Transcriptomics: Open Access

Transcriptome Analysis of *Leucaena leucocephala* and Identification of Highly Expressed Genes in Roots and Shoots

Kazue L Ishihara, Michael DH Honda, Dung T Pham and Dulal Borthakur*

Department of Molecular Biosciences and Bioengineering, University of Hawaii at Manoa, Hawaii, USA

Abstract

Leucaena leucocephala (leucaena) is a fast-growing tree legume highly tolerant to various abiotic and biotic stresses. Because of its abilities to withstand high temperature and prolonged drought and to grow as a disease-free plant, it is an interesting model plant to investigate genetics of stress resistance. The high-level stress resistance may be correlated with higher expression of certain genes in the root, which is the primary site for nutrient and water uptake and also infection by soil-borne pathogens. The objectives of this study were to characterize the transcriptome of leucaena and to identify root-specific genes that may be involved in drought tolerance and disease resistance. Transcriptomes of leucaena were analyzed through Illumina-based sequencing and *de novo* assembly, which generated 62,299 and 61,591 unigenes (≥500 bp) from the root and shoot, respectively. Through a 4 x 180k microarray analysis, the expression of 10,435 unigenes were compared between the root and shoot. Upregulated sequences in the root were mostly represented by unigenes that were related to secondary metabolism, while in the shoot, upregulated sequences were mostly represented by unigenes that were involved in carbohydrate and lipid metabolism. The unigenes sharing homology with terpenoid biosynthesis genes and a nicotianamine synthase gene were upregulated more than 100-fold in the root, which indicates that these genes may have important roles in high stress tolerance of leucaena. Cataloging of actively transcribed sequences in the root and shoot will lead to identification of genes for drought tolerance and disease resistance in leucaena.

Introduction

Leucaena leucocephala (leucaena) is a fast-growing tree legume widely grown in the tropics and subtropics. It can be grown under unirrigated conditions in areas with relatively warm climates, including Australia, Southern India, Africa, Central and South America, Philippines, and Taiwan [1]. It occupies two million ha in the Pacific Basin and Rim countries alone [2]. Because of its widespread success as a multipurpose tree suitable for agroforestry, it is also known as a "miracle tree" [3]. It is highly tolerant to various abiotic and biotic stresses, including drought and diseases, and it has high adaptability to various ecological conditions. It can grow successfully even in relatively less fertile soils because of its ability to fix nitrogen in symbiosis with Rhizobium tropici [4]. As a nitrogen-fixing, deep-rooted tree, it is often grown for green manure, shade, firewood, windbreak, and controlling erosion [2].

It is the most widely used forage legume because it has an unusually high protein content of ~18% in its foliage, providing important source of proteins for farm animals [5]. Because of its high protein content, it is often known as the "alfalfa of the tropics." Compared to alfalfa, which is a common leguminous fodder suitable for temperate regions, leucaena, as a drought-resistant legume, is more suitable even as an unirrigated crop, for tropical regions. Leucaena is a better alternative in drought-affected areas where alfalfa does not grow well. The newly developed Hawaii-bred high-yielding leucaena varieties produce an annual fodder dry weight of ~30 tons/ha. The Hawaii-bred leucaena varieties are also very successful as a fodder legume in Australia [3]. Its young leaves, pods, and seeds are used not only as a fodder, but also as vegetables for human consumption in Central America, Indonesia, and Thailand [6-8]. Due to its fast-growing nature and high-biomass productivity, leucaena's potential as raw material for pulp and paper industries has gained attention as well [9,10]. Leucaena is also useful for soil remediation; recent studies show that it can effectively remove textile dyes or heavy metal contaminations from soils [11,12].

Recently, leucaena has gained more attention as a drought- and disease-resistant forest legume [3]. Because of the current trend of global warming, the drought tolerance nature of leucaena especially has gained more importance from the plant biologists. Leucaena may provide a good source of genes conferring tolerance to various abiotic and biotic conditions. However, its genome has not been sequenced, and relatively few genes have been characterized. It is also a challenge to sequence its complex allotetraploid genome with 104 chromosomes [13]. Transcriptome sequencing can provide a better alternative for identifying genes for both drought tolerance and disease resistance from leucaena.

In the present study, Illumina de novo sequencing technology was utilized to characterize the transcriptomes of the root and shoot of leucaena, and a microarray analysis was performed to find differentially expressed genes in the two tissues. The objectives were to enrich the gene resource of leucaena with the sequencing data and to identify root- and shoot- specific genes. We especially focused on genes highly expressed in the root because it is where the plant uptakes nutrient and water to survive and also where certain abiotic and biotic stresses, such as drought and soil-borne pathogens, affect the plant first. Therefore, root-specific genes may hold a key to understanding the genetic and biochemical basis of high adaptability and stress tolerance of leucaena. To the best of our knowledge, this is the first exploration to characterize the transcriptome of leucaena. The transcriptome sequencing and expression analyses in the root and shoot of leucaena will offer valuable resources and contribute to future research to identify unique genes in this "miracle tree."

*Corresponding author: Dulal Borthakur, Department of Molecular Biosciences and Bioengineering, University of Hawaii at Manoa, Hawaii, USA, Tel: (808) 956-6600; E-mail: dulal@hawaii.edu

Materials and Methods

Plant Materials and RNA Extraction

Seeds of *L. leucocephala* var. K-636 were collected from the Waimanalo Research Station, University of Hawaii, Waimanalo, Hawaii. For scarification, they were immersed in concentrated sulfuric acid for 10 min, rinsed with water 5 times, and incubated in petri dishes with wet filter paper at 28ºC until they germinated (3-5 days). The germinated seedlings were then planted into a tray containing a vermiculite-perlite mixture and maintained at 25 °C ± 2 °C with a 16/8-h light/dark photoperiod with an irradiance of 30 μmol s^{-1} m^{-1} with Hoagland solution. After a month, when the plants were 12 to 15 cm tall, roots and shoots were harvested and immediately placed in liquid nitrogen prior to RNA extraction.

To extract RNA, roots and shoots of leucaena were separately ground to a fine powder in liquid nitrogen with a mortar and a pestle that were baked for 6 h at 300 °C prior to use. The modified method using Qiagen RNeasy Plant Kit (Valencia, CA, USA) and Fruit-mate (Takara, Japan) was used to extract RNA from shoots as described by Ishihara et al. [14]. The same method was performed to extract root RNA, except Buffer RLC was used instead of Buffer RLT. The quantity and quality of the RNA were assessed at wavelengths of 230, 260, and 280 nm using a NanoDrop Spectrophotometer ND-1000 (NanoDrop Technologies, DE, USA). To confirm the quality, the RNA was analyzed based on RNA Integrity Numbers (RIN) obtained through an Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA).

Sequencing, and Assembly, and Functional Annotations

SeqWright Genomic Service, Houston, TX conducted cDNA library construction, sequencing, and assembly. Briefly, cDNA was synthesized from poly(A)-selected RNA and non-stranded RNA libraries and was sequenced by Illumina HiSeq 2000 with 100 bp paired-end reads. SOAPdenovo was used to assemble sequences obtained from Illumina [15]. The resulting assembled sequences were defined as unigenes.

The assembled unigene sequences (≥500 bp) were selected as reference transcriptomes, and they were compared against three protein databases, including the NCBI non-redundant (nr) database, the Translated European Molecular Biology Laboratory (TrEMBL) database, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, through the Basic Local Alignment Search Tool (BLAST) algorithm with a cut-off E-value of 1E-4 using the doblast server of the Noble Foundation (http://bioinfo3.noble.org/doblast/) and the WebMGA server (http://weizhong-lab.ucsd.edu/metagenomic-analysis/). Gene names were assigned to each query based on the

highest sequence similarity. A Java program Blast2Go [16] was utilized to assign Gene Ontology (GO) functional categories for the annotated unigenes.

