Edwards R, and Gatehouse A. (1999). Higher plant tyrosine-speci®c protein phosphatases (PTPs) contain novel amino-terminal domains: expression during embryogenesis. Plant Mol. Biol., 39:593 - 605

Fronzes R, Schäfer E, Wang L, Waksman G, et al. (2009). Structure of a Type IV secretion system core complex. Science, 323:266 – 268.

Fukuta M, Moriwaki T, Kedashiro S, and Tawata S (2007). Plant mimosinase is homologous to cystathionine beta-lyase. Unpublished. Submitted (20-MAR-2007)

Gerhardt L, Sachetto-Martins G, Contarini M, Margis-Pinheiroa M, et al. (1997). *Arabidopsis thaliana* class IV chitinase is early induced during the interaction with *Xanthomonas campestris*. FEBS Letters, 419: 69-75

Ghelis T, et al. (2008). Protein tyrosine kinases and protein tyrosine phosphatases are involved in abscisic acid-dependent processes in *Arabidopsis* seeds and suspension cells. Plant Physiol. 148:1668–1680

Goring R, and Walker C (2004). Self-rejection, a new kinase connection. Science, 303:1474-1475

Gupta R, Huang Y, Kieber J, and Luan S (1998). Identification of a dual-specificity protein phosphatase that inactivates a MAP kinase from *Arabidopsis*. Plant J., 16:581–89

Hall J, Cobb J, Iqbal M, Mount D, et al. (2009). UVH-6, a plant homolog of the human/ yeast TFIIH transcription factor subunit XPD/RAD3, regulates cold-stress genesin Arabidopsis thaliana. Plant Mol Biol Rep, 27:217-228

Hamel F, Boivin R, Tremblay C, and Bellemare G (1997). Structural and evolutionary relationships among chitinases of flowering plants. J Mol Evol, 44:614–624

Hause B, and Schaarschmidt S (2009). The role of jasmonates in mutualistic symbioses between plants and soil-born microorganisms. Phytochemistry, 70:1589–1599.

Hauser A (2009). The type III secretion system of *Pseudomonas aeruginosa*: infection by injection. Nature, 7:654 - 665

He P, Shan L, Lin N, Martin G, Sheen J, et al. (2006). Specific bacterial suppressors of MAMP signaling upstream of MAPKKK in *Arabidopsis* innate immunity. Cell, 125:563–575.

Higuchi M, Pischke S, Mahonen P, Kakimoto T, et al. (2004). In planta functions of the *Arabidopsis* cytokinin receptor family. Proc. Natl. Acad. Sci. U.S.A. 101:8821-8826

Hoagland D, and Arnon D (1950). The water-culture method for growing plants without soil. California Agricultural Experimental Station Circular 347, the College of Agriculture. University of California-Berkeley, CA.

Huang J, Lin M, Huang D, Takahashi T, and Sugiyama M (2003). Protein tyrosine phosphorylation during phytohormone-stimulated cell proliferation in *Arabidopsis* hypocotyls. Plant Cell Physiol, 44:770-775.

Hung C, Huang D, Chien S, Yeh M, Chen Y, Chi C, et al. (2007). Protein tyrosine dephosphorylation during copper-induced cell death in rice roots. Chemosphere, 69:55-62.

Ichimura K, et al. (2000). Various abiotic stresses rapidly activate *Arabidopsis* MAP kinases ATMPK4 and ATMPK6. Plant J. 24:655–665

Ikegami F, and Murakoshi I (1994). Enzymatic synthesis of non-protein beta-substituted alanines and some higher homologues in plants. Phytochemistry, 35:1089-1104

Ikegami F, Mizuno M, Kihara M, and Murakoshi I. (1990). Enzymatic synthesis of the thyrotoxic amino acid mimosine by cysteine synthase. Phytochemistry, 29:3461-3465

Innes R (2001). Mapping out the roles of MAP kinases in plant defense. TRENDS in Plant Science, 6(9), 392 – 394

Islas-Flores I, Oropeza C, and Hernandez-Sotomayor S (1998). Protein phosphorylation during coconut zygotic embryo development. Plant Physiol, 118:257-63.

Jenkins I (2009). Signal transduction in responses to UV-B radiation. Annu Rev Plant Biol, 60: 407–431

Jones J and Dangl J (2006). The plant immune system. Nature, 444: 323 – 329

Joosten M and De Wit P (1989). Identification of several pathogenesis-related proteins in tomato leaves inoculated with *Cladosporium fulvum (*syn. *Fulvia fulva*) as 1,3-β-glucanases and chitinases. Plant Physiol, 89:945-951

Kameyama K, Kishi Y, Yoshimura M, Kanzawa N, Sameshima M, and Tsuchiya T (2000). Tyrosine phosphorylation in plant bending. Nature, 407:37

Kaomek M, Mizuno K, Fujimura T, Cairns K, et al. (2003). Cloning, expression, and characterization of an antifungal chitinase from *Leucaena leucocephala* de Wit. Biosci. Biotechnol. Biochem., 67(4):667-676

Karasuda S, Tanaka S, Kajihara H, Yamamoto Y, and Koga D (2003). Plant chitinase as biocontrol agent instead of chemical fungicides. Biosci. Biotechnol. Biochem., 67(1):221 – 224

Kariola T, Brader G, Helenius E, Li J, Heino P, and Palva E (2006). Early responsive to dehydration 15, a negative regulator of Abscisic acid responses in *Arabidopsis*. Plant Physiology, 42:1559–1573

Kasprzewska A (2003). Plant chitinases – regulation and function. Cellular & molecular biology letters, 8:809 – 824

Kerk D, et al. (2008). Evolutionary radiation pattern of novel protein phosphatases revealed by analysis of protein data from the completely sequenced genomes of humans, green algae, and higher plants. Plant Physiol. 146:351–367

Kiyosue T, Yamaguchi-Shinozaki K, and Shinozaki K (1994). ERD15, a cDNA for a dehydration-induced gene from *Arabidopsis thaliana*. Plant Physiol, 106(4): 1707

Li L Wen J, Lease K, Doke J, Tax F, and Walker J (2002). BAK1, an Arabidopsis LRR receptor-like protein kinase, interacts with BRI1 and modulates Brassinosteroid signaling. Cell, 110:213–222

Liu Z, Hall J, and Mount D (2001). Arabidopsis UVH3 gene is a homolog of the Saccharomyces cerevisiae RAD2 and human XPG DNA repair genes. Plant J., 26(3):329-38

Luan S (2003). Protein phosphatase in plants. Annu. Rev. Plant Biol., 54:63–92

MacRobbie E (2002). Evidence for a role for protein tyrosine phosphatase in the control of ion release from the guard cell vacuole in stomatal closure. Proc Natl Acad Sci USA, 99:11963-8.

Marcus J, Green J, Goulter K, and Manners J (1999). A family of antimicrobial peptides is produced by processing of a 7S globulin protein in Macadamia integrifolia kernels. The Plant Journal, 19(6):699 - 710

Martin G, Frary A, Wu T, Brommonschenkel S, Chunwongse J, Earle E, and Tanksley S (1994). A member of the tomato Pto gene family confers sensitivity to fenthion resulting in rapid cell death. Plant Cell, 6:1543-1552

McHale L, Tan X, Koehl P, and Michelmore R (2006). Plant NBS-LRR proteins: adaptable guards. *Genome Biology*, **7**:212

Melan A, Dong X, Endara E, Davis R, Ausubel M, and Peterman K (1993). An Arabidopsis thaliana lipoxygenase gene can be induced by pathogens, abscisic acid, and methyl jasmonate. Plant Physiol, 101:441–450

Morales L, Brosché M, Vainonen J, Aphalo P, et al. (2013). Multiple roles for UV resistance locus 8 in regulating gene expression and metabolite accumulation in Arabidopsis under solar ultraviolet radiation. Plant Physiology,161:744–759

Morales L, Tegelberg R, Brosché M, Keinänen M, Lindfors A, and Aphalo P (2010). Effects of solar UV-A and UV-B radiation on gene expression and phenolic accumulation in Betula pendula leaves. Tree Physiol, 30:923–934

Munné-Bosch S, and Alegre L (2004). Die and let live: leaf senescence contributes to plant survival under drought stress. Funct Plant Biol., 31:203–16

Navarro L, Zipfel C, Rowland O, Keller I, Robatzek S, Boller T, and Jones J (2004). The transcriptional innate immune response to flg22. Interplay and overlap with Avr gene-dependent defense response and bacterial pathogenesis**.** *Plant Physiol.,* 135**:**1113-1128

Nawkar G, Maibam P, Park J, Sahi V, Lee S, and Kang C (2013). UV-Induced Cell Death in Plants. Int. J. Mol. Sci. ,14:1608-1628

Negi V. 2012 (Ph.D thesis). Biochemistry of mimosine catabolism by enzymes from *Rhizobium sp.* strain TAL1145 and *leucaena leucocephala*. University of Hawaii at Manoa.

Pandey S and Somssich I (2009). The Role of WRKY Transcription factors in plant immunity. Plant Physiology, 150:1648–1655

Pham J, Liu J, Bennett M, Mansfield J, and Desikan R (2012). *Arabidopsis* histidine kinase 5 regulates salt sensitivity and resistance against bacterial and fungal infection. New Phytologist, 194:168–180

Porta H, Rueda-Benítez P, Campos F, Colmenero-Flores M, Colorado M, Carmona J, Covarrubias A, and Rocha-Sosa M (1999). Analysis of lipoxygenase mRNA accumulation in the common bean (Phaseolus vulgaris L.) during development and under stress conditions. Plant Cell Physiol, 40:850–858

Punija Z, and Zhang Y (1993). Plant chitinase and their roles in resistance fungal diseases. Journal of Nematology, 25(4):526 – 540

Rayapureddi J, et al. (2005). Characterization of a plant, tyrosine- specific phosphatase of the aspartyl class. Biochemistry, 44:751–758

Rensink W, Hart A, Buell C, et al (2005). Analyzing the potato abiotic stress transcriptome using expressed sequence tags. Genome, 48:598-605

Rizzini L, Favory J, Cloix C, Faggionato D, Jenkins I, et al. (2011). Perception of UV-B by the Arabidopsis UVR8 protein. Science, 332:103–106

Rodriguez M, Edsga D, Mundy J, et al (2010). Transcriptomes of the desiccation-tolerant resurrection plant *Craterostigma plantagineum*. The Plant Journal, 63:212–228

Rushanaedy I, Jones T, Dudley N, Borthakur D, et al. (2012). Chitinase is a potential molecular biomarker fro detecting resistance to *Fusarium oxysporum* in *Acacia koa*. Tropical Plant Biol. 5:244-252

Saito K, Kimura N, Ikegami F, and Noji M (1997). Production of plant non-protein amino acids by recombinant enzymes of sequential biosynthetic reactions in bacteria. Biol Pharm Bull, 20:47-53

Shakhbazau A, and Kartel' N (2008). Chitinases in bioengineering research. Russian Journal of Genetics, 44(8):881–889

Shinozaki K, and Yamaguchi-Shinozaki K (2007). Gene networks involved in drought stress response and tolerance. Jol of Experimental Botany, 58(2):221–227

Song W, Wang G, Chen L, Kim H, Ronald P, et al (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa*21. Science, 270(5243):1804-1806

Sugiyama N, Nakagami H, Mochida K, Daudi A, Tomita M, Shirasu K, et al. (2008). Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in *Arabidopsis*. Mol Syst Biol, 4:1-7

