Elucidating biosynthetic pathway of piperine using comparative transcriptome analysis of leaves, root and spike in *Piper longum* L.

Abstract

Piper longum (Pipli; Piperaceae) is an important spice valued for its pungent alkaloids, especially piperine. Albeit, its importance, the mechanism of piperine biosynthesis is still poorly understood. The Next Generation Sequencing (NGS) for *P. longum* leaves, root and spikes was performed using Illumina platform, which generated 16901456, 54993496 and 22900035, respectively of high quality reads. In *de novo* assembly *P. longum* 173381 numbers of transcripts were analyzed. Analysis of transcriptome data from leaf, root and spike showed gene families that were involved in the biosynthetic pathway of piperine and other secondary metabolites. To validate differential expression of the identified genes, 27 genes were randomly selected to confirm the expression level by quantitative real time PCR (qRT-PCR) based on the up regulation and down regulation of differentially expressed genes obtained through comparative transcriptome analysis of leaves and spike of *P. longum*. With the help of UniProt database the function of all characterized genes was generated.

Keywords- *Piper longum*, medicinal plants, piperine, NGS, Illumina, biosynthetic pathways, transcriptome analysis, secondary metabolites

Introduction

Piper longum popularly known as pippali (family: Piperaceae) is an important medicinal plant ranking third after *P. nigrum* (black papper) and *P. betle* (betel) in economic importance (Jose and Sharma, 1985). It grows extensively in hot and humid Indo-Malaya region (Manoj *et al.*, 2004). *P. longum* has several important bioactive compounds viz. alkaloid, flavonoids, glycosides, tannins, phenols and sterols (Rami *et al.*, 2013). Piper fruit has several alkaloids, piperine being the most important one and is the reason for its pungency (Zaveri *et al.*, 2010). The plant is a powerful stimulant for both digestive and respiratory systems and has a rejuvenating effect on the lungs (Charaka by Sharma, 1996). *Piper* fruit (spikes) plays an important role in multiple aliments as increased thyroid hormone level, thermogenic response, immune stimulatory, anti-ulcer, anti-amoebic, anti-oxidant, hepatoprotective, cardiovascular and anti-inflammatory activities (Dahanukar and Karandikar, 1984; Warrier *et al.*, 1995; Chauhan *et al.*, 2011; Sharma *et al.*, 2012; Prasad *et al.*, 2018). The abundance of medicinal value of this plant resulted in extensive and indiscriminate collection from the wild threating the very existence of it necessitating developing methods to conserve *P. longum* (Nair 2000; Rani and Dantu, 2012).

Collection pressure on *P. longum* could be relieved by either finding alternate source plants or through biotechnological intervention where piperine content in leaves could be increased. Piperine accumulation has been found to be highest in the spikes, followed by roots and least amount has been detected in the leaves of *P. nigrum* (Semler *et al.*, 1988) and *P. longum* (Prasad *et al.*, 2018, 2019). However, piperine biosynthesis in *P. longum*, its transportation within the plant and storage are still poorly understood. Although, Piperidine conversion to piperine using Piperoyl-CoA has been worked out in *P. nigrum* and steps of lysine conversion to piperidine have been elucidated in the bacterium *Pseudomonas aeruginosa* (Fothergill and Guest, 1977; Greisler and Gross *et al.*, 1990; Bunsupa *et al.*, 2012).

Despite the fact that the acetyl or malonyl CoA-based reaction of feruloyl-CoA to be a conceivable component for the C2-extention of the cinnamic acid- derived precursor ensuring piperoyl-CoA development. In these steps no data is available regarding molecular and enzymatic aspect. In another case, the related amide capsaicin derived from buturyl-CoA by C2- elongations which are encoded by Pun1 gene (Stewart et al., 2005). Explanation of detailed information of genome related to the pungency of capsaicin (Kim et al., 2014).

Recent advances in transcriptome analysis using Next Generation Sequencing has elucidated metabolic pathways in several medicinal plants, such as, *Artemisia annua* (Paddon and Keasling, 2014), *Withania somnifera* (Gupta *et al*., 2015), *Panax ginseng* (Zhang *et al*., 2017), *Picrorhiza kurroa* (Shitiz *et al*., 2015) to name a few. Transcriptome analysis of the seed (Hu *et al*., 2015) and root (Gordo *et al.,* 2012) of *P. nigrum* has generated a large data bank of gene sets involved in various processes. However, such a transcriptomic study in *P. longum* is still lacking. In non-model plants such as, *P. longum*, deep genome sequencing using Next Generation Sequencing is the only viable method for generating large molecular data set that can identify proteins (Yu *et al.,* 2011; Low and Heck, 2016; Low *et al*., 2019).

A detailed investigation of the genome of long pepper transcriptome using RNA-seq can give us a great insight into this species. The genomic data generated could be useful in understanding the genes involved in biosynthetic pathway of not only piperine but also other secondary metabolites of importance. The knowledge thus generated could be useful to understand ecology, evolution of this species and also in biotechnological interventions (Ekblom and Galindo, 2011; Bohan *et al*., 2017; Derocles *et al*., 2018).

In the present study transcriptome of long pepper leaves, roots and spikes were analyzed using Illumina HiSeq 2000 platform. Approximately 17, 55 and 23 million clean reads were generated from leaves, roots and spikes, respectively resulting in 1,73,381 unigenes. Analysis of the transcriptome data showed gene families that were involved in the biosynthesis of piperine and other secondary metabolites and in housekeeping genes. This is the first dataset of sequence analysis of long pepper leaves, roots and spikes and will be a useful genomic library for any molecular research.

Material and Methods

Plant materials: *Piper longum* plants were grown in net house of Department of Botany, Dayalbagh Educational Institute, Agra (U.P.), India. Young leaves, roots and spikes were collected in November from 4 years-old plants with the help of sterilized blade and transferred into RNAlater (Sigma-Aldrich, USA) for future use.

RNA isolation and library construction: Total RNA was isolated from three different tissues i.e. leaf, root and spike using Trizol (Invitrogen) method and Xcelgen plant RNA kit. The quality of the isolated RNA was checked on 1% Formaldehyde Denaturing Agarose gel and quantified using Nanodrop8000 spectrophotometer. The paired-end sequencing libraries were prepared using Illumina TruSeq Stranded mRNA Library Preparation Kit as per the protocol. Total isolated RNA was subjected to Oligo dT beads to enrich mRNA fragments, then subjected to purification, fragmentation and priming for cDNA synthesis. The fragmented

mRNA was converted into first-strand cDNA, followed by second-strand cDNA synthesis, A-tailing, adapter-index ligation and finally amplified by recommended number of PCR cycles. After obtaining the Qubit concentration for the library and mean peak size from Bio-analyzer 2100 (Agilent Technologies), library was loaded into Illumina platform for cluster generation and sequencing. Library quality and quantity check was performed using Agilent DNA High Sensitivity Assay Kit.

Transcriptome sequencing and analysis: RNA was pooled for sequencing and preparation was multiplexed on a single flow cell of an Illumina TruSeq. The final library was denatured and appropriate working concentration was loaded on to flowcell. Briefly, double stranded cDNA was subjected to Covaris shearing followed by end-repair fragments of overhangs. The fragments were A-tailed, adapter ligated and then enriched by limited number of PCR cycles. Prepared library was loaded into Illumina platform for cluster generation and sequencing was performed using paired-end (PE) 2x150 bp library on Illunima platform (Illumina TruSeq RNA 500) (Bolger *et al*., 2014). Subsequently, all assembled transcript contigs were validated using CLS Genomics Workbench 6 by plotting high quality reads back to the assembled transcript contigs. All CDS (protein coding sequences) were predicted from the assembled transcripts/ scaffolds using Transdecoder (rel16JAN2014) with default parameters.

Functional annotation and gene ontology (GO) analysis: The functional annotation were performed on the predicted CDS of plant sample by aligning the CDS to non redundant protein database of NCBI using Basic Local Alignment Search Tool (BLASTx) (Altschul *et al.*, 1990) with a minimum E-value less than 1e-5. CDSs were predicted with the help of Transdecoder software. BLASTx 2.2.30+ software was used for functional annotation detection standalone. The total CDS was found to have BLASTx hit against 'NR database of NCBI'. A BLAST explore was performed against the non-redundant protein (NR) and nucleotide sequences (NT) databases in the National Center for Biotechnology Information (NCBI) with an E value less than 1e-5. For annotating the transcript tags, the best match output from each BLAST was used. After each BLAST search annotation tags with no matches and ones with predicted annotations were extracted for next sequential BLAST search. GO assignment was used to classify the functions of the predicted CDS. The GO mapping also provided ontology of defined terms representing gene product properties which were group into three main chategories: molecular function (MF), biological process (BP) and cellular component (CC) (Ashburner *et al*., 2000). GO mapping was carried out using Blast2GO Pro software (Gotz *et al*., 2008) to retrieve GO terms for all the BLASTx functionally annotated CDS which included use of BLASTx result accession IDs to retrieve gene names or symbols.

Classification of assembled transcriptomes: KAAS (KEGG Automatic Annotation Server - http://www.genome.jp/kegg/ko.html) was used to functionally annotate the CDS by BLAST comparisons against KEGG gene database (Moriya *et al*., 2007). The BBH (Bi-directional best hit) option was used to assign KEGG Orthology (KO) terms. The KEGG Orthology database was used for pathway mapping. The unigenes were assignments of polypeptides produced from combined assembly and mapped into metabolic pathways according to KEGG (Kyoto Encyclopedia of Genes and Genomes).