Identification of Differentially Expressed Sequences Through Microarray

For 4 x 180k microarray analysis, 10,435 cDNA sequences (≥500 bp) were randomly selected from the root and shoot transcriptomes. For each sequence, minimum of five 60-bp probes were designed through Agilent Technologies (Santa Clara, CA, USA). Each probe had three to four replicates in a DNA chip. The designs of probes and microarray were deposited into GEO (Record No. GSE76810). The microarray analysis was performed by the Roy J. Carver Center for Genomics, University of Iowa. The array was scanned using a NimbleGen MS 200 Microarray Scanner (Roche NimbleGen, Inc.). The NimbleScan software v.2.6 (Roche NimbleGen, Inc.) extracted raw intensities from the images generated by the scanner, which were corrected for background noise and normalized between arrays using a Robust Multichip Average (RMA) algorithm included in the NimbleScan software. The normalized probe intensity values were averaged to give a single intensity value per transcript and per sample.

Experimental Validation Through qRT-PCR

Total RNA extracted from the root and shoot was treated with TURBO DNAfree Kit (Invitrogen, Carlsbad, CA) to remove any genomic DNA contamination. First-strand cDNA was synthesized from 2 μg of DNase-treated RNA using M-MLV Reverse Transcriptase (Promega, WI, USA) with random hexamers according to the manufacturer's instructions. The quantitative real-time (qRT)-PCR analysis was performed to confirm differential expression of 11 unigenes, using a 10 μL PCR reaction consisting of 0.25 μL forward primer (10 μM), 0.25 μL reverse primer (10 μM), 5 μL PowerUpTM SYBR® Green Master Mix (Applied Biosystems, Foster City, CA), and 1 μL of first strand cDNA. Reaction conditions were 50°C for 2 min, 95°C for 3 min, 40 cycles of 95°C for 15 s, 55°C for 30s, and 72°C for 30s, followed by melting curve analysis of the amplicon to confirm the specificities of primers. Each assay consisted of three biological and three technical replicates and was performed using StepOne Real-Time PCR System (Applied Biosystems). The PCR protocol produced a PCR efficiency of 90% to 110% for each primer set. The primer sequences used for this study are listed in Table 1.

Selection of Internal Reference Gene for qRT-PCR analysis

To select internal reference genes for relative quantification of target gene expressions, six reference candidate genes: ef1α, β-actin, 18S rRNA, 5.8S rRNA, ubiquitin-5, and tubulin-1. were tested on

Table 1: Primer sequences of target unigenes used in the qRT-PCR analysis for the confirmation of the microarray expression results.

Accession No.	Putative function	Forward Primer (5 --> 3)	Reverse Primer (5 --> 3)
GDSA01221554	Apyrase	CTAACATGCACACTCTTGGAGCTTT	TGCGGTTCTAGGTACTTCGTTCTC
GDSA01195532	Senescence-related protein 1	TAAGGCGATGACTCTAGGGTTAGG	GTCTTCTTCAGATCTCGCATGCTC
GDRZ01208625	Chalcone synthase	CCTGCGATTCTTGACCAAGTTGAG	GCAAGCACTCGACATATTCCCATAC
GDRZ01209121	Sombrero-like	CCACAGTTGTTCAGCCCAGAAT	AGTAGGTTGTGAGAACAGTGAAGGA
GDRZ01209218	MKS1-like	CCTTCTGATTTCATGAGCCTCGTTC	CCCACCATCGCTATCATTAGAGACT
GDRZ01209285	Terpene synthase	GATCTCCATTCACTAGCTTTGGTCTTC	CTCGCTGCTGAAGTTTCCTTTCT
GDRZ01209307	Peroxidase 21-like	GTCCGAGCATTGAGATGAAGACAG	CACGAGGGAAATGGAGTCGTTTATG
GDRZ01212539	Neomenthol	AGTTTGCTTTAATGTGGGAGGATCG	AGTGGAGACAAACACCAGCTTTG
GDRZ01213129	Nicotianamine synthase	ACCACTTGGTTTCCACACACTTT	ATCCTCTTCTCGAGCTCTGAATCC
GDSA01140474	Isoliquiritigenin 2-O'-methyltransferase	ACCTTGATTGGGTTGAAACCTTGA	GGCTACAAATACCCAAACCTTCTCC
GDSA01146543	Cysteine proteinase	CGGAGGAACGCGTAATGGAAA	ACTGTGTACTATCTCCCTTGGTCAG

Table 2: Summarized assembly statistics for unigenes in leucaena.

Statics	Number	
	Root	Shoot
Total number of paired-end reads	111,417,073	104,137,306
Total number of assembled unigenes	62,299	61,591
Total length of unigenes (bp)	39,742,487	40,318,375
Mean length of unigenes (bp)	805.9	809.0
Median length of unigenes (bp)	684	686
Max length of unigenes (bp)	11,152	8,094
N50 length of unigenes (bp)	790	791

Table 3: Length distribution of *de novo* assembled unigenes and annotation frequencies.

Length (bp)	Root		Shoot	
	Number of unigenes	Frequency (%)	Number of unigenes	Frequency (%)
500-599	21,292	73.7	20,817	75.4
600-699	13,264	77.4	13,061	79.8
700-799	8,463	80.5	8,451	82.5
800-899	5,548	82.5	5,512	83.9
900-999	3,682	84.7	3,660	85.6
1,000-1,999	9,107	87.5	9,104	88.5
2,000-2,999	752	91.2	752	93.1
3,000+	191	97.9	234	96.2
Total	62,299	79.16	61,591	80.9

Table 4: Summary for the annotation of unigenes of leucaena (cutoff <1.0E-4).

Database	Number of annotated unigenes	
	Root	Shoot
nr	49,224	49,766
TrEMBL	49,056	49,629
KEGG	14,040	14,573
Total annotated	49,314	49,840

first strand cDNA samples from the root and shoot with the primer sequences described in Negi et al. [17]. The qRT-PCR protocol was performed as described above. The cycle threshold (Ct) values of the candidate genes were used to evaluate their expression stability by using NormFinder applet for MS Excel [18]. NormFinder allows the user to determine intra- and inter-group variances as well as the stability value of each candidate gene. Using a reference gene, the fold change of target gene expression levels comparing the root and shoot was determined using the $2^{-\Delta\Delta Ct}$ method [19]. Statistical significance was determined using Student's one-tailed t-test with significant differences for $p < 0.05$.

Results

Sequence Analysis and Assembly

From the Illumina HiSeq 2000 sequencing, 111,417,073 paired-reads with a total length of 22.5 Gb were generated from the root, and 104,137,306 paired-reads were generated with a total length of 21Gb from the shoot. The raw reads were deposited to the NCBI Sequence Read Archive (SRA) with accession numbers SRR2517689 for the root and SRR2517688 for the shoot. When the raw reads were assembled through SOAP *de novo*, 62,299 unigenes (≥500 bp) were generated with a total length of 39.7 Mb with an average length of 805.9 bp and N50 of 790 bp for the root. For the shoot, 61,591 unigenes (≥500 bp) were generated with a total length of 40.3 Mb, an average length of 686 bp, and N50 of 791 bp (Table 2). These sequences were deposited to the NCBI Trancriptome Shotgun Assembly (TSA) database, and accession numbers GDRZ00000000 and GDSA00000000 were obtained for the root and shoot transcriptomes, respectively.