The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature ,108:796–815

Thilmony R, Chen Z, Bressan R and Martin G (1995). Expression of the tomato Pto gene in tobacco enhances resistance to *Pseudomonas syringae* pv *tabaci* expressing avrPto. Plant Cell, 7:1529-1536

Trojanek J, Pia EK, Scoble J, Muszynska G, and Engstrom L (1996). Phosphorylation of plant proteins and the identification of protein-tyrosine kinase activity in maize seedlings. Eur. J. Biochem., 235:338-344

Underwood W, et al. (2007). The *Pseudomonas syringae* typeIII effector tyrosine phosphatase HopAO1 suppresses innate immunity in *Arabidopsis thaliana*. Plant J., 52:658–672

Van Damme E, Culerrier R, Barre A, Peumans W, et al. (2007). A novel family of lectins evolutionarily related to class V chitinases: an example of neofunctionalization. Plant Physiology, 144:662–672

Vandenborre G, Smagghe G, and Van Damme M (2011). Plant lectins as defense proteins against phytophagous insects. Phytochemistry, 72:1538-50

Vasanthaiah H, Basha S, and Katam R (2010). Differential expression of chitinase and stilbene synthase genes in Florida hybrid bunch grapes to *Elsinoe¨ampelina* infection. Plant Growth Regul, 61:127–134

Wirtz M, and Hell R (2006). Functional analysis of the cysteine synthase protein complex from plants: Structural, biochemical and regulatory properties. Jol of Plant Physiology, 163:273—286

Xu Q, Fu H, Gupta R, and Luan S (1998). Molecular characterization of a tyrosine- specific protein phosphatase encoded by a stress-responsive gene in *Arabidopsis*. Plant Cell, 10:849–57

Xu Q, and Deng X (2010). Cloning and phylogenetic analyses of serine/ threonine kinase class defense-related genes in a wild fruit crop 'chestnut rose'. BMC Research Notes, 3:202

Yamaji Y, Maejima K, Komatsu K, Shiraishi T, Namba S, et al. (2012). Lectin-mediated resistance impairs plant virus infection at the cellular level. Plant Cell, 24:778-793

Yang K, Liu Y, and Zhang S (2001). Activation of a mitogen-activated protein kinase pathway is involved in disease resistance in tobacco. The National Academy of Sciences Current Issue, 98(2):741–746

Yoshimura S, Yamanouchi U, Katayose Y, Toki S, Wang Z, Kono I, Kurata N, Yano M, Iwata N, and Sasaki T (1998). Expression of *Xa*1, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. Proc. Natl Acad. Sci. USA, 95:1663–1668

Zhang S, et al. (1998). Activation of the tobacco SIP kinase by both a cell wall-derived carbohydrate elicitor and purified proteinaceous elicitins from *Phytophthora spp.* Plant Cell 10: 435–450

Chapter 4

BIOCHEMICAL CHARACTERIZATION OF *LEUCAENA*

4.1. Introduction

Leucaena contains mimosine, which is a toxic amino acid harmful to animals. The goal of this chapter was to characterize the mimosine contents of (i) wild type leucanea and (ii) sixteen transgenic leucaena lines that were developed in our lab. Although Jadd Correia was responsible for developing the transgenic plants, I was also involved in tissue culture and genetic transform. We have improved the existing protocol for tissue culture and *Agrobacterium*-mediated genetic transformation to obtain these transgenic plants. The genetic transformation protocol will be described in Jadd Correia's thesis. Here I describe very briefly about these plants so that I can explain the biochemical characterization of the transgenic plants.

Genetic transformation in leucaena has been developed in Dr. Borthakur's Lab previously. Jube and Borthakur (2009, 2010) successfully developed four transgenic leucaena plants expressing the *gus* gene, five transgenic plants carrying *pydA* and four transgenic plants carrying *pydB*. However, the efficiency of transformation was low and therefore needs to be improved. To improve efficiency of genetic transformation of leucaena, I have participated with Jadd Correia who was developing transgenic leucaena plants expressing a *pydA-pydB* fusion gene encoding a dioxygenase (*pydA*) and a hydrolase (*pydB*) from the root-nodule bacterium *Rhizobium* sp. strain TAL1145. We worked on (a) improving tissue culture protocol for shoot and root regeneration; (b) genetic transformation of leucaena using immature embryo as the explants and *Agrobacterium tumefaciens*-mediated transformation method. The goal of this project was to develop transgenic leucaena with reduced mimosine content. It is known that 3-hydroxy-4-pyridone (3H4P) is an intermediate of both mimosine degradation and mimosine biosynthesis. It was argued that by expressing *pyd*A that encodes a dioxygenase specific for 3H4P, the amount of precursor available for mimosine biosynthesis could be reduced in transgenic leucaena. However, earlier study in our lab showed that another gene, *pyd*B, encoding a hydrolase is also required for complete degradation of 3H4P (Awaya et al. 2005). Therefore, a binary plasmid containing a *pydA-pydB* fusion gene was constructed to

transfer *pydA-pydB* to leucaena (Jube and Borthakur 2009). We used this binary plasmid to develop 16 transgenic leucaena plants. PCR analysis by Jadd Correa showed that these plants contained the *pydA-pydB* transgene.

To characterize these transgenic leucaena biochemically, I performed high-performance liquid chromatography (HPLC) analysis to determine mimosine contents.

4.2. Materials and methods

(a) Mimosine isolation:

Earlier reports indicated that mimosine concentration is highest at the youngest shoots (Endrinal B. and Mendoza E. 1979). Therefore, we used leaflets from only young leaves of the leucaena plants for mimosine isolation. One g young leucaena leaves were harvested and immediately frozen in liquid nitrogen. The frozen leaves were powered by using clean cold mortar and pestle to grind quickly. The ground power was then transferred into a 50-ml Falcon tube, to which 5 ml of 0.1N HCl was immediately added. The mixture was homogenized by vortex for 1 min and then kept at room temperature in 1 h before centrifuging at 15,000 rpm for 10 min at 4°C. The aqueous phase was carefully transferred to a syringe and filtered with 0.45 μm-pore-size filter (Corning Incorporated, New York, NY).

(b) HPLC system:

In cooperation with Dr. Li's Lab (University of Hawaii at Manoa), we used HPLC system LC-10AS liquid chromatograph (Shimadzu Inc., Columbia, MD), which includes a UV-VIS detector (SPD-10A), a SIL-10AF auto sampler, and a Dionex acclaim 120 C_{18} column (4.6 x 250 mm). The mobile phase performed was 0.02 M orthophosphoric acid in filtered deionized water, with a flow rate of 1 mL per min. Mimosine was detected at a wavelength of 280 nm. Each injection volume was 20 μL, and the run was 20 min per sample. The range of concentrations of pure mimosine and 3H4P at 1 ppm, 5 ppm, 10 ppm, 30 ppm, 60 ppm and 100 ppm was prepared in 0.1N HCl and used in the HPLC analysis to obtain the standard curves for mimosine and 3H4P, respectively. Mimosine was purchased from BIOMOL International, Plymouth Meeting, PA, and 3H4P was prepared by Vishal Negi from mimosine and purified by HPLC (Negi et al. 2013). Isolation of mimosine from the transgenic lines and the wild type

plants were repeated three times, which were used as three replicates. Mimosine contents were calculated as a percentage of the total dry weight of 1.0 g freshly harvested young leaves.
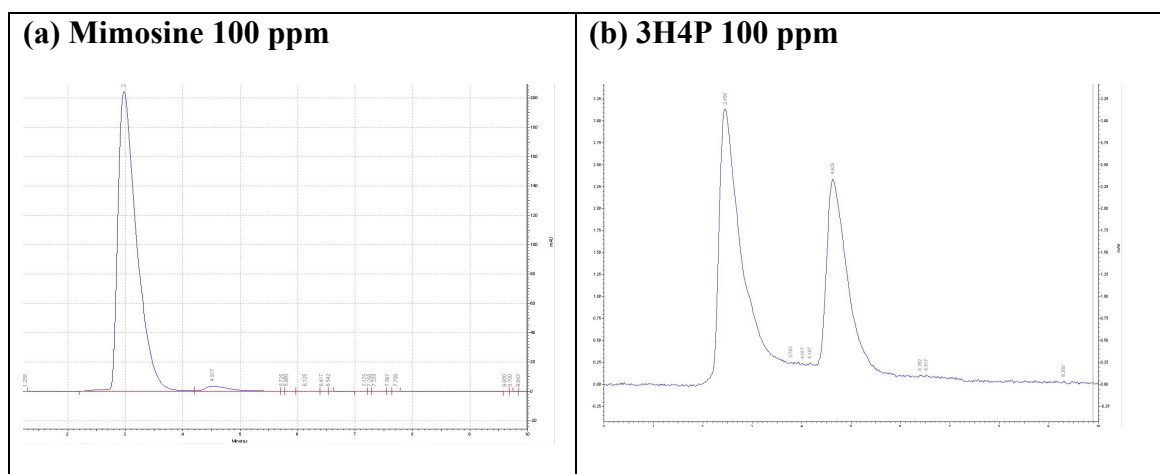
4.3. Results

The peaks for mimosine and 3H4P appeared at 3 min and 4.5 min, respectively (Figures. 4.1a and 4.1b). The peak for HCl showed at 2.5 min. The areas of the mimosine and 3H4P peaks for six different concentrations were plotted as concentration (X)-area (Y) graphs (Figures. 4.1c and 4.1d).

(a) Mimosine contents of the pydA-pydB transgenic leucaena pants

To determine if the wild type K636 leucaena plants grown in the growth room produced the same amounts of mimosine as the plants grown outside, we isolated mimosine from the young shoots from a leucaena plant grown outside. The outside grown plant contained 2.3% mimosine while the wild-type plant grown in the growth room contained only 1.1% mimosine (Figures. 4.2a and 4.2b).