Differential gene expression analysis: Differential gene expression analysis has been carried out for a total predicted CDS among (a) *P. longum* leaf (as Control) vs spike (as Tested), (b) *P.*

3

longum leaf (as Control) vs root (as Tested) and (c) *P. longum* root (as Control) vs spike (as Tested) using DESeq package (Wang *et al.,* 2010). Criteria which was taken for up regulated filtration is log2FC>2 AND p<0.05 and for down regulated filtration is log2FC<-2 AND p<0.05. Differentially expressed genes identified in *P. longum* leaf, root and spike were analyzed by hierarchical clustering (Eisen *et al*., 1998). A heat map was constructed using the log-transformed and normalized value of genes based on both Pearsonun centered correlation distance and on complete linkage method.

Transcription factors: For the identification of transcription factor families represented in transcriptome, all the predicted CDS were searched against all the transcription factor protein sequences in the Plant Transcription Factor database (PlnTFDB; http://planttfdb.cbi.pku.edu.cn/download.php) using BLASTx with e-value less than 1e-5.

Identification of Simple Sequence Repeats (SSRs): For identification of SSRs, all the assembled transcripts / scaffolds were searched with Perl Script MISA. SSRs generated from transcriptome sequence of leaves, root and spikes of same plants were used in Microsatellite program (MISA). Scrutiny contains microsatellites from di-nucleotide to hexa- nucleotide. Mono-nucleotide repeat motifs were not considered in the analysis due to chances of homopolymer tailing in ESTs generated. Unigenes were used as reference data in SSRs detection. SSRs having a flanking of 150 bp were filtered and used for primer designing (Temnykh *et al.*, 2000).

Quantitative gene expression analysis: The total RNA was isolated from samples using PureLink® RNA Mini Kit (Invitrogen, Thermo Fisher Scientific) and quality of RNA was checked on 1% denatured agarose gel using MOPS buffer. The cDNA synthesis from each RNA samples was carried out by using cDNA synthesis kit SuperScript® III First-Strand Synthesis System for RT-PCR (Invitrogen, US). PCR reactions were set up in 10µl reaction mixture in 96-well PCR plates (Don *et al*., 1991). PCR products were separated on 4% agarose gels (Smith *et al*., 2000). Relative and quantitative expression in leaves and spikes was analyzed using Real-Time PCR Detection Machine (Stratagene Mx3005P, Agilent Technologies) and KAPA SYBR® FAST qPCR Master Mix (2X) kit (KAPA BIOSYSTEMS,US). For relative analysis quantitative PCR was performed using cDNA templates; 10µl of 2X KAPA SYBR FAST qPCR Master Mix; 200nM of each primer (forward and reverse); 1X 50X ROX high/low and rest of nuclease free water for 20µl complete reaction volume. For relative analysis quantitative PCR was performed using the corresponding cDNA templates and SYBR master mix as per kit's instruction. For each set of primer, a control reaction was also included having no template. GAPDH gene was used as internal control to estimate the relative transcript level of the genes studied (Zhang *et al*., 2018). All the 27 genes including internal gene (GAPDH) was analyzed by qRT-PCR with the help of cycle parameters 95°C for 10 min followed by 40 cycles of 95°C for 30 sec, 55°C for 30 sec, 72°C for 30 sec and for melting curve 95°C for 1 min, 55°C for 30 sec and 95°C for 30 sec. Data from qRT-PCR was analyzed using comparative ΔΔct and fold changes in the transcript level were calculated using $2^{-\Delta\Delta ct}$ method (Livak and Schmittgen 2001 and Ranjan *et al*., 2012) considering the Ct value of GAPDH as the internal control. Each experiment was repeated using three biological replicates and three technical replicates, the data was analyzed statistically (± Standard Deviation).

Results

Sequencing and de novo transcriptome assembly: All the plant samples were collected and immediately stored in RNAlater. Total RNA was isolated using Trizol (Invitogen) method. RNAs of equal quality were mixed for Illumina TruSeq RNA sample preparation Kit sequencing. Raw reads of totaling 5,07,04,36,800 from leaves, 8,16,08,80,210 from roots and 6,87,00,10,500 from spikes were obtained. The NGS for *P. longum* leaves, root and spikes were performed using 2x150 bp on the Illumina Platform that resulted in generation of high quality data. The high quality reads statistics of *P. longum* leaves contained 1,69,01,456 paired reads, roots contained 5,49,93,496 paired reads and spikes contained 2,29,00,035 paired reads. *De novo* master assembly of pooled high quality paired end reads of *P. longum* roots, leaves and spikes samples were accomplished and total number of 1,73,381 transcripts were obtained (Table 1). Total transcriptome, N50 and maximum length of transcripts size in the libraries were 99612836, 722 and 12994 bases, respectively. The size distribution of the raw reads and assembled transcripts from libraries were characterized (Figure 1).

Feature of the combined assembly: To identify gene families of the samples of *P. longum,* the CDS obtained from assembly of each transcriptome were tagged, pooled and assembled again using Illumina TruSeq RNA sequencer. It has been considered that the super assembly of all primary assemblies is generated using different programs which provided better results in terms of consistency and size of transcripts. This analysis was carried out to get larger transcripts and to study digital differential expression of different tissues. CDS were predicted from the 1,73,381 assembled transcript sequences using Transdecoder (rel16JAN2014) at default parameters which resulted in identification of 58,773 CDS (Protein coding sequence) and total size of CDS was 3,87,34,722. The maximum and N50 length of CDS was 12990 and 747 bases. The size distribution of assembled CDS is displayed in Figure 2.

Functional Annotation of Predicted CDSs: Functional annotation of the CDS generated from the combined assembly of three different tissues of *P. longum* was carried out using Standalone BLASTx 2.2.30+. The annotation was performed on the predicted CDS of *P. longum* sample by aligning the CDS to non-redundant protein database of NCBI using BLASTx with e-value less than 1e- 5. BLASTx analysis statistics of the predicted CDS with BLAST hits was 50277 and without BLAST hits was 8496 out of total 58773 CDS sequences (Table 2). Out of these different databases, top-BLAST hit was observed in *Nelumbo nucifera,* about 11934 CDS, followed by *Vitis vinifera* 3816 CDS and *Elaeis guineensis* 3496 CDS (Figure. 3).

Gene annotation and functional classification: GO assignments were used to classify the functions of the predicted CDS. The GO mapping also provided ontology of defined terms representing gene product properties which were grouped into three main domains: Biological Process (BP), Molecular Function (MF) and Cellular Component (CC). BLASTx result accession IDs were searched directly in the gene product of GO database. As the international standardized gene functional classification system, GO and COG classifications were conducted for transcriptome data annotation. Among all of the categories, metabolic processes (24.6%) and cellular processes (21.9%) in the biological processes, membrane (13.8%) and cell (13.6%) in the cellular component, and binding (23.7%) and catalytic activity (22.4%) in the molecular function represented the major subcategories (Figure 4). In *P. longum*, 18508 CDS in biological process, 14429 CDS in Cellular component and 21103 CDS in Molecular function were obtained (Table 3).

5

Functional characterization and scanning of piperine-related gene using KEGG: KEGG pathway was used to identify the functional profiles of genes. The pathway analysis was carried out for all the samples using KEGG automatic annotation server. All the CDS were compared against the KEGG database using BLASTx with threshold bit-score value of 60 (default). The mapped CDS represented metabolic pathways of carbohydrates, lipids, nucleotides, amino acids, secondary metabolites, terpenoids, polyketides etc. The mapped CDS also represented the genes involved in genetic information processing, environmental information processing and cellular processes. A total of 2097 CDS (3.5%) were mapped into metabolism category followed by other pathways for *P. longum* (Figure 5). Among the CDS count 113 (0.2%), 290 (0.5%) and 103 (0.2%) were involved in biosynthesis of other secondary metabolites, amino acid metabolism and metabolism of terpenoids and polyketides, respectively. This data provided a valuable resource that might be used to study specific processes and pathways in the development of *P. longum*. Few genes were identified as piperidine pathway such as Lysine biosynthesis and Amino Acid biosynthesis related genes. In this pathway Tropane, piperidine and pyridine alkaloid biosynthesis [PATH: ko00960] represented 14 CDS.

Differential gene expression analysis, primer designing and validation: Based on the common hits accession of functionally annotated CDS in samples, a total of 58773 CDS between leaf, root and spike were predicted. Differential gene expression analysis were divided into three different catagories: leaf vs root, leaf vs spike and root vs spike.

(a) **Differential gene expression analysis between *P. longum* leaf (Control) and *P. longum* roots (Treated):** Differential gene expression analysis was carried out using commonly occurring NR blast hit accessions in the two samples. Within the set of 30296 commonly differentially expressed genes, a total of 581 genes were identified as significantly differentially up-regulated expressed and 1307 genes as significantly differentially down-regulated based on conditions p-value<0.05 (Table 4).

(b) Differential gene expression analysis between *P. longum* leaf (Control) and *P. longum* Spike (Treated): Within the set of 32216 commonly differentially expressed genes, a total of 625 genes were identified as significantly differentially up-regulated expressed and 1246 genes as significantly differentially down-regulated based on conditions p-value<0.05 (Table 5).

(c) Differential Gene Expression Analysis between *P. longum* roots (Control) and *P. longum* Spike (Treated): Within the set of 30225 commonly differentially expressed genes, a total of 993 genes were identified as significantly differentially up-regulated expressed and 878 genes as significantly differentially down-regulated based on conditions p-value<0.05 (Table 6).

Detection of SSR markers: For detection of SSRs all the transcript contigs were searched with Perl script MISA (Microsatellite Searching Tool). A total of 8041 SSRs were found in assembled transcripts of *P. longum* (Table 7). Out of 8041 SSRs, 2619 SSRs were filtered to satisfy the 150 bp flanking criteria for *P. longum*. In these SSRs, the trinucleotide repeat motifs were the most abundant, accounting for 4506 SSRs (56.04%), followed by 3209 (39.91%)

dinucleotide repeat motifs, 216 (2.69%) tetranucleotide repeat motifs, 79 (0.98%) pentanucleotide repeat motifs and 31 (0.38%) hexanucleotide repeat motifs (Table 8).