The length distributions of the assembled unigenes were very similar for both root and shoot transcriptomes. Of the root transcriptome sequences, 52,249 sequences (83.4%) were 500-999 bp; 9,107 sequences (14.6%) were 1,000-1,999 bp, and 943 sequences (1.5%) were ≥2,000 bp (Table 3). The longest sequence in the root transcriptome (Acc. No. GDRZ01240663) was 11,152 bp, which had high identity (91%) with a predicted protein of transformation/transcription domain-associated protein from G. max (Acc. No. XP_006590726.1). Similar to the root transcriptome, the shoot transcriptome data had 51,501 sequences (83.6%) with the length 500-999 bp, 9,104 sequences (14.8%) with the length between 1,000-1,999 bp, and 986 sequences (1.6%) with the length ≥2,000 bp. The longest sequence from the shoot transcriptome had a length of 8,094 bp (Acc. No. GDSA01234712), which had 70% identity to the predicted protein of small subunit processome 20 from Glycine max (Acc. No. XP_006601933.1). The longest sequences from both root and shoot transcriptomes had 99% coverage with the known protein sequences (data not shown).

Functional Annotations of Assembled Sequences

All the assembled unigenes were searched against several protein databases, including the nr database, the TrEMBL database, and the KEGG database using the BLAST algorithm (E-value <1E-4). Almost equal proportions of unigenes in the root and shoot showed homology with sequences in the databases: a total of approximately 49,000 unigenes (~79%) were annotated with the three databases in each of the root and shoot transcriptomes (Table 4). There was a higher annotation frequency for the unigenes with greater lengths (Table 3). In both root and shoot transcriptomes, over 93% of unigenes with the length of ≥2,000 bp showed homologous matches to protein sequences in the searched databases, whereas the annotation rates were ~88% for unigenes between 1,000-1,999 bp and only ~77% for unigenes between 500-999 bp. The majority (over 43%) of the unigenes without annotations from the databases were 500-599 bp. The reason for this was most likely their short sequence lengths, resulting statistically insignificant matches.

GO Classification

Among the unigenes annotated by TrEMBL, 27,501 and 27,780 were assigned with one or more GO terms for the root and shoot transcriptome data, respectively, and classified into three GO functional categories: "biological process," "cellular component," and "molecular function" (Figure 1). The distributions were similar for both root and shoot. In the "biological process" category, the unigenes were further clustered into 20 subcategories. Of those, "metabolic process" was the most represented (~12,500 unigenes); the second was "cellular process" (~11,000 unigenes), and the third was "single-organism process" (~7,700 unigenes). Under the "cellular component" category, the unigenes were assigned to 16 subcategories; the most abundant classes were "cell" (~7,100 unigenes), "membrane" (~6,800 unigenes), and "organelle" (~4,600 unigenes). The unigenes under the molecular function category were sorted into 6 subcategories. Highly represented group was "binding activity" (~10,400 unigenes), "catalytic activity" (~9,700 unigenes), and "transporter activity" (~1,300 unigenes).

KEGG Pathway Classification

The KEGG database provides systemic functional information of biochemical pathways and functions of gene products in addition to annotations of sequences. From the KEGG-annotated unigenes, 5,081 and 5,275 sequences from the root and shoot transcriptomes, respectively, were grouped into KEGG biochemical pathways (Figure 2). Both transcriptomes had almost the same distributions
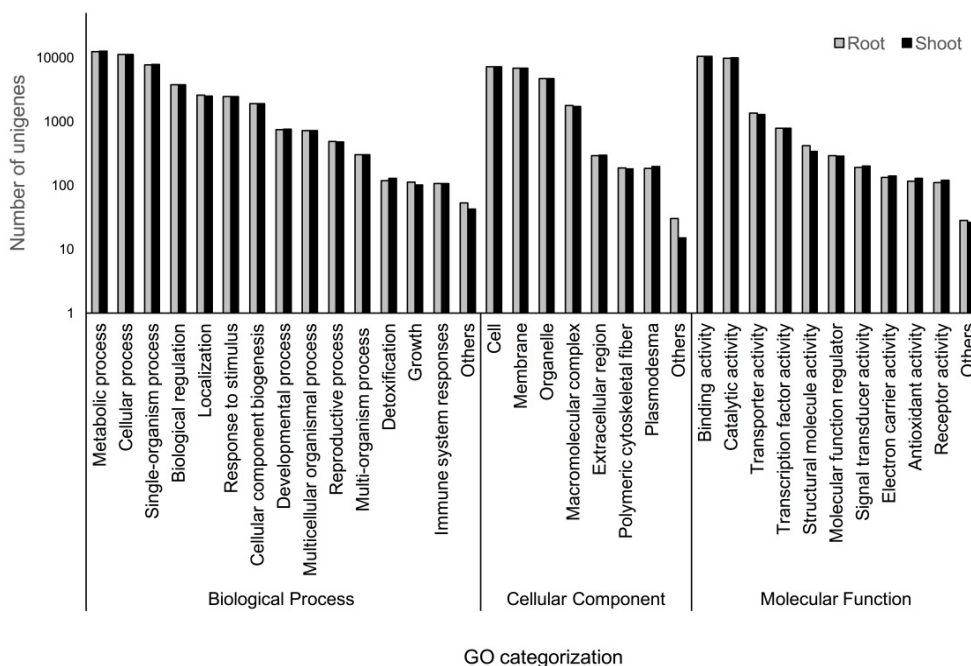
Figure 1: Gene Ontology (GO) functional categorization of the unigenes from the leucaena roots and shoot tissues (≥500 bp). The 27,501 and 27,780 unigenes annotated by TrEMBL were assigned to one or more GO terms for the root and shoot transcriptome data, respectively. There are three GO functional categories: "biological process," "cellular component," and "molecular function," which were further divided into more specific functional groups.
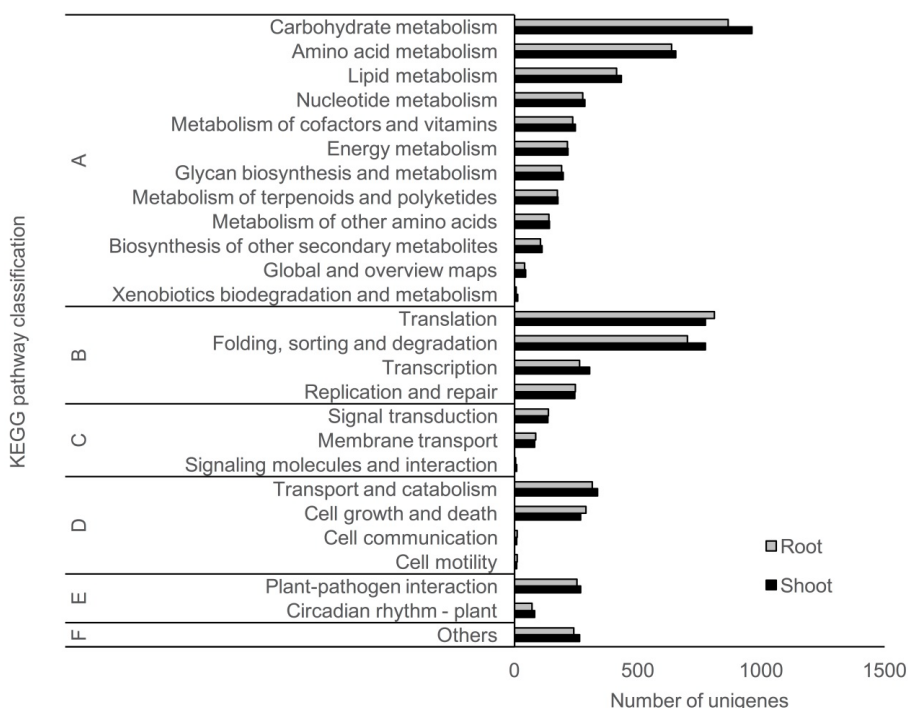