Biochemical analysis by HPLC was performed to determine mimosine content of sixteen independent transgenic leucaena plants and one wild type growing under laboratory and one wild type outside laboratory, used as positive controls.
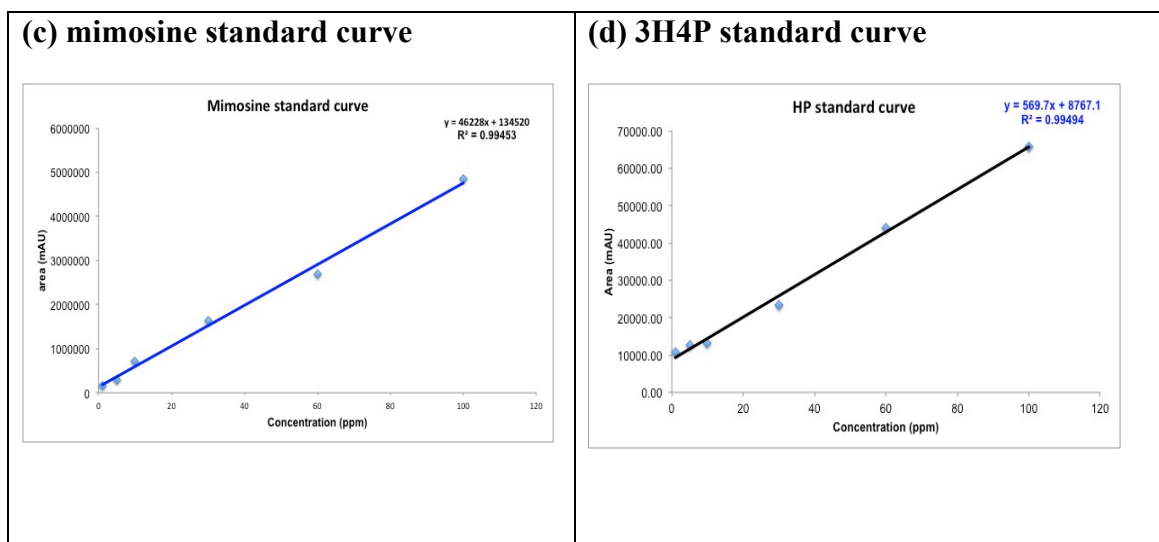

(a) Mimosine 100 ppm

(b) 3H4P 100 ppm

(c) mimosine standard curve

(d) 3H4P standard curve

Figure 4.1: Chromatograph of mimosine 100 ppm, HP 100 ppm, mimosine ad 3H4P standard curve
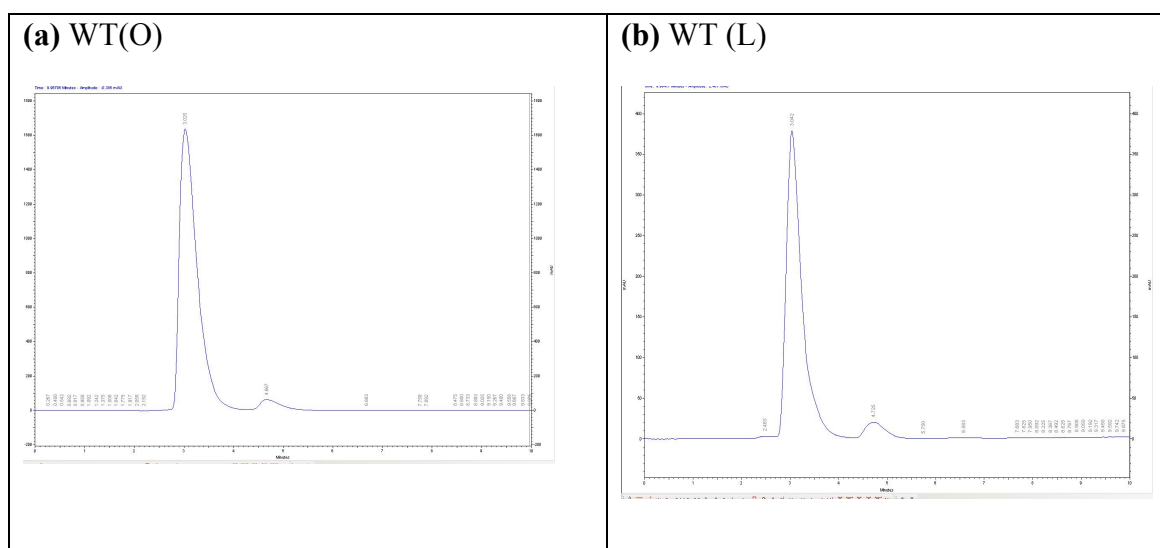

(a) WT(O)

(b) WT (L)

Figure 4. 2: Chromatograph of wild type (grown outside) and wild type (lab growth)

The results from HPLC analysis showed that the mimosine content of the laboratory-grown wild-type leucaena was slightly less than that from the wild-type leucaena grown outside. The transgenic leucaena plants showed varying levels of mimosine contents, which were less than that in the wild-type plant (Figure 4.3 a).   The mimosine contents of the transgenic lines varied between 4.10 to 59.53 % of that of the laboratory-grown wild-type plant (Figure 4.3 b). Transgenic plant # 1, 3 and 5 had the least amount mimosine, which were 4.10 –

6.96 % of that of the wild-type leucaena. These results prove that expression of the *pyd*A-*pyd*B fusion transgene reduces mimosine content in transgenic leucaena (Table 4.1)

Table 4.1: Mimosine content in wild type and transgenic leucaena plants

Explants	% mimosine (in total dry weight)*	% mimosine (comparison with lab grown WT)*
WT (outside grown)	2.38 ± 0.08	
WT (lab grown)	1.51 ± 0.15	100
Explant 1	0.10± 0.02	6.96 ± 1.57
Explant 2	0.46 ± 0.11	30.55 ± 6.53
Explant 3	0.06 ± 0.04	4.10 ± 2.00
Explant 4	0.64 ± 0.24	42.52 ± 15.00
Explant 5	0.09 ± 0.02	5.87 ± 2.03
Explant 6	0.59 ± 0.17	39.06 ± 11.34
Explant 7	0.71 ± 0.12	47.45 ± 10.76
Explant 8	0.50 ± 0.09	32.72 ± 2.67
Explant 9	0.51 ± 0.20	33.81 ± 10.81
Explant 10	0.60 ± 0.09	39.72 ± 5.70
Explant 11	0.69 ± 0.16	46.67 ± 15.59
Explant 12	0.88 ± 0.17	58.66 ± 12.20
Explant 13	0.68 ± 0.16	46.37 ± 14.89
Explant 14	0.77 ± 0.13	51.17 ± 10.47
Explant 15	0.88 ± 0.19	59.53 ± 18.25
Explant 16	0.79 ± 0.09	53.04 ± 10.15

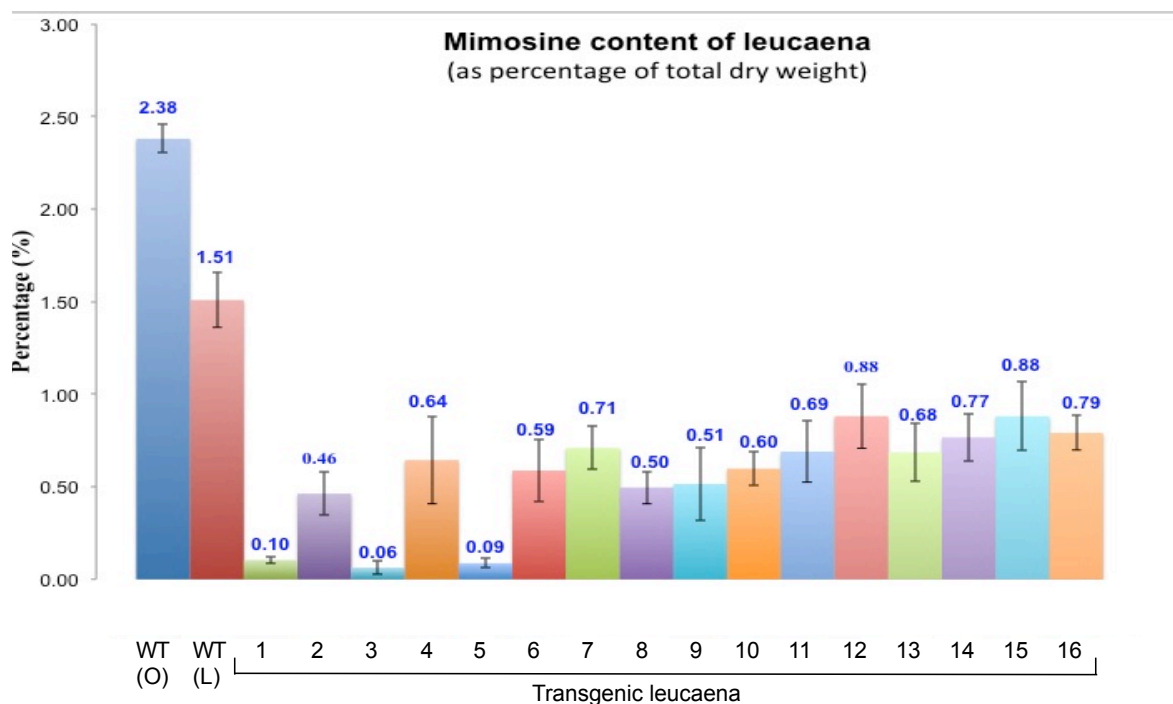*, data are means ± standard deviations from three biological replicates.

Figure 4.3 (a): Mimosine contents of wild-type and transgenic leucaena plants. The amounts of mimosine were expressed as percentage of total dry weight of the plants. WT(O): wild-type leucaena grown outside; WT(L): wild-type leucaena grown in the laboratory. The error bars represent standard deviations of three biological replicates.

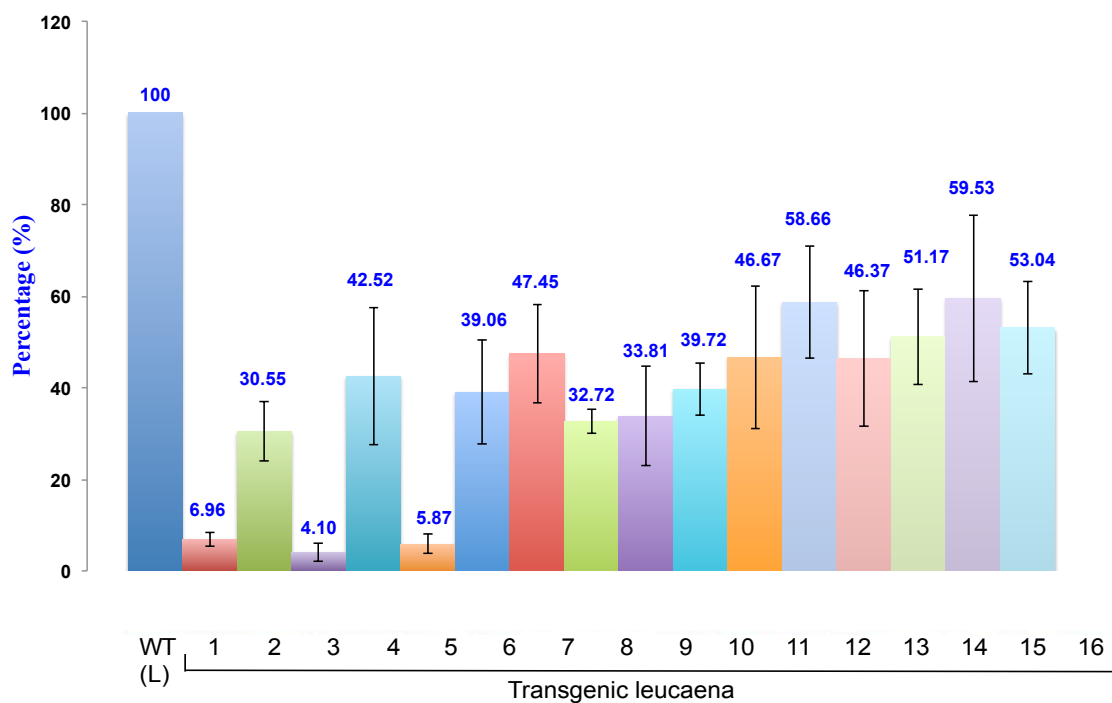
Figure 4.3 (b): Mimosine contents of 16 transgenic leucaena plants, expressed as percentages of mimosine content of a laboratory-grown wild-type (WT) leucaena plant. The error bars represent standard deviations of three biological replicates.
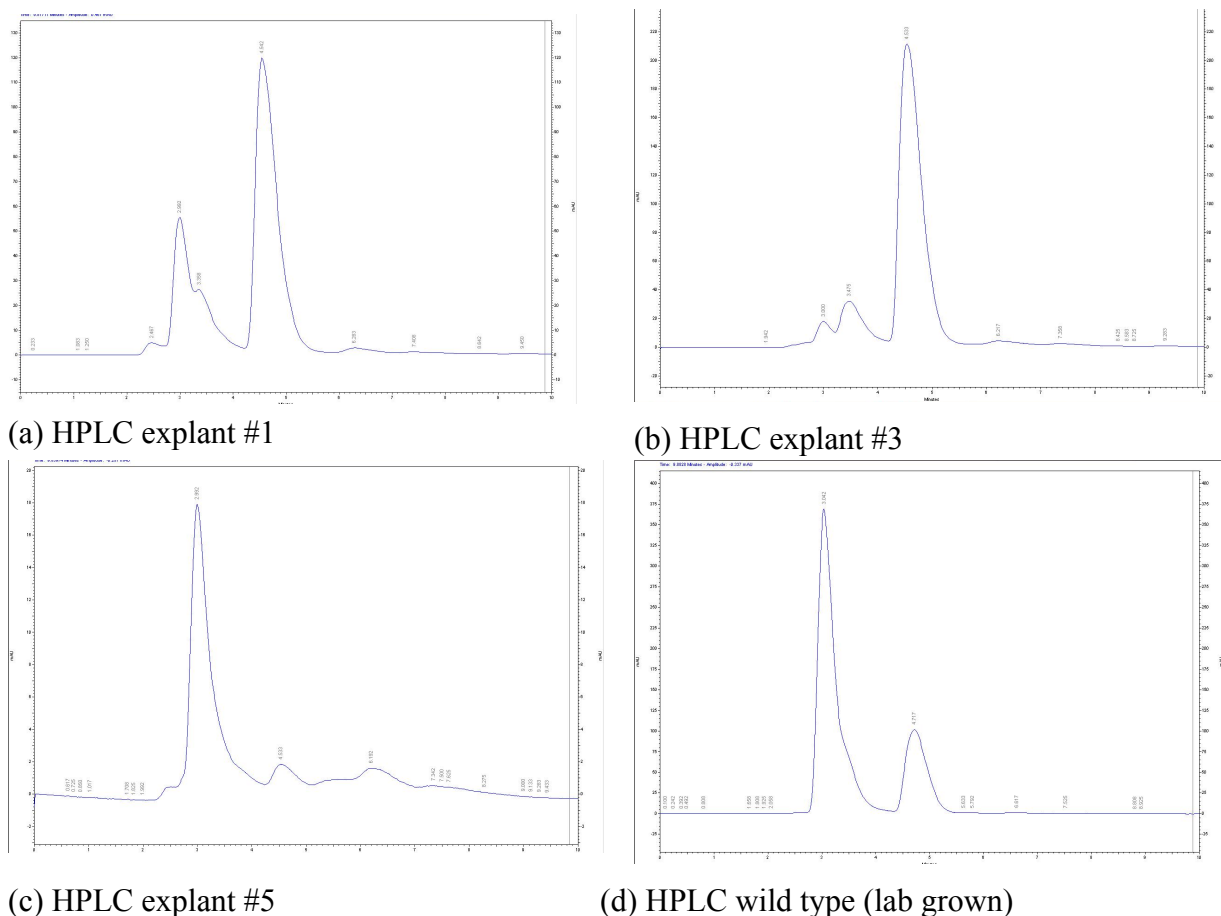
(a) HPLC explant #1   (b) HPLC explant #3