Transcription factors (TFs): Transcription factor encoding transcripts were analyzed by CDS comparison to known transcription factor gene families. In total, 21235 transcription factor genes, distributed in at least 65 families, were identified. In *P. longum* most abundant factors identified were bHLH, NAC and MYB_related which contain 2264, 1508 and 1370 CDS respectively. During the annotation analysis of the transcriptome data of *P. longum*, we have identified several TFs belonging to different families (Figure 6).

Identification of candidate genes involved in the piperine biosynthesis pathway: A total of 4730 CDS counts were assigned to 377 different pathways in KEGG database. Biosynthesis of secondary metabolites were categoried into 28 sections including phenylpropanoid biosynthesis (61 CDS), flavonoid biosynthesis (17 CDS), Tropane, piperidine and pyridine alkaloid biosynthesis (14 CDS), Isoquinoline alkaloid biosynthesis (11 CDS), Stilbenoid, diarylheptanoid and gingerol biosynthesis (9 CDS), Streptomycin biosynthesis (8 CDS) and biosynthesis of other secondary metabolites (Figure 8). Higher proportion of CDS belonged to the pathway of Translation, Signal transduction and Carbohydrate metabolism. Majority of CDS were classified into the metabolism, and the number of CDS related to different secondary metabolisms. To explore the regulatory mechanisms for the accumulation of piperine in *P. longum,* the expression profile of genes involved in piperine biosynthesis were analyzed (Figure 7). A total 14 expressed CDS encoding Tropane, piperidine and pyridine alkaloid biosynthesis enzymes were identified in *P. longum* (Table 9).

Validation of Expression of Differentially Expressed genes- To validate differential expression, 27 genes were selected (Table 10, 11) to confirm the expression level by Quantitative Real Time PCR (qRT-PCR) m. Validation of expression of selected genes through qRT-PCR suggests that differentially expressed genes identified through digital gene expression analysis might be differentially expressed between samples (Figure 8).

Development piperine biosynthesis-related gene- Piperoyl CoA and piperidine are the two precursor of piperine using acyltransferase. However, piperine is related with three main groups:

Group A: Gene associated with phenylpropanoid pathway (map00940). In this pathway cinnamoyl-CoA is developed for piperoyl CoA biosynthesis.

Group B: Genes related with L-lysine metabolism (map01064). In this pathway lysine is converted into piperidine using series of reaction.

Group C: Genes allied with acyltransferase, which perform catalytic activity in between piperoylcoenzyme A and piperidine.

The spike of plant shows the importance of arogenate dehydratase (ADT), amino transferase (PPA-AT), p-coumarate 3-hydroxylase (C3H), cinnamate 4- hydroxylase (C4H/CYP7) HCT, caffeic acid-3-O-methyltransferase (CAOMT), p-coumarate- CoA-ligase (4CL). All these genes are concerned in the reaction of phenylpropanoids to piperoyl CoA. The concern

7

transcripts associated with different enzymes and compounds are shown in the Table 12. Finally, the piperine synthesis and their related genes are observed expansion of multiple gene families.

Discussion

In nature 12,000 alkaloids are known and many more are yet to be discovered (Facchini *et al.*, 2012). Transcriptome databases of specific plant species generate numbers of biosynthetic genes encoding enzymes with valuable catalytic activities and variants with similar functions, but different biochemical features. As compared to complete genome sequencing, transcriptome sequencing is a good substitute for rapid and efficient access to the expressed genes and characterizing phenotypes (Wang *et al.*, 2010). There is a benefit of *de novo* assembly and gene models could be obtained that otherwise may not be assembled by any other method as the reads may not be mapped back to the current genomic scaffolds due to improper scaffold assembly. A differential expression of gene analysis for different tissues was investigated in the present study with Digital Gene Expression (DGE) profiling technology (Strickler *et al.,* 2012) to understand the functional genomics for *P. longum*. Previous studies showed that de novo transcriptome assembly using high quality Illumina RNA-Seq reads were more useful for genome assembly. However, if the gene transcripts are fragmented and gene annotation is incomplete, de novo assembly can help in the reading of gene map back to the assembly and also in gene model representation. If a genome assembly is present, it can be useful to unite *de novo* when the databases are available (Huang *et al*., 2016).

For Next Generation Sequencing (NGS) with the help of RNA-Seq data, transcriptome assembly is the first important step. However, in non-model plants NGS is a challenge, especially due to the high level of ploidy, large gene size, enormous quantity of data, large proportion of repeat sequences and low genome complexities produced by transcriptome assembly (Wang *et al.,* 2010). Aim of our study was to provide important and useful information to the database which is helpful to the researchers accomplishing transcriptome analysis in non-model plants. Gene transcripts show better expression characteristics in plants and also provide metabolic engineering alternative for optimization of synthetic biosystems designed to produce valuable plant metabolites (Xiao *et al.,* 2013).

The Next Generation Sequencing study in non-model plants as *P. longum* will lead to identification of pathways and gene related to production of natural products such as piperine and would also help to discover additional pathways. Transcriptome analysis is cost effective approach of sequence determination and improves the efficiency and speed of gene discovery (Jose and Sharma, 1985; Vasavirama and Upender, 2014). *P. longum* is an important plant used for medicinal purposes but transcriptome and genetic information are not available in NCBI database. Leaves, roots and spikes of this plant are used for medicinal purpose. Despite having immense medicinal value, there is very little information available regarding biosynthesis of piperine from this plant (Vasavirama and Upender, 2014) and lack the complete genetic data (Paterson *et al*., 2006; Jiao *et al*., 2011; Sheila *et al*., 2012). In GenBank database, genome of black pepper (*P. nigrum*) is represented by only 184 sequences (Jaramillo *et al*., 2001; Menezes *et al*., 2009; Sen *et al*., 2010; Joy *et al*., 2007).

8

In the comparative characterizations of the leaf, root and spike transcriptome of *P. longum* valuable resources were generated for new gene discovery and development of SSR markers for further study. To the best of our knowledge, there is no other report on Illumina paired-end sequencing technology for *P. longum* leaf, root and spike transcriptome de novo sequencing and assembly without reference genome. The results demonstrated that Illumina paired end sequencing can be used as a fast and efficient approach to gene discovery and molecular marker development for non-model organism, especially those with larger genome (Wang *et al.*, 2010; Martin and Wang, 2011; Xiao *et al.*, 2013; Xu *et al.,* 2015; Hagel *et al.*, 2015; Jain, 2011). This information represents a major trancriptomic level resource for *P. longum* and will be useful for comparative transcriptomic studies in *P. longum*. These analyses would provide genomic information to understand patterns of adaptive variation across the genome and to identify the genetic basis of adaptive traits (Parchman *et al.*, 2010).

In the present analysis, a total of 99 million high quality reads were generated from leaves, roots and spikes of *P. longum* using Illumina TruSeq RNA. From these reads, 1,73,381 transcripts were sequenced and analyzed generating 8,041 SSRs. It gives a new fingerprinting for taxonomic contrasts, species mapping for this species. These analyzed data would help to describe the mechanisms and genes involved in alkaloid biosynthesis in non-model plants. Transcriptomic studies on medicinal plants as *Cassia angustifolia* (Reddy *et al.*, 2015), *Panax vietnamensis* (Zhang *et al.*, 2015), *Phyllanthus amarus (*Bose and Chattopadhyay, 2016) and *Chelidonium majus* (Pourmazaheri *et al.*, 2019) has revealed important information regarding biosynthetic pathways of the active principles present in these plants. In these studies, only few transcripts were annotated by GO distribution as limited database is available for non-model plant sequences resulting in low hits in gene identification. SSRs are more informative and reliable markers in organism analysis (Idrees and Irshad, 2014; Madhumati, 2014; García *et al.*, 2018; Ahmad *et al.,* 2018).

The functional annotation of the dataset of CDS revealed synthesis of Lysine in *P. longum*. In spike, the activity of amino acid especially Lysine is very high as compared to leaf and root due to presence of piperine. Several differentially expressed transcripts were identified and analyzed in the three plant tissues of *P. longum*. Comparative transcriptome study of *P. longum* identified exclusive gene in leaf (2741 CDS) and root (1595 CDS) by the investigation of leaf and root tissue sample. Same analysis was found in leaf (821 CDS) and spike (1033 CDS) by the comparative analysis of both these tissues. Again a differential gene expression analysis identified exclusive genes in root (3024 CDS) and spike (1667 CDS).

Gene Ontology (GO) unites genes and gene product attributes across all the species (The Gene Ontology project, 2008) and also provideds information about the role of metabolic alkaloids. GO analysis identified 28 transcripts involved in the secondary metabolite pathway specially related to the piperine production including Lysine biosynthesis. In future, these transcripts will be important resources for genetic manipulation of *P. longum*. The metabolic pathway of piperine are not known and little information is available on biosynthesis of tropane, piperidine and pyridine alkaloid biosynthesis in plants (Sui *et al.*, 2011, Liu *et al.,* 2015; Zhang *et al.,* 2015; Liao *et al.,* 2016; Mazumdar and Chattopadhyay, 2016; Liu *et al.,* 2017; Shou-ming *et al.,* 2017; Li *et al.,* 2018; Lei *et al.*, 2018).