Figure 2: KEGG pathway classification of unigenes from the leucaena root and shoot tissues. A total of 5,081 and 5,275 unigenes from the root and shoot transcriptomes, respectively, were categorized into six major KEGG biochemical pathways: A-metabolism, B-genetic information processing, C-environmental information processing, D-cellular processes, E-organismal systems, F-others.

in major categories. Major KEGG biochemical pathway groups were "metabolism" (~3,300 unigenes), "genetic information processing" (~2,000 unigenes), "cellular processes" (~600 unigenes), "organismal system" (~350 unigenes), and "environmental information and processing" (~230 unigenes). Although overall numbers of the major

pathway groups were similar, the subcategories within the "metabolism" group varied. The shoot transcriptome, for example, consisted of ~10% higher number of unigenes for "carbohydrate metabolism," which includes "glycolysis/gluconeogenesis," "citrate cycle," "starch and sucrose metabolism," and "pyruvate metabolism" compared to the root
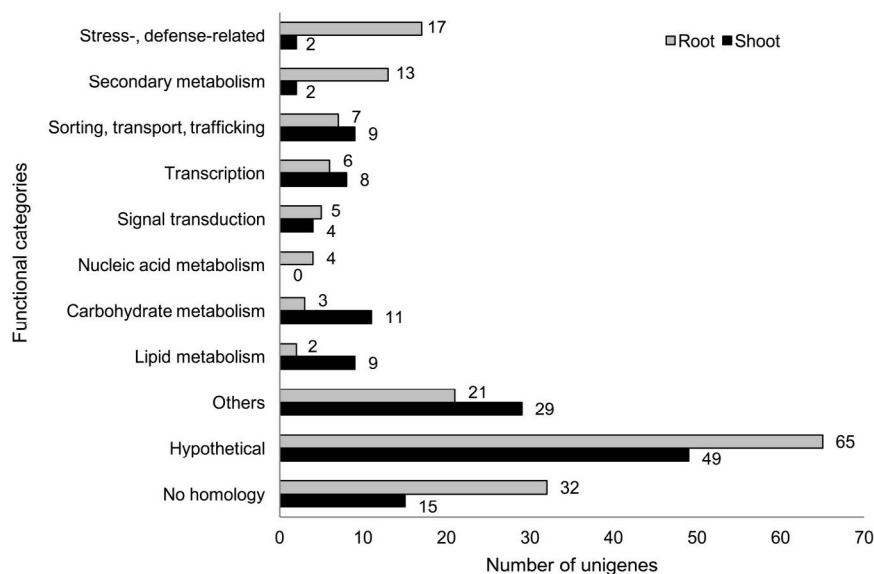
Figure 3: Functional annotation of unigenes upregulated in the root and shoot of leucaena (≥5-fold), showing the number of unigenes in each category. The microarray analysis showed that 175 and 138 unigenes were upregulated in the root and shoot, respectively.

transcriptome. On the hand, the root transcriptome has ~10% more unigenes grouped into subcategories related to terpenoid metabolism, such as "ubiquinone and other terpenoid-quinone biosynthesis," "terpenoid backbone biosynthesis," and "diterpenoid biosynthesis" (Table S1).

Microarray and qRT-PCR Analyses

To confirm that the assembled unigenes were indeed expressed in *L. leucocephala*, 10,435 unigenes were randomly selected from the root and shoot transcriptome data for a microarray analysis. The microarray data was deposited into GEO (Record No. GSE76810). In this analysis, expression levels between the root and the shoot were compared in order to identify tissue specific genes. The sequences that showed at least five-fold differences were analyzed and categorized into functional groups (Figure 3). There were 175 unigenes with higher expression levels in the root than in the shoot, of which 79 unigenes showed over 10-fold increase in expression levels (Tables 5 and S2), while 138 unigenes were upregulated in the shoot with 65 of them upregulated more than 10-fold (Tables 6 and S3). Categorization of those unigenes indicated that a greater number of genes related to stress (17 unigenes) and secondary metabolism (13 unigeues) were upregulated in the root, compared to the shoot (2 unigenes in each category of stress-related and secondary metabolism). On the other hand, there were more unigenes categorized in carbohydrate and lipid metabolisms in the shoot (11 and 9 unigenes, respectively) than in the root (3 and 2 unigenes, respectively; Figure 3). Unigenes that had no homology in the known protein database also showed differential expression; 32 unigenes with no homology were upregulated in the root with the highest fold change of 446.4 (Table S4), while 15 were upregulated in the shoot (Tables S5).

For qRT-PCR analysis, an internal control was selected based on the stabilities of the candidate genes, evaluated by NormFinder. The ef1α had the lowest stability value with a small intra- and inter-group variations, compared to the other five candidates and the "best combination of the genes" (ef1α and ubiquitin-5), so it was selected as the internal control (Figure 4). Through the expression analysis, a unigene that had homology to a nicotianamine synthase was confirmed to have significant upregulation of over 600-fold in the root compared

to the shoot (p <0.001; Figure 5). The qRT-PCR analysis also confirmed significant upregulation of six unigenes that shared homology with genes that may be involved in secondary metabolite biosynthesis. Those sequences were homologous to neomenthol dehydrogenase (171.6-fold, p <0.01), sesquiterpene synthase (131.0-fold, p <0.01), isoliquiritigenin 2-O'-methyltransferase (6-fold, p <0.01), peroxidase 21-like (8.4-fold, p <0.01), senescence-related gene 1 (4.4-fold, p <0.05), and chalcone synthase (2.1-fold, p <0.05). The expression of four other unigenes, including those sharing homology with SOMBRERO-like (48.9-fold, p <0.05), MSK1-like (9.2-fold, p <0.01), cysteine proteinase (7.8-fold, p <0.05), and apyrase (5.1-fold, p <0.01), was also confirmed through qRT-PCR.

Discussion

Transcrioptome Sequencing and Assembly

The next-generation sequencing (NGS) provides a rapid, cost-effective, and labor-saving approach to construct and characterize transcriptomes of organisms including non-model species without known genomic sequences [20-22]. Leucaena is such a non-model legume species, whose genome has not been sequenced. As of now, in legumes, the genome sequences have been available only for few species, including *Glycine max*, *Lotus japonicus*, *Medicago truncatula*, *Cajanus cajan* (pigeonpea) and *Cicer arietinum* (chickpea) [23-27]. Genetic information of forage tree legumes, such as leucaena, is still limited in public databases. Therefore, in this study, we sequenced and *de novo* assembled the transcriptomes of the root and shoot of leucaena. With its high tolerance to drought and diseases, leucaena may provide a rich source of genes for agroforestry improvement programs. Through this study, a total of over 60,000 unigenes were identified from each transcriptome of the root and shoot, and approximately 80% of them were annotated with three protein databases. To the best of our knowledge, leucaena is the first forage tree legume that has been thus far characterized through transcriptome analysis. These sequences will be useful as a reference database of mRNA, which will facilitate future genetic studies of leucaena.

Because the published transcriptome sequences of various plant

Table 5: Annotated unigenes upregulated in the root compared to the shoot of leucaena (≥10-fold).