(c) HPLC explant #5   (d) HPLC wild type (lab grown)

Figure 4.4: HPLC chromatograms for mimosine of transgenic leucane plants (a) explant #1, (b) explant #3, (c) explant #5 and (d) wild type (lab grown)

(b) 3H4P contents of the pydA-pydB transgenic leucaena pants

The *pydA-pydB* transgenic leucaena plants are expected to have reduced mimosine as well as 3H4P, which is the substrate for dioxygenase and hydrolase encoded by the *pydA* and *pydB* genes. Therefore, we determined the 3HP4 contents of the transgenic leucaena plants (Table 4.2). The 3H4P contents in wild-type and transgenic leucaena plants as percentage of total dry weight are shown in Figure 4.5a. The 3H4P contents in transgenic leucaena plant as percentage of 3H4P of the lab-grown wild type plant is shown in Figure 4.5b. These figures showed variation in the levels of 3H4P in the wild-type and transgenic leucaena. Suprisingly, the 3H4P content in the transgenic plants #1 and #3 were quite high, considering that these plants had only low amounts of mimosine. Only the other hand transgenic plant #5 had only negligible amount of HP. Thus, plant #5 had very low amounts of mimosine as well as 3H4P. This plant may be considered as the most desirable plant among the 16 transgenic plants. Ten

of the remaining 13 transgenic plants also had low amounts of 3H4P. Transgenic plants #9, #12 and #14 contained moderate amounts (24-37% of the wild-type) of 3H4P.
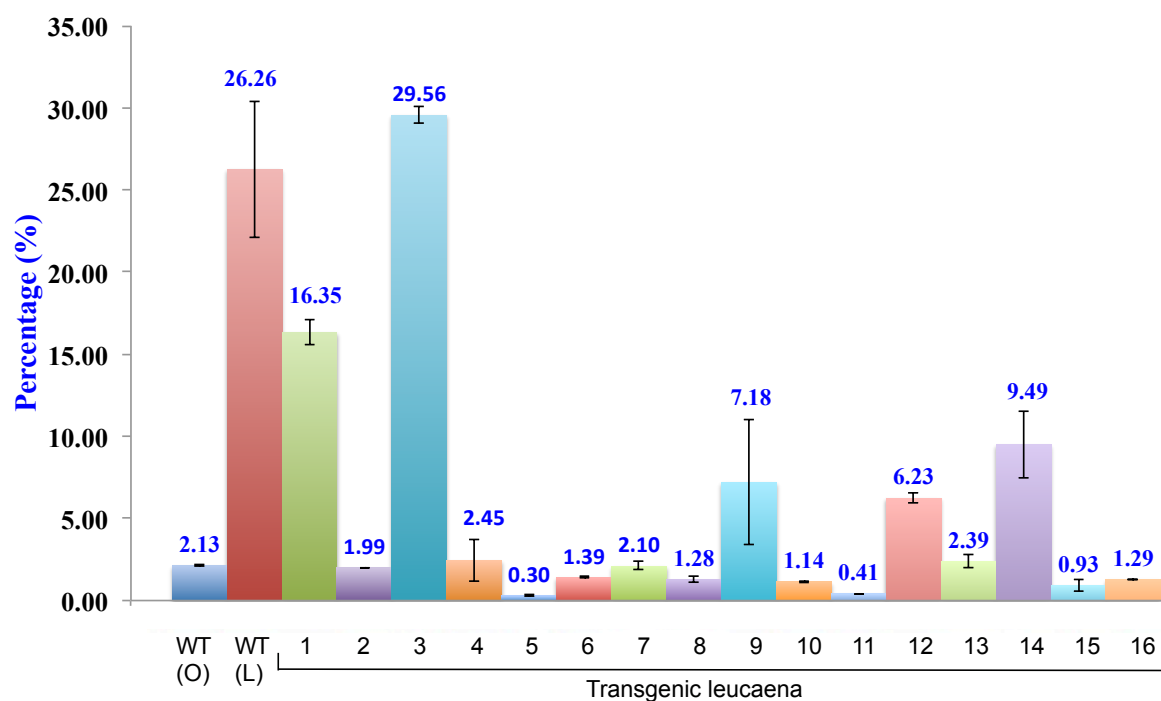

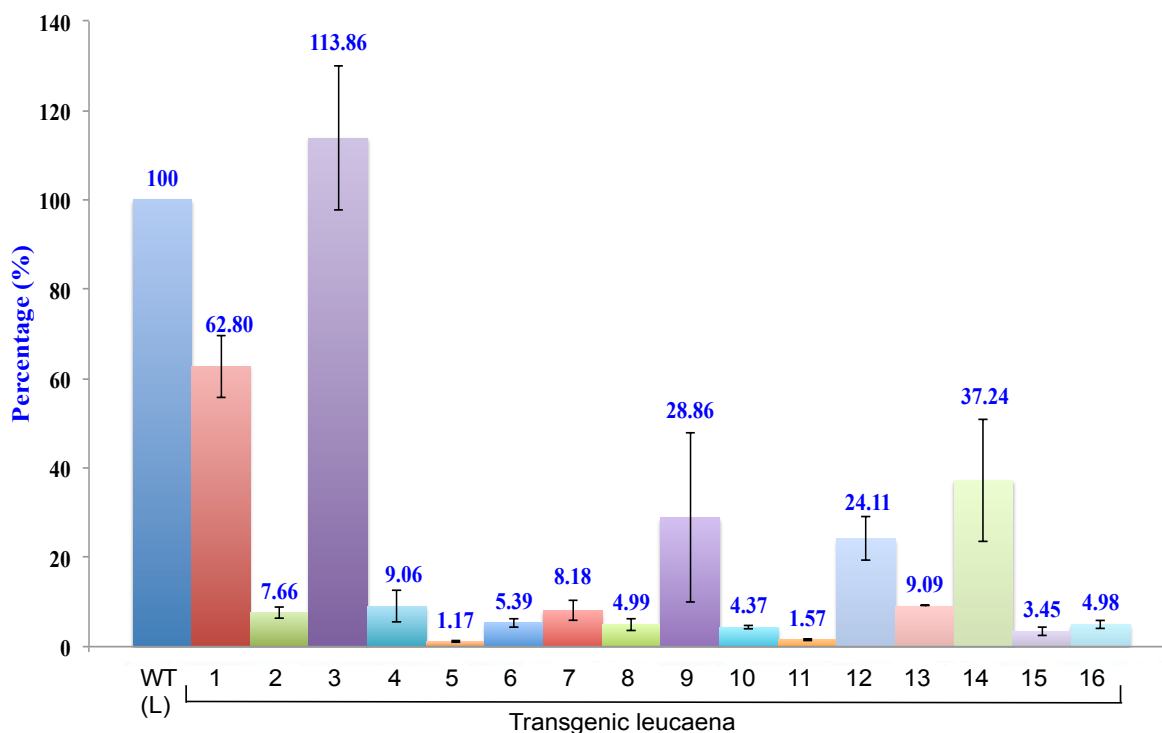**Figure 4.5 (a):** 3H4P content in leucaena in total dry weight

Figure 4.5 (b): 3H4P content in leucaena in comparison with wild type (lab grown)

Table 4.2: 3H4P content in leucarna in total dry weight and comparison with wild type (lab grown)

Explants	% 3H4P (in total dry weight)*	% 3H4P (comparison with lab grown WT)*
WT (outside grown)	2.13 ± 0.03	
WT (lab grown)	26.26 ± 4.14	100
Explant 1	16.35 ± 0.76	62.8 ± 7.01
Explant 2	1.99 ± 0.02	7.66 ± 1.29
Explant 3	29.56 ± 0.49	113.86 ± 16.09
Explant 4	2.45 ± 1.30	9.06 ± 3.51
Explant 5	0.3 ± 0.04	1.17 ± 0.35
Explant 6	1.39 ± 0.05	5.39 ± 1.04
Explant 7	2.1 ± 0.26	8.18 ± 2.27
Explant 8	1.28 ± 0.16	4.99 ± 1.41
Explant 9	7.18 ± 3.8	28.86 ± 19.04
Explant 10	1.14 ± 0.04	4.37 ± 0.54
Explant 11	0.41 ± 0.01	1.57 ± 0.21
Explant 12	6.23 ± 0.28	24.11 ± 4.88
Explant 13	2.39 ± 0.38	9.09 ± 0.03
Explant 14	9.49 ± 2.05	37.24 ± 13.67
Explant 15	0.93 ± 0.39	3.45 ± 0.94
Explant 16	1.29 ± 0.03	4.98 ± 0.90

4.4. Discussion

Leucaena contains mimosine in all parts of the plant including stem, shoots, seeds, pods, leaves and nodules (Soedarjo and Borthakur 1998). The highest amount of mimosine is present in young shoot tips (Soedarjo and Borthakur 1996). We used young shoot tips for determining mimosine content because: (i) it is easy to isolate mimosine from young shoot tips, (ii) this portion of plants is preferred by grazing cattle. Early reports indicated that mimosine content of leucaena dry matter to be 3-5% (Soedarjo and Borthakur 1996). However, in this study, we found that mimosine content of wild type plants growing outside was much less than

mimosine content of leucaena foliage reported in the literature. This variation in estimate of mimosine content may be due to different growth conditions, seasonal variation in mimosine content, and environmental conditions such as temperature, rainfall and soil nutrition. It was not surprising that mimosine content of lab-grown wild-type plant was less than in the wild type plant grown outside. The leucaena plants in the lab were grown in pots at 25°C under fluorescent lights. This growing condition is very different from the natural growth conditions outside.