Piperine is a lysine-derived alkaloid and in the first step Lysine through decarboxylation produces cadaverine (Bunsupa *et al.,* 2014). Cadaverine is converted into piperidine (Wink and Hartmann, 1979, Wink, 1987, Okwute *et al.*, 2013). Some of CDS related to enzymes involved in these steps also showed different expression (Gordo *et al.*, 2012 and Joy *et al.*, 2013). The number of Illumina reads in each CDS is directly proportional to the abundance of specific cDNA in the library. This study helped us to study differential expression of genes related to pathway and their specific role in tissue-specific piperine biosynthesis and molecular mechanism of leaf, root and spike tissue. TFs control gene expression in response to various internal and external signals by suppressing and enhancing downstream genes (Reddy *et al.*, 2015). These factor families known to manage secondary metabolism play an important role in regulation of piperine biosynthesis. TFs family plays critical roles in interactions with other molecules (Takatsuji, 1998; De *et al.*, 1999).

A large number of SSR markers were identified from the transcriptome data and it is an important molecular tool to be used in gene mapping, genetic diversity and molecular breeding. SSR markers also provide useful reference data for further research. Transcriptional factors regulate gene expression of external and internal signals for activating or suppressing downstream genes (De Folter *et al.*, 2005). Transcripts were found to be putative transcription factor encoding regions not belonging to any particular transcription factor family. These families were known to regulate secondary metabolism and play important role in piperine biosynthesis. C3H transcription factors belong to zinc finger motif family that plays a major role in interactions with other molecules (Takatsuji, 1998; De *et al.*, 1999). Functional characterization of the candidate genes will not only help the biochemical mechanism of compound, but also provide molecular and biochemical target for biotechnological interventions.

In this study, comparative study of piperine biosynthesis related with tropane, piperidine and pyridine alkaloid biosynthesis pathway were addressed, and 15 functional genes were found (Reddy *et al.,* 2015, Ishihara *et al.,* 2015, Sun *et al.*, 2017 and Lei *et al.,* 2018). In the first attempt, more genes were significantly down-regulated than up-regulated as compared with leaf and root. In the second attempt, again more genes were down-regulated than up-regulated as compared with leaf and spike. In the last attempt, more genes were up-regulated than down-regulated as compared with root and spike. Pathway based analysis for the leaves and spike of *P. longum* is helpful to further understand the function and gene interactions. Lysine are the most important and basic structure for the piperine biosynthetic pathway (Szoke et al., 2013, Hu et al., 2015, Chopra, et al., 2016). In biosynthetic process of piperine, N-heterocycle piperidine and thioester piperoyl- CoA was transferring their group (Geissmann and Crout, 1969). The decarboxylation of L-lysine in the presence of pyridoxal phosphate (PLP) formed cadaverine. Then oxidative deamination via enzyme oxidative deamination was produced amino aldehyde and after cyclization provides imine, Δ1- piperideine. And the product were further reduced to piperidine and after the reaction with piperoyl-CoA final product piperine was created (Simon and Henry, 2013). Still this process has not been proven yet, thus many enzymatic analysis depend upon the biosynthesis of similar compounds like coumaroylagmatine (Bird and Smith, 1983, 1989) and feruloyltyramine (Negrel and Martin, 1984, Negrel and Jeandet, 1987) were carried out. But the reaction between piperoyl-coenzyme A and 2-piperidine that produces piperine was shown in the shoots of *P. nigrum* (Semler et al., 1897; Semler and Gross, 1988).

The pathway of piperine is starts with shikimate being converted to shikimate 3-phospthate by shikimate kinase (2.7.1.71), and formed 5-O-(1-carboxyvinyl)-3-phosphoshikimate with the help of 3-phosphoshikimate 1-carboxyvinyl transferase (2.5.1.19) (KEGG pathway) (Legrand et al., 2016). Then by using chorismate synthase (4.2.3.5), chorismate produced. L-phenylalanine is synthesized after two steps from chorismate (Figure 9). The phenylalanine being converted to cinnamic acid by phenylalanine ammonia lyase (PAL), and diverges into several branches at q-coumaroyl CoA (Dewick, 2002; Ehlting et al., 2006). The first enzyme in the phenylpropanoid pathway is phenylalanine ammonia-lyase (PAL). Three other enzymes of phenylpropanoid and lignin metabolism, were shown to increase, notably 4-coumarate: CoA ligase (4CL), caffeic acid/5-hydroxyconiferaldehyde O-methyltransferase (COMT) and cinnamyl alcohol dehydrogenase (CAD) (Hoffmann et al., 2003). Thus, 4 enzymes involved in the phenylpropanoid and lignin pathways are PAL, p-hydroxycinnamoyl-CoA: quinate shikimate p-hydroxycinnamoyltransferase (HCT), cinnamoyl-CoA reductase (CCR) and CAD (Andersen et al., 2008; Peng et al., 2008).

Piperine consists of 3,4-methylenedioxyphenyl piperidine connected by a C5 amide bridge. 3,4-Methylenedioxyphenyl and part of the bridging chain is known to bifurcate from the phenylpropanoid pathway (Richet et al., 2012). The biosynthesis of piperine in that 3,4-methylenedioxycinnamoyl-CoA, derived through phenylpropanoid pathway, would undergo one C2 chain extension using malonyl-CoA to arrive at piperoyl-CoA (Schnabel et al., 2020). The presence of (E,E)-piperoyl-CoA: piperidine N piperoyltransferase, catalyzing the ultimate reaction in the piperine pathway to form piperine in the shoots of the plant (Guzman, 2014). The above-mentioned putative pathway proposed for CoA thioester formation of phenylpropenoate(s) at the certain stage of phenylpropanoid pathway to activate the carboxyl group of the acid(s) (Chen et al., 2013). This type of activation is commonly carried out by 4-coumarate: CoA ligase (4CL) (Jin et al., 2020). Members in 4CL family can be classified into three subfamilies according to their catalytic functions and amino acid sequences.

Class I: 4CLs are typically involved in lignin biosynthesis, whereas
Class II: 4CLs participate in plant's response to environmental stress and in the biosynthesis of specialized metabolites.
Class III: Coumarate:CoA ligase-like (CLL) clade was recently described

In piperine, the aromatic part of the molecules is linked by an amide bond piperidine. The piperidine heterocycle is derived from L-lysine, whereas the aromatic part of piperine is likely to be derived from phenylpropanoid metabolism. Piperoyl-CoA is required for the piperine synthase activity. Although the acetyl or malonyl CoA-based elongation of feruloyl-CoA appears to be a possible mechanism for the C2-extension of the cinnamic acid-derived precursor resulting in piperoyl-CoA formation. Based on computational analysis, we now report the piperic acid is accepted as a substrate. PipCoA ligase might be required for the accumulation of piperine during early fruit ripening and may provide the piperoyl-CoA ester for the subsequent piperine transferase reaction, the key to piperine.

In an article the spike specific expression related genes contains lysine decarboxylase (LDC), glycosyltransferases (GFTs), cytochrome P450s (CYP); shikimate hydroxycinnamoyl transferases (HCT) in the phenylpropanoid pathway. The amino acid pathways,

11

acyltransferases (ATs) and phenylpropanoid are ubiquitous in plant secondary metabolism (Hu et al., 2019). In case of Capsicum species, the capsaicinoids are derived from phenylpropanoid (Aza-González et al., 2011). Lysine-derived quinoline alkaloids are create by the synthesis of phenylpropanoid and lysine metabolism in Papaver somniferum, Macleaya cordata (Liu, X. et al., 2017), Nelumbo nucifera and Carica papaya (Bennett et al.,1997).

From the functional annotation of *P. longum* plant two genes were found related to the pyridoxal phosphate (PLP) viz., Cys Met pyridoxal phosphate-dependent enzyme and Pyridoxal phosphate (PLP)-dependent transferases superfamily isoform 1. Both are important genes and they may help in further study of piperine biosyhthesis.

The shape of inflorescence architecture is depended on the meristems produced in the inflorescence apex which defines the relative position of flowers. In *P. longum* plant the fruit is a spike form of inflorescence due to their floral transition. The development of inflorescence may be mostly expressed by the function and mutual combination of three genes: APETALA 1 (AP1), LEAFY (LFY) and TERMINAL FLOWER 1 (TFL1) (Shannon and Meeks-Wagner, 1993; Liljegren *et al*., 1999; Blazquez *et al*., 2006, Benlloch *et al.,* 2015, Zhang *et al.,* 2015, Cheng *et al*., 2018, Ma *et al.,* 2018). These genes help to mentain the balance between the floral meristem and inflorescence (Alvarez *et al.,* 1992, Ratcliffe *et al.,* 1999; Blazquez *et al.*, 2006). LFY gene is involved in the identification, regulation and development of floral meristem (Schultz and Haughn, 1991, Shannon and Meeks-Wagner, 1993). In our study, functional annotation of the transcripts revealed two genes related to LFY gene, one gene for LEAFY and again to gene function like TFL gene. The inflorescence of plant is a complex network that entails multiple pathways. Abundant genes actively participate in the process of inflorescence development, viz *STERILE APETALA* (*SAP*), *LEAFY* (*LFY*) and few floral homeotic genes, *AGAMOUS* (*AG*) and *APETALA2* (*AP2*). In C Class, gene AG, LFY is an up-regulated and in Class A, AP2 and SAP are down-regulated, which is involved in the development of reproductive organ (Krogan *et al.*, 2012). These genes, we have found, are present in piper plant as well with four CDS in SAP, eight CDS in APETALA2 and seven CDS in AGAMOUS. Both WUSCHEL (WUS) and ULTRAPETALA 1 genes play important roles in flowering process such as activation of AG and ovule development (Lenhard et al., 2001). In the present study 21 annotations have been assigned to WUSHEL genes while three annotations have been assigned to ULTRAPETALA 1. AGL (MADS-box) transcripts are necessary for the development of reproductive organs (de Folter *et al*., 2005). In our study 19 genes have been annotated to AGL and 35 genes involved in LRR-RLK. GIGANTEA gene help in the end control of flowering, Circadian Clock and 27 annotations have been assigned to GIGANTEA in our study on *P. longum* (Peláez et al., 2019, Zhou et al., 2019). By using KEGG mapping of the best hits CDS, we have identified large number of CDS involved in different process viz. genetic information processing, metabolism, cellular processes, environmental information processing and organizational systems.