Acc. No.	Ratio	SE	Putative function	Blast Hit Acc. No.
GDRZ01213129	1011.8	54.4	Nicotianamine synthase	XP_006598174.1
GDSA01213076	155.5	22.4	Cytosolic sulfotransferase	XP_002267209.1
GDRZ01212539	144.9	14.7	(+)-neomenthol dehydrogenase-like	XP_006588341.1
GDRZ01209285	122.2	4.7	Sesquiterpene synthase	XP_003618357.1
GDRZ01208003	82.5	7.6	Cytochrome P450	XP_003617731.1
GDRZ01209506	67.1	3.8	Leucine-rich repeat receptor-like	XP_003589757.1
GDRZ01106044	41.7	65.9	Tryptophan aminotransferase-like	AFG31374.1
GDSA01186189	35.9	1.4	TMV resistance protein N	XP_003614280.1
GDSA01215187	35.1	12.1	Patatin-11-like	XP_006601293.1
GDRZ01209121	30.3	8.7	SOMBRERO-like	XP_004488594.1
GDRZ01106763	25.7	4.4	Transcription factor	XP_006572975.1
GDSA01193213	25.4	4.0	OMT3 protein	AGK93044.1
GDRZ01149910	24.2	2.8	Cysteine proteinase	AGV54734.1
GDSA01197972	21.7	6.0	Copalyl pyrophosphate synthase	AAB87091.1
GDSA01142011	19.1	4.6	Cytochrome P450	XP_003534174.2
GDSA01236328	18.8	0.5	Patellin-6-like	XP_004514256.1
GDRZ01211270	18.2	0.8	Cysteine-rich receptor-like protein kinase	XP_006605289.1
GDRZ01208625	16.1	2.1	Chalcone synthase	ACB78187.1
GDSA01140474	15.7	0.4	Isoliquiritigenin 2-O'-methyltransferase	XP_004496361.1
GDSA01155181	15.4	0.4	Laccase-6-like	XP_003529882.1
GDSA01200385	15.0	0.4	Peptidyl-prolyl cis-trans isomerase cyp5-like	XP_004164532.1
GDRZ01212404	14.4	0.0	Elongation factor 1 alpha, partial	ADK90073.1
GDSA01219052	14.0	2.6	Cytochrome P450 71A1-like	XP_004501856.1
GDRZ01208241	13.5	1.5	U3 small nucleolar RNA-associated protein 13	XP_004504147.1
GDRZ01144141	13.4	4.5	LRR receptor-like Ser/Thr-protein kinase	XP_003535558.1
GDSA01104248	13.1	1.2	Rpp4 candidate	XP_003589916.1
GDRZ01209934	12.7	2.5	Fructose-bisphosphate aldolase	XP_004497605.1
GDSA01235763	11.3	0.3	F3'H	AEX07282.1
GDSA01150359	11.3	1.1	Major facilitator superfamily protein	XP_007011003.1
GDSA01210265	11.0	0.6	Laccase-14-like	XP_003537502.1
GDSA01209254	10.9	0.4	Glycine-rich protein	XP_003523459.1
GDRZ01107573	10.7	0.2	CM0216.320.nc	BAF98216.1
GDSA01233718	10.6	0.8	Protection of telomeres 1 protein	ACJ49159.1
GDSA01193636	10.2	0.6	Inactive purple acid phosphatase	NP_001276313.1
GDSA01151501	10.0	0.4	Isoliquiritigenin 2-O'-methyltransferase	XP_004496361.1

Table 6: Annotated unigenes upregulated in the shoot compared to the root of leucaena (≥10-fold).

Acc. No.	Ratio	SE	Putative function	Blast Hit Acc. No.
GDRZ01156473	130.6	7.3	40S ribosomal protein S3, putative	XP_002514061.1
GDSA01236003	121.1	15.2	WAT1-related protein	XP_004485529.1
GDSA01158382	97.2	2.8	Hypothetical protein	XP_007132014.1
GDSA01169979	92.3	12.9	GDSL esterase/lipase	XP_004485880.1
GDRZ01156568	90.9	16.1	Cytochrome P450	EXB50285.1
GDSA01105057	74.3	5.6	Chlorophyll a-b binding protein 16	EXB44448.1
GDSA01144134	64.9	4.0	Axial regulator YABBY	XP_003542627.1
GDSA01148410	55.8	6.9	Aldehyde dehydrogenase	XP_004486968.1
GDSA01182796	54.3	6.1	Beta-amyrin synthase	Q9MB42.1
GDSA01234114	53.7	3.7	UDP-glycosyltransferase	XP_003531850.1
GDSA01235589	51.8	2.1	GDSL esterase/lipase	XP_003546013.1
GDSA01175641	51.3	1.3	Pentatricopeptide repeat-containing protein	XP_006605878.1
GDSA01234291	46.8	1.6	UDP-glycosyltransferase	XP_002274748.2
GDSA01148623	43.4	1.5	GDSL esterase/lipase	XP_003549836.1
GDSA01167380	34.2	2.4	sGS-II	AAB82745.1
GDSA01161196	32.0	4.6	Polyol transporter	XP_006577924.1
GDSA01161960	30.8	1.2	Lipoxygenase	BAB84352.1
GDSA01173290	27.2	1.9	Protein HOTHEAD-like	XP_003549865.1
GDSA01180243	24.4	2.3	Homeobox protein 24, putative	XP_007046919.1
GDSA01210192	20.5	0.4	pbsQ-like protein	XP_004490109.1
GDSA01233524	20.4	0.8	Xyloglucan endotransglucosylase/hydrolase	XP_003517365.2
GDSA01236287	19.7	0.0	Long chain acyl-CoA synthetase	XP_003554561.1