All transgenic plants produced reduced amounts of mimosine compared to the wild type plants. However, the mimosine content varied widely among the transgenic plants. Only three transgenic plants (#1, 3, 5) showed big reduction in mimosine content compared to the lab-grown wild-type plant. The amounts of mimosine in the remaining 13 transgenic plants were 30.55-59.53% of the amount of mimosine produced by the wild type plant. This variation indicated different levels of expression of the transgene in these plants. The variation in the transgene expression may arise due to position effects of transgene insertion; transgene is generally inserted randomly at different positions in different chromosomes. Some locations of chromosomes may not be transcriptionally active; therefore genes inserted in certain locations may not be expressed or expressed at low levels. Transgenic plants 1, 3, and 5 contain much lower amounts of mimosine compared to the wild type plant or other transgenic plants. It is possible that these three plants have more than one copy of the transgene. Jube and Borthakur (2010) found that one transgenic leucaena plant with two copies of *pydA* gene. To confirm if these leucaena plants carry more than one copy of *pydA-pydB*, Sourthern analysis of these plants will have to be carried out in the future. The Southern analysis may show different sizes of hybridizing bands, which will indicate that the positions of the transgene insertion vary among these plants. The different levels of transcription of transgenes in different transgenic plants will be determined by reverse transcription PCR and real-time PCR in the future. Finally, the overall results of transcription and translation of the *pydA-pydB* transgene in these plants will be determined Western analysis in the future.
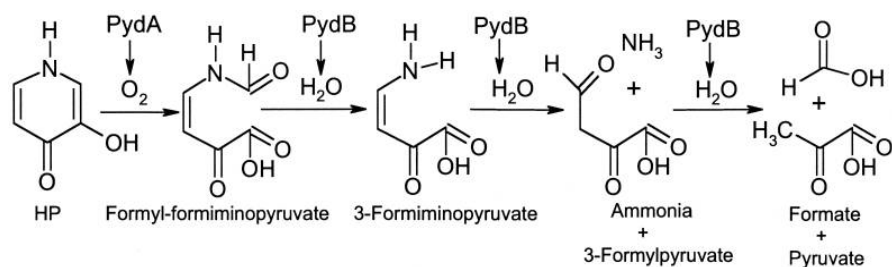
Figure. 4.6: The proposed pathway of HP degradation determined by the *pyd* genes in *Rhizobium* sp. strain TAL1145 (Adapted from Awaya et al. 2005).

We expected that transgenic plant expressing the *pydA-pydB* transgene will contain reduced amounts both mimosine and 3H4P. We found only one plant (#5) that satisfied this expectation. Plants #1 and #3, although contained only low amounts of mimosine, they contained surprisingly high amounts of 3H4P. 3H4P measurements were based on areas of the peak corresponding to the 3H4P elution time in the HPLC chromatograph. However, it is possible that peak that appeared at ~4.5 min in HPLC is a derivative of 3H4P, which was produced from 3H4P through the dioxygenase reaction. The product of dioxygenase reaction on 3H4P may be formyl-forminopyruvate (Awaya et al. 2005; Figure 4.6). Formyl-forminopyruvate has slightly higher molecular mass than 3H4P due to incorporation of one molecule of O_2. It is possible that formyl-forminopyruvate has the same elution time in HPLC as 3H4P. We are planning to determine the identity of this compound through MS-MS spectrometry in the near future.

References

Adams R, Cristol S, et al. (1945). The Structure of Leucenol. I. Journal of the American Chemical Society, 67(1): 89-92

Awaya J, Fox P, and Borthakur D (2005). *pyd* genes of *Rhizobium* sp. *strain* TAL1145 are required for degradation of 3-Hydroxy-4-Pyridone, an aromatic intermediate in mimosine metabolism. J. Bacteriol. 187(13):4480

Brewbaker J and Hylin J (1965). Variation in mimosine content among *Leucaena* species and related *Mimosaceae*. Crop Science, 5(4): 348-349

Endrinal B and Mendoza E (1979). *Mimosine levels* of *different* parts and *height classes* of *Leucaena leucocephala*. (Lam.) de Wit . Phil . J . Crop Sci . 4(1):48-52

Hegarty P, Court R, et al. (1964). The determination of mimosine and 3, 4-dihydroxypyridine in biological material. Australian Journal of Agricultural Research, 15(1):168-179

Jones J (1979). The value of *Leucaena leucocephala* as a feed for ruminants in the tropics. World Animal Review, 31:13-23

Jube S, and Borthakur D (2010). Transgenic *Leucaena leucocephala* expressing the *Rhizobium* gene *pyd*A encoding a meta- cleavage dioxygenase shows reduced mimosine content. Plant Physiology and Biochemistry, 48(4):273-278

Jube S, and Borthakur D (2009). Development of an *Agrobacterium*-mediated transformation protocol for the recalcitrant tree-legume *Leucaena leucocephala* using immature zygotic embryos. Plant Cell, Tissue and Organ Culture (PCTOC): Journal of Plant Biotechnology, 96: 325-333

Negi V, Bingham J.P, Li Q, and Borthakur D (2013). *mid*D-encoded 'rhizomimosinase' from *Rhizobium sp. strain* TAL1145 is a C–N lyase that catabolizes L-mimosine into 3-hydroxy-4-pyridone, pyruvate and ammonia. Amino Acids, 44:1537–1547

Soedarjo M, and Borthakur D (1996). Simple procedures to remove mimosine from young leaves, seeds and pods of *Leucaena leucocephala* used as food. Int J Food Sci Technol, 31:97–103

Chapter 5

TISSUE CULTURE REGENERATION OF LEUCAENA LEUCOCEPHELA

5.1. Introduction

Leucaena leucocephala (leucaena) is known to be a recalcitrant legume species, which is difficult for genetic transformation.cDifficulties in *Agrobacterium tumefaciens*-mediated transformation may be partly due to problems in tissue culture regeneration of leucaena from vegetative tissues of the plant. So far, successful tissue culture regeneration protocols for leucaena involve immature embryo (Jube et al. 2009, 2010; Pal and Borthakur 2012), cotyledonary nodes and nodal segments (Dhawan and Bhojwani 1985, Shaik et al. 2009), and nodal explants (Rastogi et al. 2008).

Vegetative propagation of leucaena had been attempted by several groups of researchers in the past. *L. leucocephala* has been reported to be successfully propagated by air-layering (Osman 1986) and grafting (Brennan and Mudge 1998). Osman (1986) used foam-air-layer of stems of 2-4 month old leucaena seedlings to induce root development on layered region of the branches. Air-layering was done by using moist foam-rubber cubes. The air-layered branches that formed adventitious roots were excised and planted as new seedlings.

Brennan and Mudge (1998) had developed graft guide (single-bud splice) technique by performing single bud splice grafting for clonal propagation of leucaena. Small diameter shoots and for maximizing available scion wood were used for grafting. The younger rootstock seedlings (3-6 moth-year old leucaena) were used as the rootstock. The grafted explants were grown in the greenhouse under controlled moisture and temperature before transferring into fields.

Dick et al. (1999) had successfully induced roots in leucaena stem cuttings that carried single-node with leaves or without leaves, from one-year-old greenhouse-grown leucaena. Single-node cuttings were dipped in a commercial rooting powder (Seradix, May & Baker Ltd.; active ingredient 0.8% IBA) and placed in rooting media containing quartz grit (0.001 to 12.7 mm) in a temperature-controlled glasshouse. Thirty nine percent of the stem cuttings without leaves and 71% of stem cuttings with leaves formed roots.

Shi et al. (2006) developed a vegetative propagation method for leucaena using stem

cuttings from plants of different ages diploid, triploid and tetraploid leucaena. They used a mixture vermiculite and perlite as the rooting medium in pots. They found that only cuttings with half or more leaves formed roots. Both one-node and two-node cuttings showed the same rooting ability.

Dhawan and Bhojwani (1985) used single node segments of 2-month-old leucaena for in vitro vegetative propagation of leucaena. They regenerated seedlings from 2.2 cm long epicotyl of 2-week-old plants grown under aseptic condition. They also found that addition of 400 μM glutamine to the growth medium reduced precocious leaf drop.

Rastogi et al. (2008) developed an in vitro regeneration protocol for leucaena using single node from mature tree-derived nodal explants via organogenesis. They first induced the nodal explants to form calli, which were then induced to form somatic embryos and shoots.

Shaik et al. (2009) used cotyledonary nodes and cotyledons from one-week-old aseptically grown seedlings for in vitro regeneration in leucaena by using cytokinins thidiazuron (TDZ) and N6-benzyladenine (BA) with half full MS media for root induction. Multiple shoots from cotyledonary nodes and the regenerated shoots successfully rooted in the presence of the rooting auxin α-Naphthelinacetic acid (NAA).

From the literature review described above, it is apparent that there is so far no published report on tissue culture regeneration of leucaena using tissues from plants grown under non-sterile conditions. In nearly all the reports, the explants for tissue culture were taken from either immature pods or from young seedlings grown under sterile conditions in the laboratory. Although Rastogi et al. (2008) used nodal explants from mature trees, they used callus formation and organogenesis as intermittent steps before developing shoots and the procedure took. Their methods took about 6 months time and intense efforts to obtain the regenerated seedlings. Therefore, it is important to develop a relatively easy tissue culture regeneration protocol, without organogenesis as an intermittent step, for leucaena using explant material from plants grown outside or in the laboratory under non-sterile conditions. This becomes particularly important when researchers develop transgenic leucaena plants that need to be evaluated for various physiological, biochemical and agronomic characteristics. Each transgenic leucaena plant obtained through *Agrobacterium tumefaciens*-mediated transformation arises due to a unique insertion of the transgene into the leucaena genome.

95

Therefore, the locations of transgene insertions in different transgenic leucaena plants containing the same transgene may be also different. In deed, through Southern hybridization, Jube and Borthakur (2010) have shown that the size of the transgene-containing fragments varied in different transgenic leucaena plants, and some plants contained more than one transgene insertions. These plants may also have different characteristics; for example, the transgenic leucaena plants obtained by Jube and Borthakur (2010) each had different amounts of mimosine. Our own experience with 16 transgenic plants (chapter 4) show that they are different from each other in mimosine content. Therefore, to study any of these plants in great detail for their growth and biochemical characteristics under different environmental conditions, each of these plants needs to be multiplied through vegetative propagation.