However, synthesis of piperine is extracted from decarboxylation and amine oxidation of lysine. And also they recommended as polymerisation of two tyrosines in benzylisoquinoline alkaloid biosynthesis. The derivative of phenylpropanoid and lysine metabolism is the two precursors which produce piperine after catalysed by acyltransferase. Thecombination of phenylpropanoid and lysine metabolism particularly decarboxylation and amine oxidation of lysine and acyl transformation reveal the feature of piperine synthesis.

Sequencing of long pepper has developed our consideration of the different piperine biosynthesis of Piper longum. Our analysis provides valuable impact that may create a base for the future research.

Acknowledgments

Reference

- Ahmad, A., Wang, J. D., Pan, Y. B., Sharif, R., & Gao, S. J. (2018). Development and use of simple sequence repeats (SSRs) markers for sugarcane breeding and genetic studies. *Agronomy*, *8*(11), 260.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol.; 215: 403–410. PMID: 2231712.
- Alvarez, J., Guli, C.L., Yu, X.H. and Smyth, D.R., 1992. terminal flower: a gene affecting inflorescence development in Arabidopsis thaliana. *The Plant Journal*, *2*(1), pp.103-116.
- Andersen, J. R., Zein, I., Wenzel, G., Darnhofer, B., Eder, J., Ouzunova, M., & Lübberstedt, T. (2008). Characterization of phenylpropanoid pathway genes within European maize (Zea mays L.) inbreds. *BMC Plant Biology*, *8*(1), 2.
- Aparupa Bose Mazumdar and Sharmila Chattopadhyay, 2016. Sequencing, *De novo* Assembly, Functional Annotation and Analysis of *Phyllanthus amarus* Leaf Transcriptome Using the Illumina Platform, *Front. Plant Sci. 6:1199, 1-22.*
- Benlloch, R., Berbel, A., Ali, L., Gohari, G., Millán, T. and Madueño, F., 2015. Genetic control of inflorescence architecture in legumes. *Frontiers in plant science*, *6*, p.543.
- Bennett, R. N., Kiddle, G. & Wallsgrove, R. M. Biosynthesis of benzylglucosinolate, cyanogenic glucosides and phenylpropanoids in Carica papaya. Phytochemistry 45, 59–66 (1997).
- Bird, C.R.; Smith, T.A. Agmatine coumaroyltransferase from barley seedlings. Phytochemistry, 1983, 22, 2401-2403. [http://dx.doi.org/10.1016/0031-9422(83)80127-7]
- Bird, C.R.; Smith, T.A. The biosynthesis of coumarylagmatine in barley seedlings. Phytochemistry, 1989, 20, 2345-2346. [http://dx.doi.org/10.1016/S0031-9422(00)82662-X]
- Blazquez, M.A., Ferrandiz, C., Madueño, F., and Parcy, F. (2006). How floral meristems are built. *PlantMol.Biol.* 60, 855–870.doi:10.1007/s11103-006- 0013-z.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics. (2014); 30: 2114–2120.
- Bose Mazumdar, A. and Chattopadhyay, S., 2016. Sequencing, de novo assembly, functional annotation and analysis of Phyllanthus amarus leaf transcriptome using the Illumina platform. *Frontiers in plant science*, *6*, p.1199.

- Bunsupa S, Katayama K, Ikeura E, Oikawa A, Toyooka K, Saito K, et al. (2012) Lysine decarboxylase catalyzes the first step of quinolizidine alkaloid biosynthesis and coevolved with alkaloid production in leguminosae. The Plant Cell 24: 1202–1216.
- Bunsupa, S., Komastsu, K., Nakabayashi, R., Saito, K., & Yamazaki, M. (2014). Revisiting anabasine biosynthesis in tobacco hairy roots expressing plant lysine decarboxylase gene by using 15N-labeled lysine. *Plant Biotechnology*, *31*(5), 511-518.
- Chauhan K, Solanki R, Patel A, Macwan C and Patel M (2011) Phytochemical and Therapeutic potential of *Piper longum*Linn. A review. *IJRAP*. 2(1):157-161.
- Chen, H.C., Song, J., Williams, C.M., Shuford, C.M., Liu, J., Wang, J.P., Li, Q., Shi, R., Gokce, E., Ducoste, J. and Muddiman, D.C., 2013. Monolignol pathway 4-coumaric acid: Coenzyme a ligases in populus. trichocarpa: Novel specificity, metabolic regulation, and simulation of coenzyme a ligation fluxes. *Plant physiology*, *161*(3), pp.1501-1516.
- Cheng, X., Li, G., Tang, Y. and Wen, J., 2018. Dissection of genetic regulation of compound inflorescence development in Medicago truncatula. *Development*, *145*(3), p.dev158766.
- Chopra, B., Dhingra, A.K., Kapoor, R.P. and Prasad, D.N., 2016. Piperine and its various physicochemical and biological aspects: A review. *Open Chemistry Journal*, *3*(1).
- Chris J. Paddon& Jay D. Keasling (2014) Artemisinin biosynthesis pathway in the plant *Artemisia annua*.Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development, *Nature Reviews Microbiology*, 12,355–367.
- Chun Sui, Meng Chen, Jiesen Xu, Jianhe Wei, Yue Jin, Yanhong Xu, Jing Sun, Ke Gao, Chengmin Yang, Zheng Zhang, Shilin Chen and Hongmei Luo, 2015. Comparison of root transcriptomes and expressions of genes involved in main medicinal secondary metabolites from *Bupleurum chinense* and *Bupleurum scorzonerifolium*, the two Chinese official *Radix bupleuri* source species, Physiologia Plantarum 153: 230–242.
- Dahanukar SA and Karandikar SM (1984) Evaluation of anti-allergic activity of *Piper longum*. *Indian Drugs*. 21:377–83.
- de Folter, S. *et al*. Comprehensive interaction map of the *Arabidopsis* MADS box transcription factors. *Plant Cell* 17, 1424–1433, https://doi.org/10.1105/tpc.105.031831 (2005).
- De Folter, S., Immink, R.G., Kieffer, M., Pařenicová, L., Henz, S.R., Weigel, D., Busscher, M., Kooiker, M., Colombo, L., Kater, M.M. and Davies, B., 2005. Comprehensive interaction map of the Arabidopsis MADS box transcription factors. *The Plant Cell*, *17*(5), pp.1424-1433.
- De J, Lai WS, Thorn JM, Goldsworthy SM, Liu X, Blackwell TK, et al. Identification of four CCCH zinc finger proteins in Xenopus, including a novel vertebrate protein with four zinc fingers and severely restricted expression. Gene. 1999; 228: 133–145.
- De J, Lai WS, Thorn JM, Goldsworthy SM, Liu X, Blackwell TK, et al. Identification of four CCCH zincfinger proteins in Xenopus, including a novel vertebrate protein with four zinc fingers and severelyrestricted expression. Gene. 1999; 228: 133–145. PMID: 10072766
- Delaney KJ, Xu R, Zhang J, Li QQ, Yun KY, Falcone DL, et al. Calmodulin interacts with and regulatesthe RNA binding activity of an Arabidopsis polyadenylation factor subunit. Plant Physiol. 2006; 140:1507–1521.

- Dengqun Liao, Pengfei Wang, Chan Jia, Peng Sun, Jianjun Qi, Lili Zhou & Xianen Li, 2016. Identification and developmental expression profiling of putative alkaloid biosynthetic genes in *Corydalis yanhusuo* bulbs, Scientific Reports | 6:19460, 1-14.

- Dewick, P. M. (2002). *Medicinal natural products: a biosynthetic approach*. John Wiley & Sons.pp- 1-546.

- Dolly rani and Prem Kumar Dantu, 2012. Direct shoot regeneration from nodal, internodal and petiolar segments of Piper longum L. and in vitro conservation of indexed plantlets. Plant Cell Tissue and Organ Culture 109(1), 9-17.

- Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucl. Acids Res. 1991; 19, 4008.

- Ehlting, J., Hamberger, B., Million-Rousseau, R., & Werck-Reichhart, D. (2006). Cytochromes P450 in phenolic metabolism. *Phytochemistry Reviews*, *5*(2-3), 239-270.

- Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome wide expression patterns. ProcNatlAcadSci USA 95: 14863–14868.

- Facchini, P.J., Bohlmann, J., Covello, P.S., De Luca, V., Mahadevan, R., Page, J.E., Ro, D.K., Sensen, C.W., Storms, R., Martin, V.J.J., 2012. Synthetic biosystems for the production of high-value plant metabolites. Trends in Biotechnology 30, 127–131.

- Fothergill JC and Guest JR (1977) Catabolism of L-lysine by *Pseudomonas aeruginosa*. *J Gen Microbiol*. 99(1):139-55.

- Gagne SJ, Stout JM, Liu E, Boubakir Z, Clark SM, Page JE. Identification of olivetolic acid cyclase from *Cannabis sativa* reveals a unique catalytic route to plant polyketides. ProcNatlAcadSci USA. 2012;109: 12811–12816.

- Gahlan P, Singh HR, Shankar R, Sharma N, Kumari A, Chawla V, et al. De novo sequencing and characterization of *Picrorhizakurrooa* transcriptome at two temperatures showed major transcriptome adjustments. BMC Genomics. 2012;13: 126.