| GDSA01235002	19.0	13.3	GMC-type oxidoreductase, putative	XP_003629014.1
GDSA01200937	16.7	2.7	Chloroplast import apparatus	XP_002268213.1
GDSA01217062	16.5	4.9	Protein HOTHEAD-like	XP_006607079.1
GDSA01170676	16.3	0.5	Vinorine synthase-like	XP_003538872.1
GDSA01147873	16.1	3.0	Abscisic acid 8'-hydroxylase 2 isoform	XP_003534678.1
GDSA01167031	15.3	1.0	Glycerol-3-phosphate acyltransferase	XP_007026769.1
GDSA01217884	15.1	1.5	BTB/POZ domain-containing protein	XP_006590635.1
GDSA01185590	14.7	0.8	Fructose-1,6-bisphosphatase	XP_004147899.1
GDSA01210839	14.6	1.7	Chitinase homologue	CAA88593.1
GDSA01235414	13.0	0.4	Mannan endo-1,4-beta-mannosidase 4-like	XP_003553696.1
GDSA01054472	12.4	1.1	ABC transporter G family	XP_003541427.1
GDRZ01157225	12.2	0.4	Neutral invertase isoform 2	XP_007035890.1
GDSA01195932	12.2	1.7	Adenylate isopentenyltransferase	ABD93934.1
GDSA01117917	11.5	0.1	Receptor-like	XP_004498297.1
GDSA01234520	11.5	0.4	GT-2 factor	XP_003592112.1
GDRZ01108853	10.7	0.3	Beta-glucosidase	XP_002512142.1
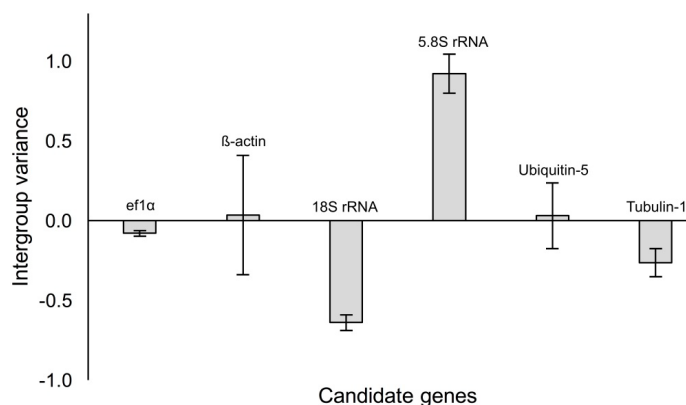
Figure 4: Intergroup variance values of candidate reference genes as determined by Normfinder. Error bars indicate average of intragroup variance values. The ef1α gene was selected for internal control of the qRT-PCR analysis based on its lowest inter- and intragroup variance values.
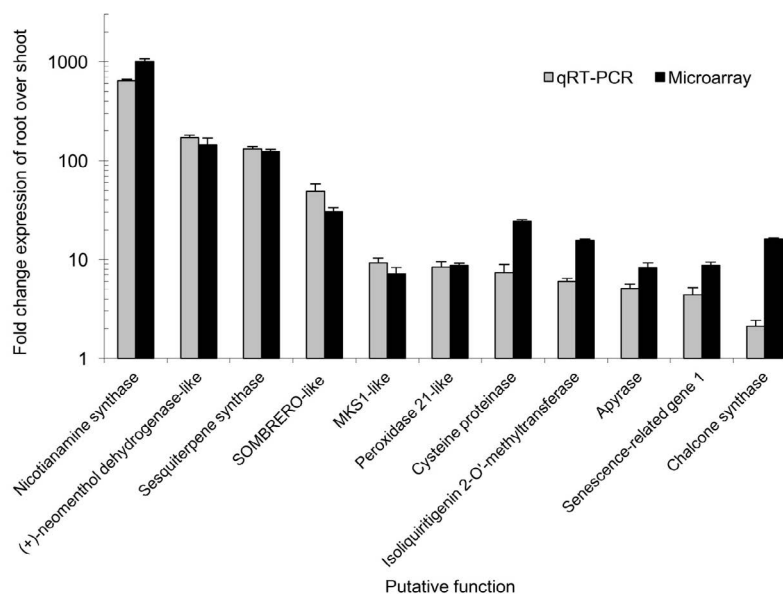

Figure 5: Validation of the microarray data by qRT-PCR, showing the correlation between the data obtained from the two methods. Fold expression change of the root compared to the shoot of leucaena obtained from microarray and qRT-PCR analyses were shown with p-values: nicotianamine synthase ($p<0.001$), neomenthol dehydrogenase ($p<0.01$), sesquiterpene synthase ($p<0.01$), SOMBRERO-like ($p<0.05$), MSK1-like ($p<0.01$), peroxidase 21-like ($p<0.01$), cysteine proteinase ($p<0.05$), isoliquiritigenin 2-O'-methyltransferase ($p<0.01$), apyrase ($p<0.01$), senescence-related gene 1 ($p<0.05$), and chalcone synthase ($p<0.05$). Error bars represent standard errors.

species were obtained through the use of several high-throughtput sequencing technologies and assembly software, it is difficult to evaluate the leucaena transcriptomes by comparing the number or average length of unigenes with other transcriptomes. Majority of unigenes in the leucaena transcriptomes were annotated, and their distribution pattern of the GO and KEGG pathway classifications is similar to that of the transcriptomes of other legumes species such as *Acacia koa* and chickpea [28,29]; therefore, the transcriptome sequences presented here appear to be a comprehensive representation of the entire transcriptome of leucaena.

Overview of the Gene Expression Analysis in the Root and Shoot

Microaray analysis of ~10,000 unigenes showed differential expression of some of the unigenes in the root and shoot. Functional categorization of differentially expressed sequences revealed that more genes related to stress and secondary metabolism were upregulated in the root, while more genes involved in carbohydrate and lipid metabolisms were upregulated in the shoot. This is consistent with the KEGG pathway classification, in which more unigenes from the root transcriptome were shown to be related to terpenoid metabolism, while more unigenes from the shoot transcriptome was categorized into carbohydrate metabolism. The upregulated genes related to carbohydrate and lipid metabolisms in shoots included those for energy metabolism, such as fructose-1,6-bisphosphatase (14.7-fold), which is involved in gluconeogenesis or the Calvin cycle, and long-chain acyl-CoA synthetase (19.6-fold), which converts free fatty acids to acyl-CoA, a key step for lipid metabolism. Acyl-CoA synthetases yield fatty acyl-CoA and facilitate β-oxidation, which feeds into gluconeogenesis, generating more sugar. In *Zea mays* (maize) and *Gossypium arboretum* L. (cotton), the amounts of soluble sugars like sucrose and hexose increased in response to drought stress [30-32], and they play an important role in osmotic adjustment [32,33]. Further studies are necessary to confirm the roles of these upregulated genes in leucaena shoot.

Upregulation of Genes Involved in Terpenoid Biosynthesis in the Root

Many secondary metabolites, such as phenylpropanoids and terpenoids, are known to have antimicrobial properties to protect plants from pathogens [34,35]. From the microarray and qRT-PCR analyses, six sequences that may be involved in the secondary metabolite biosynthesis were found to be upregulated in the root. Two of them were upregulated more than 100-fold in the root, and they shared homology with two terpenoid biosynthesis genes, neomenthol dehydrogenase and sesquiterpene synthase. Neomenthol dehydrogenase is involved in biosynthesis of monoterpene menthol in plants and is considered to provide basal resistance against pathogens; Choi et al. [36], for example, demonstrated that overexpression of the gene from *Capsicum annuum* (pepper) in *Arabidopsis thaliana* increased its resistance to the hemibiotrophic pathogen Pseudomonas syringae and the biotrophic pathogen *Hyaloperonospora parasitica*. Sesquiterpene synthase catalyzes the cyclization of farnesyl diphosphate to sesquiterpenes, many of which are known to have antimicrobial activities [37-39]. In cotton, one sesquiterpene synthase, called delta-cadinene synthase, is one of the key enzymes for the biosynthesis of gossypol, a phytoalexin, which protects cotton from blight-causing pathogens and bollworms [40,41]. Also, a unigene homologous to isoliquiritigenin 2'O-methyltransferase, another gene involved in the terpenoid biosynthesis, was highly expressed in the root. Its homolog in alfalfa is involved in the biosynthesis of a flavonoid inducer of Rhizobium nodulation genes [42].

Tissue-specific expressions of genes involved in terpenoid biosynthesis have been observed in other plants, especially in medicinal plants with antimicrobial properties. Generally, plant tissues used for medicinal purposes have higher expression of genes involved in terpenoid biosytnehsis. For example, in the root of *Valeriana fauriei* (valerian), which is used for medicinal purposes, the terpenoid biosynthesis genes were highly upregulated [43]. Another plant *Asparagus racemosus*, whose tuberous roots are used in traditional Indian medicine, showed upregulation of terpenoid biosynthesis genes in roots compared to leaves [44]. On the other hand, in *Cymbopogon winterianus* (citronella), whose leaves are known to have antimicrobial properties, shows higher expression of terpenoid biosynthesis genes in leaves compared to roots [45]. Also, terpenoids accumulate in roots in response to drought in many plants, including *Tanacetum vulgare* (tansy) and maize [46,47]. Most likely these compounds act as antioxidants to protect plants against oxidative stress caused by drought. Consistent with these published reports, relatively high expression of these genes involved in terpenoid biosynthesis in the root of leucaena may be related to the disease-free and highly stress-tolerant nature of leucaena.