At the Waimanalo Research Station of the University of Hawaii, Dr. James Brewbaker and his colleagues have developed several triploid leucaena lines by intercrossing between tetraploid *L. leucocephala* and diploid *L. retusa*, between *L. leucocephala* and diploid *L. diversifolia*, and between tetrapoid and diploid species of *L. diversifolia* (Brewbaker and Sorenson 1990). These triploid plants show hybrid vigor and are much larger than the parent species. Unfortunately, as any triploid plant, these plants do not form viable seeds and are sterile. The only way such plants can be propagated is through vegetative propagation. There are only a few triploid leucaena plants and they are only at the Waimanalo Research Station. Although these plants have many superior attributes, they could not be planted elsewhere. Therefore, it is highly essential that a tissue culture regeneration protocol be developed for easy and rapid multiplication of these triploid plants.

5.2. Materials and Methods

(a) Explant material:

Shoot tips of 40-50 mm in length and 2-3 mm in diameter, and stem sections containing single node with approximately 20 mm internode segments on both sides of the node, from a six-month-old transgenic leucaena plant and a wild type leucaena (variety K636) that had been grown in the growth room at 25 ˚C with 16 h light/ 8 h dark photoperiod were selected by using clean scissors to cut at 45˚ angles.

(b) Surface sterilization:

Following excision from the mother plant, the explants were dipped into 150 ml sterilized water containing 1% sodium hypochlorite Tween 20 (1 drop per litter). The samples were stirred in the sterilizing solution for 30 min, and then carefully rinsed with 5 times with sterilized water. Using a sterilized scissor, 2-3 mm sections of the open ends of the explants were removed. A sterile blade was then use to make sharp cuts at 45° at the open ends. The segments were immediately transferred to ~100 ml of sterile antioxidant solution, which contained ascorbic acid 75 mg/L and citric acid 50 mg/L, and kept emerged for 15 min. The explants were placed on sterilized filter paper to remove excess moisture before transferring to culture media.

(c) Tissue culture medium for shoot induction:

The explants were placed on half MS salt medium containing 0.89 μM benzyladenine (BA) and 1.1μM naphthalene acetic acid (NAA), 2% phytogel, and 10% of activated charcoal in Magenta boxes. The Magenta boxes with explants were placed under 16 h light and 8 h dark photoperiod with 60 μM m^{-2} Sec^{-1} light intensity at 25 °C. The plates were checked for possible contamination every day. After 4 weeks, news shoots were transferred to rooting media.

(d) Tissue culture media for root induction:

The explants with shoot were transferred to Magenta boxes containing root-inducing medium. The root-inducing medium contained 2/3rd MS salt, 0.2 mg/L NAA, 1.0 mg/L IBA, 0.1 mg/L kinetin, 30 g/L sucrose and 10% activated charcoal.

(e) Transfer to soil:

The rooted explants were transferred to 6x6 pots containing horticultural soils. Before transferring to the pot, the explants were cleaned with sterile water to completely remove the rooting medium.

5.3. Results

Since one of the mother plants from which the explants were obtained was a newly developed transgenic leucaena plant, only limited number of explants could be excised without seriously hurting the mother plant. Two shoot tips and five nodal explants were taken from the transgenic leucaena plant and equal number of explants was taken from a non-transgenic plant.

The surface sterilization treatment was fully effective since no contamination was observed in shoot tip and nodal explants. Leucaena explants are known to produce phenolic compounds, which give brown color to the media. I used activated charcoal to absorb any phenolic compound released by the cut ends of the explants. I did not observe any browning on the medium. After two weeks of culturing in the shoot-inducing medium, all the leaves from the shoot tip explants fell down. However, after 1-2 weeks of loosing the leaves, the shoot-tip explants showed growth of very small new buds; only one bud was observed per shoot-tip explant. The nodal explants showed some variations in shoot induction. Among the five nodes, the oldest node formed a bud for shoot formation after one week of culturing in shoot-induction medium, while the remaining four younger nodes took 21-25 days for developing small buds for shoot formation. There were no differences between the explants from transgenic and non-transgenic plants for any of the steps in shoot induction (Figure 5.1)

Surprisingly, 2 of the 4 wild-type explants formed roots after 9 weeks of culturing; while transgenic explants have shown no root formation until now (9 weeks). However, transgenic explants are growing and appear to be healthy. One of the two wild-type explants that formed roots has been moved to pot containing vermiculite wetted with Hoagland plant nutrient solution (Figure 5.2)

(a)
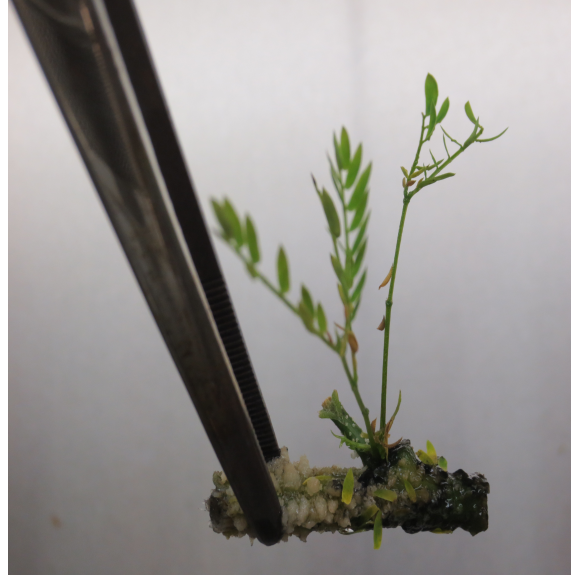
(b)
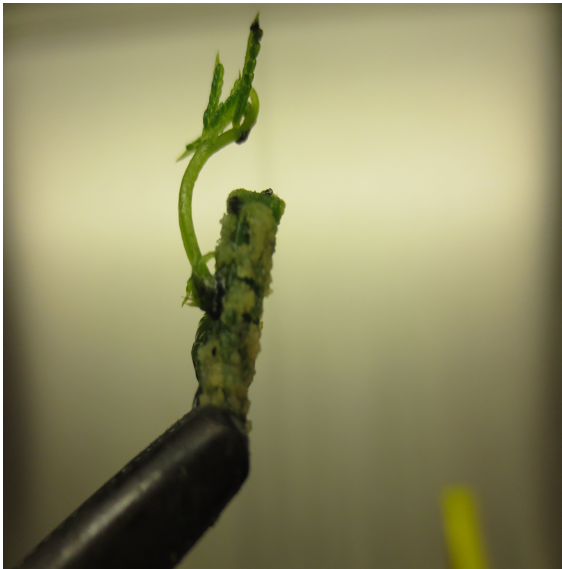
(c)

(d)

Figure 5.1: The new shoots induction after 4 weeks micro-propagation from transgenic leucaena number 3 (a, b) and lab grown wild type (c, d)
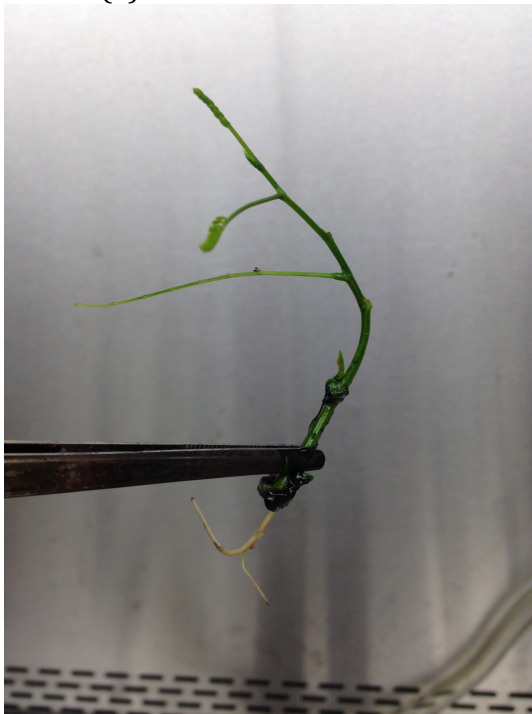
(a)

(b)

(c)

(d)

Figure 5.2: The new shoots induction after 9 weeks micro-propagation from transgenic leucaena number 3 (a, b) and lab grown wild type (c, d)

5.4. Discussion

We have been successful in establishing explants from lab-grown wild-type leucaena and transgenenic plant #3 using shoot tips and stem sections as the starting material. Generally, for rooting of transgenic explants following transformation, we pass them through a 'shoot elongation' step for two weeks, when we allow the newly formed shoots to elongate by growing them in Magenta boxes on medium with specific concentrations of hormone, before transferring them to root-induction medium. But for regeneration of explants from existing transgenic plants, we bypassed the shoot elongation step. Another difference from the tissue culture protocol that we use in developing transgenic plants is that for the regeneration protocol we had an initial step of shoot growth by adding low concentrations of BA and NAA in the medium.

In the plant tissue culture media, we added activated charcoal to absorb phenolic compounds released by the leucaena explants. Moreover, in the root induction medium, the darkness of media due to charcoal prevents light from reaching the newly developing roots. It is known that light affects on root growth of maize by inducing production of inhibitors in the root cap (Pilet 1976). In addition, ethylene may be a mediator of light-induced inhibitor of root growth (Eliasson and Bollmark 1988). Therefore, addition of activated charcoal in culture media was helpful, and as a result two wild-type explants formed roots. We believe the remaining wild-type and transgenic explants will form roots within a few weeks.

Preliminary observation showed that transgenic explants are taking longer time for rooting than the wild-type explants. This may be caused by biochemical and physiological differences between the transgenic and non-transgenic explants. The apparent biochemical differences between transgenic and wild type is the transgenic explants contain reduced amounts of mimosine and 3H4P compared to the wild-type explants. In these explants, degradation of 3H4P, which is precursor for mimosine synthesis, at least partially disrupts the mimosine biosynthesis pathway. The biosynthetic pathway for 3H4P is unknown. Its biosynthesis may follow a pathway similar to that for niacin biosynthesis, because niacin is structurally quite similar to 3H4P in having 1N and 5 C in the aromatic ring (Figure 5.3). So degradation of 3H4P may affect biosynthesis of other compounds, which may contribute to some physiological differences between wild-type and the *pydA-pydB* transgenic leucaena. At

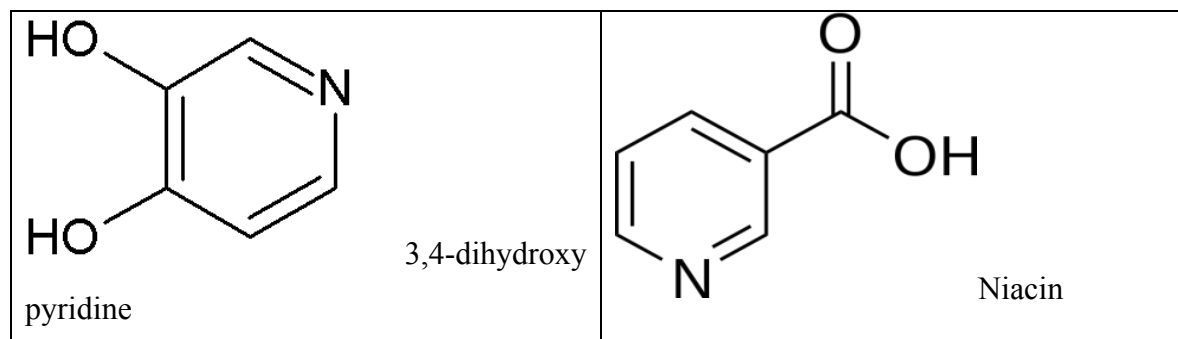this stage, we do not know about specific physiological differences between wild-type and transgenic leucaena.