- García, C., Guichoux, E., & Hampe, A. (2018). A comparative analysis between SNPs and SSRs to investigate genetic variation in a juniper species (Juniperus phoenicea ssp. turbinata). *Tree Genetics & Genomes*, *14*(6), 87.;

- Geissmann, T.A.; Crout, D.H. Organic Chemistry of Secondary Plant Metabolism; Freeman, Cooper & Company: San Francisco, 1969, p. 170.

- Gordo SM, Pinheiro DG, Moreira EC, Rodrigues SM, Poltronieri MC, Lemos OF, et al. (2012) Highthroughput sequencing of black pepper root transcriptome. BMC plant biology 12: 168.

- Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research 36: 3420– 3435.

- Guowei Zhang, Wenjuan Wang, Xiangmei Zhang, Qianqian Xia, Xinmei Zhao, Youngjoon Ahn, Nevin Ahmed, Andreea Cosoveanu, Mo Wang, Jialu Wang, Shaohua Shu, 2015. De Novo RNA Sequencing and Transcriptome Analysis of Colletotrichum gloeosporioides ES026 Reveal Genes Related to Biosynthesis of Huperzine A, PLoS ONE 10(3): 1-16.

- Guzman, J. D. (2014). Natural cinnamic acids, synthetic derivatives and hybrids with antimicrobial activity. *Molecules*, *19*(12), 19292-19349.

- He M, Wang Y, Hua W, Zhang Y, Wang Z. De novo sequencing of *Hypericumperforatum* transcriptome to identify potential genes involved in the biosynthesis of active metabolites. PLoS One. 2012;7: e42081.

- Hoffmann, L., Maury, S., Martz, F., Geoffroy, P., & Legrand, M. (2003). Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. *Journal of biological chemistry*, *278*(1), 95-103.

- Hu, L., Xu, Z., Wang, M., Fan, R., Yuan, D., Wu, B., Wu, H., Qin, X., Yan, L., Tan, L. and Sim, S., 2019. The chromosome-scale reference genome of black pepper provides insight into piperine biosynthesis. *Nature communications*, *10*(1), pp.1-11.

- Huapeng Sun, Fang Li, Zijian Xu, Mengli Sun, Hanqing Cong, Fei Qiao, Xiaohong Zhong, 2017. *De novo* leaf and root transcriptome analysis to identify putative genes involved in triterpenoid saponins biosynthesis in *Hedera helix* L. PLoS ONE 12(8), pp 1-17.

- Idrees, M. U. H. A. M. M. A. D., & Irshad, M. U. H. A. M. M. A. D. (2014). Molecular markers in plants for analysis of genetic diversity: a review. *European academic research*, *2*(1), 1513-1540.

- J. S. C. Smith, S. Kresovich, M. S. Hopkins, S. E. Mitchell, R. E. Dean, W. L. Woodman, M. Lee, and K. Porter. Genetic Diversity among Elite Sorghum Inbred Lines Assessed with Simple Sequence Repeats; Published in Crop Sci. 40:226–232 (2000).

- Jaramillo MA, Manos PS: Phylogeny and patterns of floral diversity in the genus Piper (Piperaceae). Am J Bot 2001, 88(4):706–716.

- Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al: Ancestral polyploidy in seed plants and angiosperms. Nature 2011, 473(7345):97–100.

- Jillian M. Hagel, Jeremy S. Morris, Eun-Jeong Lee, Isabel Desgagné-Penix, Crystal D. Bross, Limei Chang, Xue Chen, Scott C. Farrow, Ye Zhang, Jung Soh, Christoph W. Sensen and Peter J. Facchini. Transcriptome analysis of 20 taxonomically related benzylisoquinoline alkaloid-producing plants. BMC Plant Biology (2015) 15:227, Pp: 1-16.

- Jin, Z., Wungsintaweekul, J., Kim, S. H., Kim, J. H., Shin, Y., Ro, D. K., & Kim, S. U. (2020). 4-Coumarate: coenzyme A ligase isoform 3 from Piper nigrum (Pn4CL3) catalyzes the CoA thioester formation of 3, 4-methylenedioxycinnamic and piperic acids. *Biochemical Journal*, *477*(1), 61-74.

- Jing-Jing Zhang, He Su, Lei Zhang, Bao-Sheng Liao, Shui-Ming Xiao, Lin-Lin Dong, Zhi-Gang Hu, Ping Wang 1 , Xi-Wen Li, Zhi-Hai Huang, Zhi-Ming Gao,Lian-Juan Zhang, Liang Shen, Rui-Yang Cheng, Jiang Xu and Shi-Lin Chen.Comprehensive Characterization for Ginsenosides Biosynthesis in Ginseng Root by Integration Analysis of Chemical and Transcriptome. Molecules 2017, 22, 889; pp:1-13.

- Jose J, Sharma AK (1985) Structure and behaviour of chromosomes in Piper and Peperomia (family Piperaceae). Cytologia 50: 301–310.

- Joy N, Abraham Z, Soniya EV: A preliminary assessment of genetic relationships among agronomically important cultivars of black pepper. BMC Genet 2007, 8:42.

- Joy N, Asha S, Mallika V, Soniya EV (2013) De novo Transcriptome Sequencing Reveals a Considerable Bias in the Incidence of Simple Sequence Repeats towards the Downstream of 'Pre-miRNAs' of Black Pepper. PLoS One 8: e56694.

- Jutta G. Geisler and Georg G. Gross. The biosynthesis of piperine in *Piper nigrum.* Phytochemistry, Volume 29, Issue 2, 1990, Pages 489-492.

- Kaijie Xu, Fengli Sun, Guaiqiang Chai, Yongfeng Wang, Lili Shi, Shudong Liu and Yajun Xi. De novo assembly and transcriptome analysis of two contrary tillering mutants to learn

the mechanisms of tillers outgrowth in switchgrass (Panicumvirgatum L.). Frontiers in Plant Science | www.frontiersin.org, September 2015 | Volume 6 | Article 749, pp:1-10.

- Kalra S, Puniya BL, Kulshreshtha D, Kumar S, Kaur J, Ramachandran S, et al. De novo transcriptome sequencing reveals important molecular networks and metabolic pathways of the plant, *Chlorophytumborivilianum*. PLoS One. 2013;8: e83336.

- Kazue Ishihara, Eric K.W. Lee, Isabel Rushanaedy and Dulal Borthakur, 2015. Illumina-Based *De Novo* Transcriptome Analysis and Identifications of Genes Involved in the Monolignol Biosynthesis Pathway in *Acacia koa. American Journal of Bioinformatics,* 2015, 4 (1): 7.27.

- Kim, S., Park, M., Yeom, S.I. et al. (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. Nature, 46, 270–279.

- KirtiShitiz, Neha Sharma, Tarun Pal, HemantSood, Rajinder S. Chauhan. NGS Transcriptomes and Enzyme Inhibitors Unravel Complexity of Picrosides Biosynthesis in Picrorhizakurroa Royle ex. Benth. PLOS ONE, December 11, 2015. Pp: 1-19.

- Krogan, N. T., Hogan, K. & Long, J. A. APETALA2 negatively regulates multiple floral organ identity genes in *Arabidopsis* by recruiting the co-repressor TOPLESS and the histone deacetylase HDA19. *Development* 139, 4180–4190 (2012).

- Legrand, G., Delporte, M., Khelifi, C., Harant, A., Vuylsteker, C., Mörchen, M., Hance, P., Hilbert, J.L. and Gagneul, D., 2016. Identification and characterization of five BAHD acyltransferases involved in hydroxycinnamoyl ester metabolism in chicory. *Frontiers in Plant Science*, *7*, p.741.

- Lei, X., Xia, J., Wang, J. and Zheng, L., 2018. Comparative Transcriptome Analysis Identifies Genes Putatively Involved in 20-Hydroxyecdysone Biosynthesis in Cyanotis arachnoidea. *International journal of molecular sciences*, *19*(7), p.1885.

- Lei, X., Xia, J., Wang, J. and Zheng, L., 2018. Comparative Transcriptome Analysis Identifies Genes Putatively Involved in 20-Hydroxyecdysone Biosynthesis in Cyanotis arachnoidea. *International journal of molecular sciences*, *19*(7), p.1885.

- Lenhard, M., Bohnert, A., Jürgens, G. & Laux, T. Termination of stem cell maintenance in *Arabidopsis* floral meristems by interactions between *WUSCHEL* and *AGAMOUS*. *Cell* 105, 805–814, https://doi.org/10.1016/S0092-8674(01)00390-7 (2001).

- Li C, Zhu Y, Guo X, Sun C, Luo H, Song J, et al. Transcriptome analysis reveals ginsenosides biosynthetic genes, microRNAs and simple sequence repeats in *Panax ginseng* C. A. Meyer. BMC Genomics. 2013;14: 245.

- Li J, Jia D, Chen X. HUA1: a regulator of stamen and carpel identities in Arabidopsis, codes for anuclear RNA binding protein. Plant Cell. 2001; 13: 2269–2281. PMID: 11595801.

- Li Li, Mimi Li, Xiwu Qi, Xingli Tang and Yifeng Zhou, 2018. De novo transcriptome sequencing and analysis of genes related to salt stress response in Glehnia littoralis. PeerJ 6:e5681, 1-22.

- Liljegren,S.J.,Gustafson-Brown,C.,Pinyopich,A.,Ditta,G.S.,andYanofsky,M.F. (1999). Interactions among APETALA1, LEAFY ,and TERMINALFLOWER1 specifymeristemfate. *PlantCell* 11, 1007–1018.doi:10.1105/tpc.11.6.1007.