Upregulation of Nicotianamine Synthase in the Root

The most upregulated sequence in the root was homologous to nicotianamine synthase (NAS) with more than 600-fold difference. NASs synthesize a non-protein amino acid nicotianamine (NA) through trimerization of S-adenosylmethionine molecules [48]. NA is a chelator of bivalent metal ions, including Fe(II), Zn(II), and Cu(II) [49-51], and it is involved in intercellular and long-distant transport of these metal irons [52]. Therefore, NASs play an important role in the uptake of Fe from the rhizosphere. As one of the common ways to uptake Fe from soils, non-graminaceous plants first reduce Fe(III) to the more soluble form Fe(II) with the enzyme ferric reductase on the root surface [53-57], followed by uptake of Fe(II) through Fe(II) transporters. Then, NA delivers the metals intercellularly throughout plants. Many studies have shown the importance of NASs for plants for proper metal transport and homeostasis for their growth and development and for survival under Fe deficiency. An *Arabidopsis* mutant with full loss of NAS function was sterile, and another mutant with a reduced NAS function developed leaf chlorosis [58]. In maize, NASs showed differential expression patterns in different developmental stages and Fe availabilities; it showed upregulation in response to Fe deficiency so that the plant can uptake more Fe to survive [59]. Also interestingly, overexpression of NAS gene enhanced drought tolerance in the perennial ryegrass, *Lolium perenne* [60]. Although not studied in this expression analysis, we found 13 sequences homologous to ferric reductase in the root (≥ 500 bp; Table S4). It is critical for plants to maintain proper metal transport and homeostasis for their growth and development, and these enzymes may play an important role in leucaena's high adaptability and fast-growing nature; further studies will be necessary for their characterization in future.

Expression of Genes Sharing No Homology to the Public Databases

Leucaena is unique among other tree legumes in that it can be used as a high protein fodder and it is highly resistant to infection by plant pathogens and tolerant to environmental stresses. Thus, it is expected that leucaena will have certain genes, which are not found in other cultivated legumes such as soybeans. In this study, we found that a total of 47 unigenes that had no homology in the known protein database were differentially expressed in the root and shoot. Two of those were upregulated more than 100-fold in the root (Table S4).

Further characterization of those genes may lead to identification of novel genes unique to leucaena.

Conclusion

This is the first transcriptome-wide analysis of *L. leucocephala* using NGS technology. Illumina sequencing and SOAPdenovo assembly generated over 120,000 unigenes from roots and shoots combined, and we successfully annotated ~80% of them. Through the microarray and qRT-PCR analyses, the expression of the assembled unigenes was validated, and root- and shoot-specific sequences were identified. Many of the unigenes upregulated in the shoot were homologous to the genes involved in carbohydrate and lipid metabolisms, while those in the root shared homology with genes involved in the secondary metabolite biosynthesis. High expression levels of certain genes, such as terpenoid biosytnehsis genes, in the root may be related to leucaena's resistance to plant pathogens and tolerance to drought. Similarly, upregulation of NAS may be related to the fast-growing nature of this tree legume. Further characterization of these sequences will contribute to identification and isolation of genes for stress tolerance and disease resistance from leucaena. Our results will be a valuable resource for future genetic studies of leucaena and agroforestry improvement programs.

References

1. Brewbaker JL, Hegde N, Hutton EM, Jones RJ, Lowry JB, et al. (1985) Leucaena - Forage Production and Use. NFTA Hawaii. 39.

2. Brewbaker JL, Sorensson CT (1990) New tree crops from interspecific Leucaena hybrids. In: Janick J, Simon JE (Eds). Advances in new crops. Timber Press, Portland 283-289.

3. Shelton HM, Brewbaker JL (1994) *Leucaena leucocephala* -the most widely used forage tree legume. In: Gutteridge RC and Shelton HM (Eds) Forage Tree Legumes in Tropical Agriculture. CAB International Wallingford 15-29.

4. Martínez-Romero E, Segovia L, Mercante FM, Franco AA, Graham P, et al. (1991) Rhizobium tropici, a novel species nodulating Phaseolus vulgaris L. beans and Leucaena sp. trees. Int J Syst Bacteriol 41: 417-426.

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

6. National Academy of Science USA. (1977) Leucaena - Promising Forage and Tree Crop for the Tropics. National Academy Press Washington DC 115.

7. Bourges H, Morales de León JC (1980) The experience in Mexico on the utilization of non-conventional protein sources. Nutr Food Sci 2: 277-287.

8. Achmad MAM. The general uses of lamtoro (leucaena) in Indonesia. Leucaena Research Reports (EUA) 6: 57-58.

9. Pandey VC, Kumar A (2013) *Leucaena leucocephala*: An underutilized plant for pulp and paper production. Genet Resour Crop Evol 60: 1165-1171.

10. López F, García MM, Yánez R, Tapias R, Fernández M, et al. (2008) Leucaena species valoration for biomass and paper production in 1 and 2 year harvest. Bioresour Technol 99: 4846-4853.

11. Jayanthy V, Geetha R, Rajendran R, Prabhavathi P, Karthik Sundaram S, et al. (2014) Phytoremediation of dye contaminated soil by *Leucaena leucocephala* (subabul) seed and growth assessment of Vigna radiata in the remediated soil. Saudi J Biol Sci 21: 324-333.

12. Juson AEDS, Martinez MKM, Ching JA (2016) Accumulation and distribution of heavy metals in *Leucaena leucocephala* Lam and Bougainvillea spectabilis Wild plant systems. J Exp Biol Agric Sci 4: 1-6.

13. Boff T, Schifino-Wittmann MT (2003) Segmental allopolyploidy and paleopolyploidy in species of Leucaena benth: evidence from meiotic behaviour analysis. Hereditas 138: 27-35.

14. Ishihara K, Lee EKW, Borthakur D (2016) An improved method for RNA extraction from woody legume species Acacia koa A. Gray and *Leucaena leucocephala* (Lam.) de Wit. Int J Forestry Wood Sci 3: 31-35.

15. Luo R, Liu B, Xie Y, Li Z, Huang W, et al. (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1: 18.

16. Conesa A, Götz S (2008) Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics 2008: 619832.

17. Negi VS, Pal A, Singh R, Borthakur D (2011) Identification of species-specific genes from *Leucaena leucocephala* using interspecies suppression subtractive hybridisation. Ann Appl Biol 159: 387-398.

18. Andersen CL, Jensen JL, Ørntoft TF (2004) Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res 64: 5245-5250.

19. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25: 402-408.

20. Jain M (2011) A next-generation approach to the characterization of a non-model plant transcriptome. Curr Sci 101: 1435-1439.

21. Martin JA, Wang Z (2011) Next-generation transcriptome assembly. Nat Rev Genet 12: 671-682.

22. Morozova O, Hirst M, Marra MA (2009) Applications of new sequencing technologies for transcriptome analysis. Annu Rev Genomics Hum Genet 10: 135-151.

23. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, et al. (2010) Genome sequence of the palaeopolyploid soybean. Nature 463: 178-183.

24. Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, et al. (2008) Genome structure of the legume, Lotus japonicus. DNA Res 15: 227-239.

25. Young ND, Debellé F, Oldroyd GE, Geurts R, Cannon SB, et al. (2011) The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature 480: 520-524.

26. Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, et al. (2011) Draft genome sequence of pigeonpea (Cajanus cajan), an orphan legume crop of resource-poor farmers. Nat Biotechnol 30: 83-89.