Figure 5.3: Structures of 3,4-dihydroxypyridine (isomer of 3H4P) and Niacin

In this experiment, I performed micro-propagation of transgenic plant #3, which showed lowest mimosine content (chapter 4). This micro-propagation method will be applied immediately for another two transgenic plants with low mimosine content (#1 and #5). The successful regeneration of the transgenic plants through the protocol developed and tested in this chapter will be applied for micro-propagation of important transgenic leucaena plants in the future so that they can tested thoroughly for biochemical, physiological, and agronomic characteristics through replicated experiments.

References

Brennan E, and Mudge K (1998). Clonal propagation of *Leucaena* by single bud splice grafting with a new grafting tool, and by modified veneer crown grafting. New Forests, 15:283–297

Brewbaker J, and Sorensson C (1990). New tree crops from interspecific *Leucaena* hybrids. Advances in new crops. Timber Press, Portland, OR. 283-289.

Dhawan V, and Bhojwani S (1985). In vitro vegetative propagation of *Leucaena leucocephala* (Lam.) de Wit. Plant Cell reports, 4(6):315-318

Dick J, Magingo F, Smith R, and Mcbeath C (1999). Rooting ability of *Leucaena leucocephala* stem cuttings. Agroforestry Systems, 42:149–157

Eliasson L, and Bollmark M (1988). Ethylene as a possible mediator of light-induced inhibition of root growth. Physiologia Plantarum, 72(3):605–609

Jube S, and Borthakur D (2010). Transgenic Leucaena leucocephala expressing the Rhizobium gene pydA encoding a meta- cleavage dioxygenase shows reduced mimosine content. Plant Physiology and Biochemistry, 48(4):273-278

Jube S, and Borthakur D (2009). Development of an Agrobacterium-mediated transformation protocol for the recalcitrant tree-legume Leucaena leucocephala using immature zygotic

embryos. Plant Cell, Tissue and Organ Culture (PCTOC): Journal of Plant Biotechnology. 96: 325-333

Osman A (1986). Novel methods for the vegetative propagation of Leucaena (*Leucaena leucocephala*) under Laboratory conditions. Stagegies research finding in leucaena, 94-97

Pal A, Negi V, Khanal S and Borthakur D (2012). Immunodetection of Curcin in seed meal of *Jatropha Curcas* using polyclonal antibody developed against Curcin-L. Current Nutrition & Food Science, 8:213-219

Pilet P (1976). The light effect on the growth inhibitors produced by the root cap. Planta (Berl.) 130:245-249

Rastogi S, Rizvi S, Singh R, and Dwivedi U (2008). *In vitro* regeneration of *Leucaena leucocephala* by organogenesis and somatic embryogenesis. Biologia Plantarum, 52(4):743-748

Shaik N, Arha M, Nookaraju A, Gupta S, Khan B, et al. (2009). Improved method of *in vitro* regeneration in *Leucaena leucocephala* – a leguminous pulpwood tree species. Physiol. Mol. Biol. Plants, 15(4):311-318

Shi X, and Brewbaker J (2006). Vegetative propagation of *Leucaena* hybrids by cuttings. Agroforestry Systems, 66:77-83

Sorensson C (1987). Interspecific hybridization in *Leucaena* Bentham. MS thesis, Univ of Hawaii at Manoa.

Chapter 6

GENERAL DISSCUSION

Leucaena is called a 'miracle-tree' of the tropical and subtropical regions of the world because it has many desirable attributes, including resistance to diseases, high tolerance to drought, ability to grow in relatively poor soils, and high protein content of its foliage. Leucaena has no known diseases; it is highly resistant to pathogenic organisms. It is also resistant to insect-pests; however, it is susceptible to psyllid, and pod borer (Brebaker and Soresson 1994, National Academy of Sciences report 1977). Resistance to insects may be due to the presence of mimosine. It has been shown that psyllids' gut contains microorganisms that degrade mimosine and thereby psyllids withstand mimosine toxicity (Kamada et al 1996). It is likely that pod borers also have similar mimosine- degrading mechanism in the gut because of which it can infest leucaena seeds containing mimosine. Moreover, leucaena can grow in relatively dry conditions and survive prolonged drought stresses, in which not many other plants can survive. Its high level of drought resistance may be partially due to its deep and long root systems (National Academy of Sciences report 1977). In addition, leucaena also grows well in a wide range of soil types, from crop fields and grasslands to dry rocky mountains (Tawata et al. 2008, Shelton and Brewbaker 1994). Although it prefers neutral to slightly alkaline soils, it can also grow in soils with slightly acidic pH. Because of its high adaptability to soil types and variable environmental conditions, leucaena has established well in nearly all-tropical and subtropical countries (Shelton and Brewbaker 1994). The high-yielding varieties of leucaena developed by Dr. Brewbaker and his colleagues at the University of Hawaii at Manoa are now widely grown in Australia, South-East Asia, India, Pakistan, Africa, Central America countries and some parts of the United States including Florida, Texas and New Mexico.

Based on numerous valuable attributes of the miracle tree, it was expected that its genome contain many valuable genes that may be absent in other plants, and which can be used for developing disease-resistant and drought-tolerant varieties crop plants. We expected to identify some of leucaena resistance genes by transcriptome analysis. Total RNA was extracted and mRNA was purified from leucaena plants, converted to cDNA, and then sequenced by the Illumina method. The transcriptome sequence data obtained from both shoots and roots

contained 355 of very long sequences (>3.0 kb), indicating that the sequence data were good and reliable. These longest sequences in shoots and roots are transcription factors, DNA repair protein UVH-3, TMV resistance protein N-like, U3 small nucleolar RNA-associated protein 20-like, and dicer-like protein 4-like, which are found in all higher plants. Moreover, among >2 million sequences, 1.82% sequences were longer than 500 bp, showing that the transcriptome sequences we obtained were of good quality. Since roots are primarily affected by various drought and soil nutrition stresses, we expected that certain genes would be expressed at high levels in roots. Surprisingly, 80 early responsive to dehydration (ERD) related genes were found in shoots. We also identified 600 NB-LRR sequences (>100 bp) in the shoots and roots. This is much higher than the number of NB-LRR sequences found in other plants; for example, *Arabidopsis* contains only 125 NB-LRR. Other class of resistance genes identified was chitinase. A total of 234 chitinase sequences (>100 bp) were identified. The other interesting genes that were identified from the leucaena transcriptome sequences are shown in Fig. 6.1.

Transcriptome analysis is a suitable tool for identification of genes that are transcribed. The sequence data obtained from the transcriptome analysis of leucaena can only identify the genes that are expressed in roots and shoots. However, it does not show the level of gene expression. It also does not identify genes that express only under certain conditions. In order of identify genes which express under drought conditions, total RNA was isolated from shoots and roots of leucaena plants grown under normal and drought-stress conditions, and sent to the University of Iowa for microarray analysis. The results of microarray have not been available at the time of completing the thesis. It is expected that microarray will identify genes that express under drought condition. The genes for resistance to disease and environmental stresses may be either expressed constitutively or they are induced in certain of stress environments. The leucaena plants that were used to isolate mRNA for transcriptome analysis were grown under normal laboratory conditions on vermiculite in pots under fluorescent lights and were not exposed to any plant pathogens and stress environments. In sprite of that, we could identify so many genes for disease and environmental stresses. These genes may be expressed constitutively.

From the transcriptome analysis, we have obtained only partial sequences of transcripts. In order to characterize genes for resistance to environmental stresses, the full-length

sequences for specific genes will be needed. This can be obtained in the future through 5' and 3'-RACE of specific partial sequences.
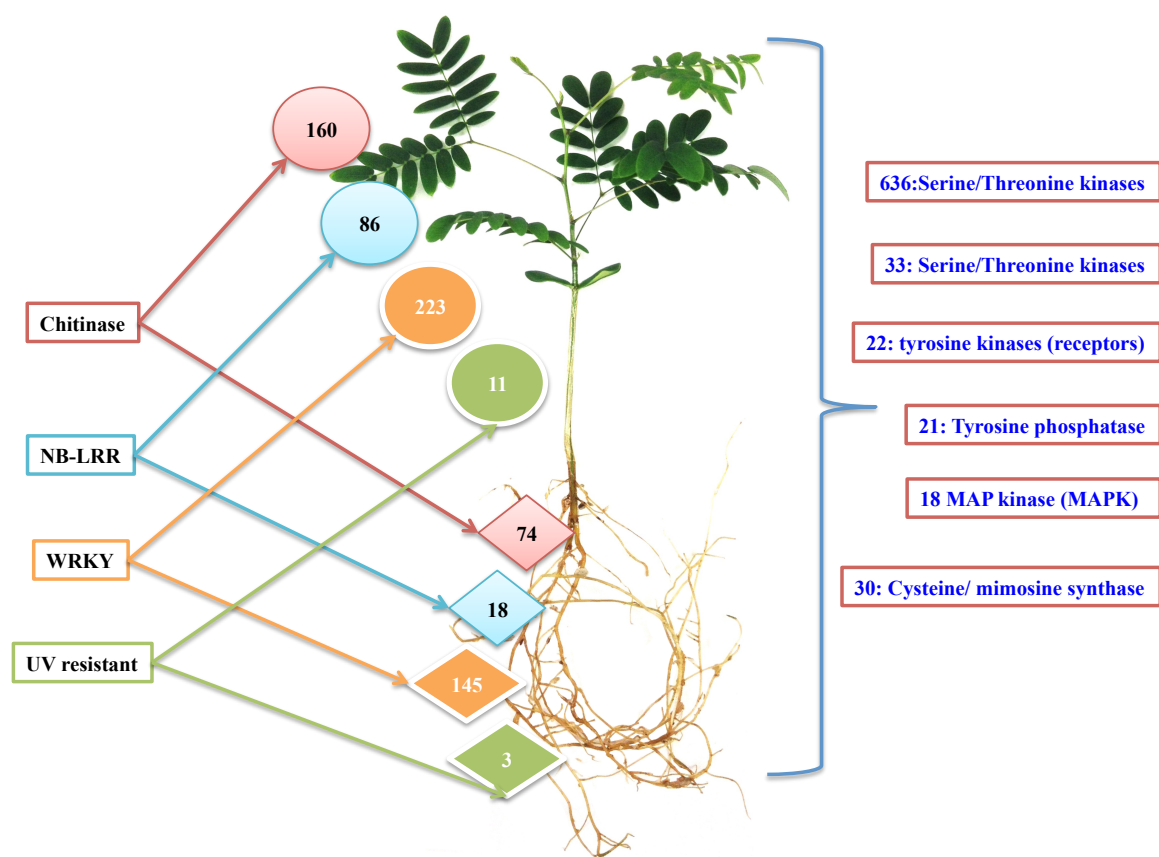
Figure 6.1: The number of various resistance genes identified from leucaena transcriptome analysis.