- Lisong Hu, Chaoyun Hao, Rui Fan, Baoduo Wu, Lehe Tan, Huasong Wu. *De Novo* Assembly and Characterization of Fruit Transcriptome in Black Pepper (*Piper nigrum*). PLOS ONE, 2015. Pp:1-14.

- Liu, S. S., Chen, J., Li, S. C., Zeng, X., Meng, Z. X., & Guo, S. X. (2015). Comparative transcriptome analysis of genes involved in GA-GID1-DELLA regulatory module in symbiotic and asymbiotic seed germination of Anoectochilus roxburghii (Wall.) Lindl.(Orchidaceae). *International journal of molecular sciences*, *16*(12), 30190-30203.

- Liu, X. et al. The genome of medicinal plant Macleaya cordata provides new insights into benzylisoquinoline alkaloids metabolism. Mol. Plant 10, 975–989 (2017).

- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-timequantitative PCR and the 2(−Delta Delta C(T)) method. Methods 25:402–408.

- Ma, L., Ma, S.W., Deng, Q., Yuan, Y., Wei, Z., Jia, H. and Ma, Z., 2018. Identification of Wheat Inflorescence Development-Related Genes Using a Comparative Transcriptomics Approach. *International journal of genomics*, *2018*.

- Madhumati, B. (2014). Potential and application of molecular markers techniques for plant genome analysis. *Int. J. Pure App. Biosci*, *2*(1), 169-188.

- Mami Yamazaki, Keiichi Mochida, Takashi Asano, Ryo Nakabayashi, Motoaki Chiba, Nirin Udomson, Yasuyo Yamazaki, Dayan B. Goodenowe, Ushio Sankawa, Takuhiro Yoshida, Atsushi Toyoda, Yasushi Totoki, Yoshiyuki Sakaki, Elsa Gongora-Castillo, C. Robin Buell, Tetsuya Sakurai and Kazuki Saito. Coupling Deep Transcriptome Analysis with Untargeted Metabolic Profiling in Ophiorrhizapumila to Further the Understanding of the Biosynthesis of the Anti-Cancer Alkaloid Camptothecin and Anthraquinones. Plant Cell Physiol. 2013;54: 686–96.

- Manoj P, Soniya EV, Banerjee NS and Ravichandran P (2004) Recent studies on well-known spice, Piper longum L. Nat Prod Rad.3(4):222-227.

- Martin JA, Wang Z (2011) Next-generation transcriptome assembly. Nature Reviews Genetics 12: 671–682.

- Mei Xiao, Ye Zhang, Xue Chen, Eun-Jeong Lee, Carla J.S. Barber, Romit Chakrabarty, Isabel Desgagné-Penix, Tegan M. Haslam, Yeon-Bok Kim, Enwu Liu, Gillian MacNevin, Sayaka Masada-Atsumi, Darwin W. Reed, Jake M. Stout, Philipp Zerbe, Yansheng Zhang, Joerg Bohlmann, Patrick S. Covello, Vincenzo De Luca, Jonathan E. Page, Dae-Kyun Ro, Vincent J.J. Marting, Peter J. Facchini, Christoph W. Sensen (2013) Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. Journal of Biotechnology 166 (2013) 122– 134.

- Menezes IC, Cidade FW, Souza AP, Sampaio IC: Isolation and characterization of microsatellite loci in the black pepper, Piper nigrum L. (piperaceae). Conserv Genet Resour 2009, 1(1):209–212.

- Michael Ashburner, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, Kara Dolinski, Selina S. Dwight, Janan T. Eppig, Midori A. Harris, David P. Hill, Laurie Issel-Tarver, Andrew Kasarskis, Suzanna Lewis, John C. Matese, Joel E. Richardson, Martin Ringwald, Gerald M. Rubin and Gavin Sherlock. Gene Ontology: tool for the unification of biology. Nat Genet. 2000 May; 25(1): 25–29.

- Ming-Hua Liu, Bin-Rui Yang, Wai-Fung Cheung, Kevin Yi Yang, He-Feng Zhou, Jamie Sui-Lam Kwok, Guo-Cheng Liu, Xiao-Feng Li, Silin Zhong, Simon Ming-Yuen Lee and Stephen Kwok-Wing Tsui, 2015. Transcriptome analysis of leaves, roots and flowers of Panax notoginseng identifies genes involved in ginsenoside and alkaloid biosynthesis, BMC Genomics (2015) 16:265, 1-12.

- Mrinalini Prasad, Anjali Singh, Arti Yadav, Teg Bahadur Singh, Deepika Goyal, Preksha Shrivastav and Prem Kumar Dantu, 2018. Analysis for piperine in leaves, roots and spikes in *Piper longum* L. Journal of Pharmacognosy and Phytochemistry 2018; 7(6): 1008-1010.

- Mrinalini Prasad, Anjali Singh, Arti Yadav, Teg Bahadur Singh, Deepika Goyal, Preksha Shrivastav and Prem Kumar Dantu (2018) Analysis for piperine in leaves, roots and spikes in *Piper longum* L. JPP 2018; 7(6): 1008-1010.

- Mukesh Jain (2011) A next-generation approach to the characterization of a non-model plant transcriptome. CURRENT SCIENCE, VOL. 101, NO. 11, 10. Pp: 1435-1439.

- Nagaraja Reddy Rama Reddy, RuchaHarishbhai Mehta, PalakHarendrabhaiSoni, JayantiMakasana, NarendraAthamaramGajbhiye, ManivelPonnuchamy, Jitendra Kumar. Next Generation Sequencing and Transcriptome Analysis Predicts Biosynthetic Pathway of Sennosides from Senna (*Cassia angustifolia* Vahl.), a Non-Model Plant with Potent Laxative Properties, PLOS ONE, 2015, pp: 1-32.

- Nair KKN (2000) Manual of nonwood forest produce plants of Kerala. Kerala Forest Research Institute, Kerala. 268–270.

- Nair P, Misra A, Singh A, Shukla AK, Gupta MM, Gupta AK, et al. Differentially expressed genes during contrasting growth stages of *Artemisia annua* for artemisinin content. PLoS One. 2013;8: e60375.

- Negrel, J.; Jeandet, P. Metabolism of tyramine and feruloyltyramine in TMV inoculated leaves of Nicotiana tabacum. Phytochemistry, 1987, 26, 2185-2190. [http://dx.doi.org/10.1016/S0031-9422(00)84681-6]

- Negrel, J.; Martin, C. The biosynthesis of feruloyltyramine in Nicotiana tabacum. Phytochemistry, 1984, 23, 2797-2801. [http://dx.doi.org/10.1016/0031-9422(84)83018-6]

- Okwute SK, Egharevba HO (2013) Piperine-Type Amides: Review of the Chemical and Biological Characteristics. International Journal of Chemistry 5: 99–122.

- Paterson AH: Leafing through the genomes of our major crop plants: strategies for capturing unique information. Nat Rev Genet 2006, 7(3):174–184.

- Peláez, P., Orona-Tamayo, D., Montes-Hernández, S., Valverde, M.E., Paredes-López, O. and Cibrián-Jaramillo, A., 2019. Comparative transcriptome analysis of cultivated and wild seeds of Salvia hispanica (chia). *Scientific reports*, *9*(1), p.9761.

- Peng, M., Hudson, D., Schofield, A., Tsao, R., Yang, R., Gu, H., Bi, Y.M. and Rothstein, S.J., 2008. Adaptation of Arabidopsis to nitrogen limitation involves induction of anthocyanin synthesis which is controlled by the NLA gene. *Journal of experimental botany*, *59*(11), pp.2933-2944.

- Pourmazaheri, H., Soorni, A., Kohnerouz, B.B., Dehaghi, N.K., Kalantar, E., Omidi, M. and Naghavi, M.R., 2019. Comparative analysis of the root and leaf transcriptomes in Chelidonium majus L. *PloS one*, *14*(4), p.e0215165.

- Qionglin HUANG, Lingling LIANG, Rui HE, Xinye MA, Ruoting ZHAN, Weiwen CHEN, 2015. The first insight into transcriptome profile of herbaceous plant Nervilia fordii based on RNA-seq, POJ 8(6):493-499.

- R Ranjan, S Patro, B Pradhan, A Kumar, IB Maiti, N Dey. Development and functional analysis of novel genetic promoters using DNA shuffling, hybridization and a combination thereof, PLoS One 7 (3), e31931. 2012.

- Rami E, Sipai S and Patel I (2013) Studies on qualitative and quantitative phytochemical, analysis of *Piper longum*Linn, *Int J Pharma Bio Sci.* 4(3): (B) 1381 – 1388.

- Ramilowski JA, Sawai S, Seki H, Mochida K, Yoshida T, Sakurai T, et al. *Glycyrrhizauralensis* transcriptome landscape and study of phytochemicals. Plant Cell Physiol. 2013;54: 697–710.

- Ratcliffe,O.J.,Bradley,D.J.,andCoen,E.S.(1999).Separationofshootandfloral   identityin *Arabidopsis*. *Development* 126, 1109–1120.

- Reddy, N. R. R., Mehta, R. H., Soni, P. H., Makasana, J., Gajbhiye, N. A., Ponnuchamy, M., & Kumar, J. (2015). Next generation sequencing and transcriptome analysis predicts biosynthetic pathway of sennosides from Senna (Cassia angustifolia Vahl.), a non-model plant with potent laxative properties. *PLoS One*, *10*(6), e0129422.

- Reddy, N.R.R., Mehta, R.H., Soni, P.H., Makasana, J., Gajbhiye, N.A., Ponnuchamy, M. and Kumar, J., 2015. Next generation sequencing and transcriptome analysis predicts biosynthetic pathway of sennosides from Senna (Cassia angustifolia Vahl.), a non-model plant with potent laxative properties. *PLoS One*, *10*(6), p.e0129422.