27. Jain M, Misra G, Patel RK, Priya P, Jhanwar S, et al. (2013) A draft genome sequence of the pulse crop chickpea (Cicer arietinum L.). Plant J 74: 715-729.

28. Ishihara K, Lee EKW, Rushanaedy I, Borthakur D (2015) Illumina-based de novo transcriptome analysis and identifications of genes involved in the monolignol biosynthesis pathway in Acacia koa. American Journal of Bioinformatics 4:7-27.

29. Hiremath PJ, Farmer A, Cannon SB, Woodward J, Kudapa H, et al. (2011) Large-scale transcriptome analysis in chickpea (Cicer arietinum L.), an orphan legume crop of the semi-arid tropics of Asia and Africa. Plant Biotechnol J 9: 922-931.

30. Pelleschi S, Rocher JP, Prioul JL (1997) Effect of water restriction on carbohydrate metabolism and photosynthesis in mature maize leaves. Plant Cell Environ 20: 493-503.

31. Burke JJ (2007) Evaluation of source leaf responses to water-deficit stresses in cotton using a novel stress bioassay. Plant Physiol 143: 108-121.

32. Mohammadkhani N, Heidari R (2008) Drought-induced Accumulation of Soluble Sugars and Proline in Two Maize Varieties. World Appl Sci J 3: 448-453.

33. Hummel I, Pantin F, Sulpice R, Piques M, Rolland G, et al. (2010) Arabidopsis plants acclimate to water deficit at low cost through changes of carbon usage: an integrated perspective using growth, metabolite, enzyme, and gene expression analysis. Plant Physiol 154: 357-372.

34. Dixon RA, Paiva NL (1995) Stress-Induced Phenylpropanoid Metabolism. Plant Cell 7: 1085-1097.

35. Gershenzon J, Dudareva N (2007) The function of terpene natural products in the natural world. Nat Chem Biol 3: 408-414.

36. Choi HW, Lee BG, Kim NH, Park Y, Lim CW, et al. (2008) A role for a menthone reductase in resistance against microbial pathogens in plants. Plant Physiol 148: 383-401.

37. Erasto P, Grierson DS, Afolayan AJ (2006) Bioactive sesquiterpene lactones from the leaves of Vernonia amygdalina. J Ethnopharmacol 106: 117-120.

38. Hashidoko Y, Tahara S, Mizutani J (1989) Antimicrobial sesquiterpene from damaged Rosa rugosa leaves. Phytochemistry 28: 425-430.

39. Komaraiah P, Reddy GV, Reddy PS, Raghavendra AS, Ramakrishna SV, et al. (2003) Enhanced production of antimicrobial sesquiterpenes and lipoxygenase metabolites in elicitor-treated hairy root cultures of Solanum tuberosum. Biotechnol Lett 25: 593-597.

40. Townsend BJ, Poole A, Blake CJ, Llewellyn DJ (2005) Antisense suppression of a (+)-delta-cadinene synthase gene in cotton prevents the induction of this defense response gene during bacterial blight infection but not its constitutive expression. Plant Physiol 138: 516-528.

41. Mao YB, Cai WJ, Wang JW, Hong GJ, Tao XY, et al. (2007) Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. Nat Biotechnol 25: 1307-1313.

42. Maxwell CA, Harrison MJ, Dixon RA (1993) Molecular characterization and expression of alfalfa isoliquiritigenin 2'-O-methyltransferase, an enzyme specifically involved in the biosynthesis of an inducer of Rhizobium meliloti nodulation genes. Plant J 4: 971-981.

43. Park YJ, Li X, Noh SJ, Kim JK, Lim SS, et al. (2016) Transcriptome and metabolome analysis in shoot and root of Valeriana fauriei. BMC Genomics 17: 303.

44. Upadhyay S, Phukan UJ, Mishra S, Shukla RK (2014) De novo leaf and root transcriptome analysis identified novel genes involved in steroidal sapogenin biosynthesis in Asparagus racemosus. BMC Genomics 15: 746.

45. Devi K, Mishra SK, Sahu J, Panda D, Modi MK, et al. (2016) Genome wide transcriptome profiling reveals differential gene expression in secondary metabolite pathway of Cymbopogon winterianus. Sci Rep 6: 21026.

46. Kleine S, Müller C (2014) Drought stress and leaf herbivory affect root terpenoid concentrations and growth of Tanacetum vulgare. J Chem Ecol 40: 1115-1125.

47. Vaughan MM, Christensen S, Schmelz EA, Huffaker A, McAuslane HJ, et al. (2015) Accumulation of terpenoid phytoalexins in maize roots is associated with drought tolerance. Plant Cell Environ 38: 2195-2207.

48. Higuchi K, Suzuki K, Nakanishi H, Yamaguchi H, Nishizawa NK, et al. (1999) Cloning of nicotianamine synthase genes, novel genes involved in the biosynthesis of phytosiderophores. Plant Physiol 119: 471-480.

49. Beneš I, Schreiber K, Ripperger H, Kircheiss A (1983) Metal complex formation by nicotianamine, a possible phytosiderophore. Experientia 39: 261-262.

50. Weber M, Harada E, Vess C, Roepenack-Lahaye Ev, Clemens S (2004) Comparative microarray analysis of Arabidopsis thaliana and Arabidopsis halleri roots identifies nicotianamine synthase, a ZIP transporter and other genes as potential metal hyperaccumulation factors. Plant J 37: 269-281.

51. Curie C, Cassin G, Couch D, Divol F, Higuchi K, et al. (2009) Metal movement within the plant: contribution of nicotianamine and yellow stripe 1-like transporters. Ann Bot 103: 1-11.

52. Takahashi M, Terada Y, Nakai I, Nakanishi H, Yoshimura E, et al. (2003) Role of nicotianamine in the intracellular delivery of metals and plant reproductive development. Plant Cell 15: 1263-1280.

53. Eide D, Broderius M, Fett J, Guerinot ML (1996) A novel iron-regulated metal transporter from plants identified by functional expression in yeast. Proc Natl Acad Sci U S A 93: 5624-5628.

54. Yi Y, Guerinot ML (1996) Genetic evidence that induction of root Fe(III) chelate reductase activity is necessary for iron uptake under iron deficiency. Plant J 10: 835-844.

55. Robinson NJ, Procter CM, Connolly EL, Guerinot ML (1999) A ferric-chelate reductase for iron uptake from soils. Nature 397: 694-697.

56. Mori S (1999) Iron acquisition by plants. Curr Opin Plant Biol 2: 250-253.

57. Vert G, Grotz N, Dédaldéchamp F, Gaymard F, Guerinot ML, et al. (2002) IRT, an Arabidopsis transporter essential for iron uptake from the soil and for plant growth. Plant Cell 14: 1223-1233.

58. Klatte M, Schuler M, Wirtz M, Fink-Straube C, Hell R, et al. (2009) The analysis of Arabidopsis nicotianamine synthase mutants reveals functions for nicotianamine in seed iron loading and iron deficiency responses. Plant Physiol 150: 257-271.

59. Zhou ML, Qi LP, Pang JF, Zhang Q, Lei Z, et al. (2013) Nicotianamine synthase gene family as central components in heavy metal and phytohormone response in maize. Funct Integr Genomics 13: 229-239.

60. Zhang ZX, Zheng YZ (2008) Overexpression of nicotianamine synthase (NAS) gene results in enhanced drought tolerance in perennial ryegrass. Biotechnol Biotechnol Equip 22: 938-941.