Table 6.1: The list of some important genes for future research

	Query ID	Length (bp)	Description
1	971835	1154	Chitinases
2	1673697	1955	Chitinase class Ib
3	1496814	426	Chitinase class II
4	1652009	885	Chitinase class III
5	1632644	1306	Acidic mammalian chitinase-like
6	1670725	3868	TIR-NBS-LRR type disease resistance protein
7	1657587	3081	NBS-LRR resistance-like protein 4G

8	1655482	2506	Cc-nbs-lrr resistance protein
9	1658426	1908	TIR-NBS-LRR-TIR type disease resistance protein
10	1664932	1078	BED finger-nbs-lrr resistance protein
11	1669243	1498	Dehydration-responsive protein RD22
12	1666699	2054	Dehydration responsive element-binding protein
13	1666957	859	WRKY-like drought-induced protein
14	1631512	654	Cold and drought-regulated protein CORA
15	114105	127	Drought-induced protein RPR-10
16	1646154	1065	Multi-sensor signal transduction histidine kinase
17	1596066	807	Signal transduction histidine kinase
18	1632515	502	Histidine kinase osmosensor protein
19	138168	597	Histidine kinase of the competence regulon, ComD
20	1673656	3978	Histidine kinase 5-like
21	1669168	4050	CBL-interacting serine/threonine-protein kinase
22	1673844	1259	LRR receptor-like serine/threonine-protein kinase FEI
23	47465	541	Microtubule-associated serine/threonine-protein kinase 4
24	1676044	1648	Calcium/calmodulin-dependent serine/threonine-protein kinase 1-like
25	1673048	1302	G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300-like
26	1543587	523	ALK tyrosine kinase receptor
27	1671721	2403	PTI1-like tyrosine-protein kinase At3g15890-like
28	1655816	1057	C-type lectin receptor-like tyrosine-protein kinase At1g52310-like
29	1665181	557	Tyrosine-sulfated glycopeptide receptor 1-like precursor
30	1666063	951	Hepatocyte growth factor-regulated tyrosine kinase substrate
31	1805304	1640	Tyrosine-protein phosphatase 3-like
32	1807762	682	Protein-tyrosine-phosphatase IBR5-like
33	1646911	559	Protein-tyrosine phosphatase mitochondrial 1-like isoform 1
34	1633608	529	Tyrosine-protein phosphatase non-receptor type 20-like
35	1667712	779	Protein-tyrosine-phosphatase MKP1-like

36	1672248	1903	MAP3k delta-1 protein kinase
37	1675796	1152	MAP3k delta-1 protein kinase
38	1669083	1601	MAP kinase kinase kinase mkh1-like
39	1669713	1409	MAPepsilon protein kinase
40	1805812	955	Big MAP kinase/bmk
41	1806653	1055	Mimosinase
42	1663223	1843	Mimosinase
43	1667384	1335	Mimosinase
44	1662801	1034	Mimosinase
45	1666921	1019	Mimosinase

Transcriptome sequencing of leucaena opens new opportunities of research on leucaena in the future by providing precious basic sequence data of transcriptionally active genes. Four main kinds of experiments may be designed to study leucaena genes in the future:

(i) Microarray analysis may be conducted using gene sequences obtained from the transcriptome analysis to identify genes that are expressed under different environmental conditions. For such microarray experiments, mRNA may be isolated from leucaena plants grown under certain environmental conditions. For example, to identify genes induced under drought-stress condition, mRNA must be isolated from plants grown under normal and drought conditions. Again, genes expressed in different parts of the plants can be identified by isolating mRNA from different parts of the plant, such as foliage, root, stem, flower, pods etc., and used as probes in microarray analysis.

(ii) Although the microarray analysis method is suitable for identification of genes that are expressed under certain environmental conditions or in certain parts of the plant, this method is still not enough to determine the level expression of the genes under certain environmental stresses; therefore, quantitative PCR (q-PCR) analysis is needed. By using q-PCR, the expression level of genes of leucaena plants grown under various environmental conditions can be determined.

(iii) For further analysis of the genes, listed in Table 6.1, it will be essential to obtain the full-length sequences for the incomplete gene sequences obtained through transcriptome

analysis. The full-length cDNA may be obtained by using 5' and 3'-RACE for the targeted genes.

(iv) Since leucaena has no known diseases, the testing NB-LRR genes may have to be done by heterologous gene expression in other plants such as *Arabidopsis*. The completed ORFs (open reading frames) of NB-LRR can be cloned in binary vectors and transfered to *Arabidopsis* to obtain transgenic plants. These transgenic *Arabidopsis* plants expressing leucaena NB-LRR genes may show resistance to fungal and bacterial pathogens that normally infect *Arabidopsis*.

In spite of having many good attributes, the miracle tree has a problem; it produces a non-protein amino acid mimosine, which is toxic to animals. Mimosine production by leucaena is perhaps a way of protecting itself from heavy infestation by insects and mammals. Mimosine is toxic to bacterial, animal and plant cells because it chelates iron and pyridoxal phosphate (Yeung et al. 2002, Negi et al. 2013). Iron is a co-factor of many enzymes in the cell, and similarly, pyridoxal phosphate is also prosthetic group required by some enzymes. In the presence of mimosine, iron and pyrirdoxal phosphate are not available for enzymes in the cell; therefore the cell cannot function normally. Mimosine also inhibits the thyroid hormone thyroxine; thereby causes goiter in animals (Jone 1979). Too much eating leucaena also causes hair loss of animals. Therefore, it is essential to develop mimosine-free leucaena varieties, which can be used as fodder for cattle without any harmful side effects. Unfortunately, there are no mimosine-free leucaena germplasms for developing new mimosine-free varieties of leucaena through conventional plant breeding. To develop mimosine-free leucaena through gene interruption using transgenic methods, knowledge of the biosynthesis pathway for mimosine will be necessary. However, the biosynthetic pathway for mimosine has not yet been discovered. It is expected that one or more genes identified in our transcriptome analysis may be involved in mimosine biosynthesis. At the present, we have no clue on how to identify such as genes from transcriptome sequences. Since mimosine biosysthesis pathways are not known, out lab has used a metabolic engineering approach for developing transgenic leucaena with reduced mimosine content. In this approach, genes for mimosine degradation was isolated from *Rhizobium sp*. TAL1145 that nodulates leucaena, and were transferred to leucaena through *Agrobacterium*-mediated transformation. Two *Rhizobium* genes *pydA* and *pydB*, encoding dioxygenase and hydrolase, were selected for this purpose (Awaya et al. 2005). The target of

these two enzymes is 3H4P, which is the precursor compound for mimosine biosynthesis. It was expected that due to the degradation of the precursor, mimosine will not be synthesized in the transgenic plants. For high-level expression of *pydA* and *pydB* in plants, their sequences were optimized for expression in plants, and a *pydA*/*pydB* fusion gene was constructed (Awaya et al. 2007, Jube et al. 2009). The *pydA*/*pydB* fusion gene was inserted to a binary vector and used to transform to leucaena through *Agrobacterium*-mediated transformation following the protocol developed by Jube and Borthakur (2009, 2010). This approach produced 16 transgenic leucaena plants, which were characterized for mimosine content in this study. Three transgenic plants (number 1,3,5) were identified, which contained reduced amounts of mimosine. These three plants may be useful as fodder for animals in the future.

Currently, we have only single plant for each of the newly developed transgenic leucaena plants. It is important to multiply these plants as soon as possible so that they can be used for further testing through replicated experiments. In this research, I have developed a tissue culture regeneration method for rapid micropropagation of leucaena without having to wait for these plants to flower and develop seeds.

References

Awaya J, Fox P, and Borthakur D (2005). *pyd* genes of *Rhizobium sp.* strain TAL1145 are required for degradation of 3-hydroxy-4-pyridone, an aromatic intermediate in mimosine metabolism. J. Bacteriol. 187(13):4480-4487

Awaya J, Walton C, and Borthakur D (2007). The *pydA-pydB* gene produces an active dioxygenase-hydrolase that degrades 3-hydroxy-4-pyridone, an intermediate of mimosine metabolism. Appl. Microbiol. Biotechnol, 75(3):583-588

Brewbaker J, and Sorenson C (1994). Domestication of lesser-known species of the genus Leucaena. "Tropical Trees: The potential for Domestication and Rebuilding of Forest Resources"; Institute of Terrestrial Ecology Symposium, 29:195-204

Jones R (1979). The value of *leucaena leucocephala* as a feed for ruminants in the tropics. World Animal Review, 31:13-23

Jube S, and Borthakur D (2009). Development of an *Agrobacterium*-mediated transformation protocol for the recalcitrant tree-legume *Leucaena leucocephala* using immature zygotic embryos. Plant Cell, Tissue and Organ Culture (PCTOC): Journal of Plant Biotechnology, 96: 325-333

Jube S, and Borthakur D (2010). Transgenic *Leucaena leucocephala* expressing the *Rhizobium* gene *pydA* encoding a meta-cleavage dioxygenase shows reduced mimosine content. Plant Physiol Biochem, 48:273-278

Jube S, Awaya J, and Borthakur D (2009). Expression of *Rhizobium pydA-pydB* fusion gene in *Nicotiana tabacum* confers resistance to the toxic aromatic compound 3-hydroxy-4-pyridone. Biologia Plantarum, 53(2):355-359

Kamada Y, Oshiro N, Oku H, Hongo F, and Chinen I (1996). Degradation of toxic mimosine by a hydrolytic enzyme of *leucaena psyllids* (jumping plant lice). Bioscience, Biotechnology, and Biochemistry, 60:351-352

Negi V, Bingham J.P, Li Q, and Borthakur D (2013). *mid*D-encoded 'rhizomimosinase' from *Rhizobium sp. strain* TAL1145 is a C–N lyase that catabolizes L-mimosine into 3-hydroxy-4-pyridone, pyruvate and ammonia. Amino Acids, 44:1537–1547

Report of study (1977). Leucaena promising forage and tree crop for the tropics. National Academy of Sciences Washington D.C.

Shelton H, and Brewbaker J (1994). *Leucaena leucocephala* – the most widely used forage tree legume. CAB international,15-29

Tawata S, Fukuta M, Xuan T, and Deba F (2008). Total utilization of tropical plants *Leucaena leucocephala* and *Alpinia zerumbet*. J. Pestic. Sci., 33(1):40-43

Yeung P, Wong F, and Wong J (2002). Mimosine, the *Allelochemical* from the leguminous tree *Leucaena leucocephala*, selectively enhances cell proliferation in dinoflagellates. Applied and environmental microbiology, 68(10):5160–5163

LIST OF CONFERENCE PROCEEDINGS

Correia J*, **Pham D**, Honda M, Chaille L, and Borthakur D (Poster presentation). Optimized *Agrobacterium*-mediated transformation of Leucaena leucocephala. The 22nd North American Symbiotic Nitrogent Fixation Conference, University of Minnesota, MN, USA (2013).

Correia J*, **Pham D**, Honda M, Chaille L, and Borthakur D. (Oral presentation). *Agrobacterium*-Mediated Transformation of *Leucaena leucocephala*. 25th Annual CTAHR & COE Student Research Symposium, University of Hawaii at Manoa, Honolulu, HI, USA. (2013)

Pham D*, Porter B, and Borthakur D (Poster presentation). Transcriptome analysis and identification of stress resistance genes in *Leucaena leucocephala*. The 22nd North American Symbiotic Nitrogent Fixation Conference, University of Minnesota, MN, USA (2013).

Pham D* and Botharkur D (Poster presentation). Transcriptome analysis and identification of stress resistance genes in *Leucaena leucocephala*. 25th Annual CTAHR & COE Student Research Symposium, University of Hawaii at Manoa, Honolulu, HI, USA. (2013)