- Richet, N., Tozo, K., Afif, D., Banvoy, J., Legay, S., Dizengremel, P., & Cabané, M. (2012). The response to daylight or continuous ozone of phenylpropanoid and lignin biosynthesis pathways in poplar differs between leaves and wood. *Planta*, *236*(2), 727-737.

- S. Temnykh, W.D. Park, N. Ayres, S. Cartinhour N. Hauck, L. Lipovich, Y.G. Cho, T. Ishii S.R. McCouch. Mapping and genome organization of microsatellite sequences in rice (Oryza sativa L.). TheorAppl Genet (2000) 100:697–712.

- Schnabel, A., Cotinguiba, F., Athmer, B., Yang, C., Westermann, B., Schaks, A., Porzel, A., Brandt, W., Schumacher, F. and Vogt, T., 2020. A piperic acid CoA ligase produces a putative precursor of piperine, the pungent principle from black pepper fruits. *The Plant Journal*, *102*(3), pp.569-581.

- Schultz, E.A. and Haughn, G.W., 1991. LEAFY, a homeotic gene that regulates inflorescence development in Arabidopsis. *The Plant Cell*, *3*(8), pp.771-781.

- Semler, U.; Gross, G.G. Distribution of piperine in vegetative parts of Piper nigrum. Phytochemistry, 1988, 27, 1566. [http://dx.doi.org/10.1016/0031-9422(88)80249-8].

- Semler, U.; Schmidtberg, G.; Gross, G.G. Synthesis of piperoyl-CoA-thioester. Z. Naturforsch., B. J. Chem. Sci., 1897, 42c, 1070-1074.

- Sen S, Skaria R, Abdul Muneer PM: Genetic diversity analysis in Piper species (Piperaceae) using RAPD markers. MolBiotechnol 2010, 46(1):72–79.

- Senthil K, Wasnik NG, Kim YJ, Yang DC. Generation and analysis of expressed sequence tags from leaf and root of *Withaniasomnifera* (Ashwgandha). MolBiol Rep. 2010;37: 893–902.

- Shannon,S.,andMeeks-Wagner,D.R.(1993).Geneticinteractionsthat regulateinflorescencedevelopmentin *Arabidopsis*. *PlantCell* 5, 639–655.doi: 10.1105/tpc.5.6.639.

- Sharma PV: In: CharakaSamhita (1996) Chaukamba Publications. ChikitsaStana (English), pp. 434–47.

- Sharma V, Renuka K, Palak V, Harisha CR and Prajapati PK (2012) Pharmacognostical and phytochemical study of Piper longum L. and Piper retrofractum VAHL.

- Simon, K.O.; Henry, O.; Piperine-Type Amides, E. Review of the Chemical and Biological Characteristics. Int. J. Chem., 2013, 5(3), 99-122.

- Singh M (2012) Comparative phytochemical & antioxidant study of aqueous extracts of *Glycyrrhizaglabra*(mulethi) &*Piper longum*(long pepper). *Int. J. Drug Res. Tech*. 2(2): 203-207.

- Soetaert SS, Van Neste CM, Vandewoestyne ML, Head SR, Goossens A, Van Nieuwerburgh FC, et al. Differential transcriptome analysis of glandular and filamentous trichomes in *Artemisia annua*. BMC Plant Biol. 2013;13: 220.

- Stewart, C. Jr, Kang, B.-C., Liu, K., Mazourek, M., Moore, S.L., Yoo, E.J., Kim, B.-D., Paran, I. and Jahn, M.M. (2005) The Pun1 gene for pungency in pepper encodes a putative acyltransferase. Plant J. 42, 675–688.

- Sui, C., Zhang, J., Wei, J., Chen, S., Li, Y., Xu, J., Jin, Y., Xie, C., Gao, Z., Chen, H. and Yang, C., 2011. Transcriptome analysis of Bupleurum chinense focusing on genes involved in the biosynthesis of saikosaponins. *BMC genomics*, *12*(1), p.539.

- Szőke, É., Lemberkovics, É. and Kursinszki, L., 2013. Alkaloids derived from lysine: piperidine alkaloids. *Natural Products: Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes*, pp.303-341.

- Takatsuji H. Zinc-finger transcription factors in plants. Cell Mol Life Sci. 1998; 54: 582–596.

- Takatsuji H. Zinc-finger transcription factors in plants. Cell Mol Life Sci. 1998; 54: 582–596. PMID:9676577

- The Gene Ontology Consortium (January 2008). "The Gene Ontology project in 2008". *Nucleic Acids Research*. 36 (Database issue): D440–4. PMC 2238979.

- Thomas L Parchman, Katherine S Geist, Johan A Grahnen, Craig W Benkman, C Alex Buerkle. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. BMC Genomics 2010, 11:180. Pp:1-16.

- Thompson MJ, Lai WS, Taylor GA, Blackshear PJ. Cloning and characterization of two yeast genesencoding members of the CCCH class of zinc finger proteins: zinc finger-mediated impairment of cellgrowth. Gene. 1996; 174: 225–233. PMID: 8890739

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

- Vasavirama K, UPENDER M (2014) PIPERINE: A VALUABLE ALKALOID FROM PIPER SPECIES. International Journal of Pharmacy & Pharmaceutical Sciences 6: 34–38.

- Wang L, Feng Z, Wang X, Wang X, Zhang X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics. 2010; 26: 136–138.

- Wang L, Li P, Brutnell TP: Exploring plant transcriptomes using ultra high-throughput sequencing. Brief Funct Genomic Proteomic2010, 9(2):118–128.

- WANG Shou-ming, LI Wei-jia, LIU Yue-xue, LI He, MA Yue, ZHANG Zhi-hong, 2017. Comparative transcriptome analysis of shortened fruit mutant in woodland strawberry (*Fragaria vesca*) using RNA-Seq, *Journal of Integrative Agriculture 2017, 16(4): 828–844*.

- Warrier PK, Nambiar VP and Raman KC (1995) Madras, India: Orient Longman Ltd. *Piper longum*. *Indian medicinal Plants*. 4:290.

- Wink M (1987) Quinolizidine alkaloids: biochemistry, metabolism, and function in plants and cell suspension cultures. Planta Med 53: 509–514. PMID: 17269092.

- Wink M, Hartmann T (1979) Cadaverine-pyruvate transamination: The principal step of enzymatic quinolizidine alkaloid biosynthesis in *Lupinus polyphyllus* cell suspension cultures. FEBS letters 101: 343–346. PMID: 446758.

- Xin Huang, Xiao-Guang Chen and Peter A. Armbruster. Comparative performance of transcriptome assembly methods for non-model Organisms. BMC Genomics (2016) 17:523.

- Xiu Yun Lei, Jing Xia, JianWen Wang and Li Ping Zheng, 2018. Comparative Transcriptome Analysis Identifies Genes Putatively Involved in 20-Hydroxyecdysone Biosynthesis in Cyanotis arachnoidea. Int. J. Mol. Sci. 2018, 19, 1885, 1-14.

- Yue Liu, Yi Wang, Fengxian Guo, Lin Zhan, Toni Mohr, Prisca Cheng, Naxin Huo, Ronghui Gu, Danning Pei, Jiaqing Sun, Li Tang, Chunlin Long, Luqi Huang & Yong Q. Gu, 2017. Deep sequencing and transcriptome analyses to identify genes involved in secoiridoid biosynthesis in the Tibetan medicinal plant *Swertia mussotii*. Scientific Reports | 7:43108, 1-14.

- Yuki Moriya, Masumi Itoh, Shujiro Okuda, Akiyasu C. Yoshizawa and Minoru Kanehisa*. KAAS: an automatic genome annotation and pathway reconstruction server. W182–W185 Nucleic Acids Research, 2007, Vol. 35, pp: w182-w185.

- Zaveri M, Khandhar A, Patel S and Patel A (2010) Chemistry and pharmacology of *Piper longum*Lin. 5(1):67-76.

- Zhang QQ, Liu W, Cai YL, Lan AF, Bian YB (2018) Validation of internal control genes for quantitative real-time PCR gene expression analysis in *Morchella*. Molecules 23(9):2331.

- Zhang, D., Kong, W., Robertson, J., Goff, V.H., Epps, E., Kerr, A., Mills, G., Cromwell, J., Lugin, Y., Phillips, C. and Paterson, A.H., 2015. Genetic analysis of inflorescence and plant height components in sorghum (Panicoidae) and comparative genetics with rice (Oryzoidae). *BMC plant biology*, *15*(1), p.107.

- Zhang, G. H., Ma, C. H., Zhang, J. J., Chen, J. W., Tang, Q. Y., He, M. H., ... & Yang, S. C. (2015). Transcriptome analysis of Panax vietnamensis var. fuscidicus discovers putative ocotillol-type ginsenosides biosynthesis genes and genetic markers. *BMC genomics*, *16*(1), 159.

- Zhang, X., Allan, A., Li, C., Wang, Y. and Yao, Q., 2015. De novo assembly and characterization of the transcriptome of the Chinese medicinal herb, Gentiana rigescens. *International journal of molecular sciences*, *16*(5), pp.11550-11573.

- Zhou, W., Chen, Q., Wang, X.B., Hughes, T., Liu, J.J. and Zhang, X., 2019. De novo assembly of the Platycladus orientalis (L.) Franco transcriptome provides insight into the development and pollination mechanism of female cone based on RNA-Seq data. *Scientific reports*, *9*(1), p.10191.

- Zhou, W., Chen, Q., Wang, X.B., Hughes, T., Liu, J.J. and Zhang, X., 2019. De novo assembly of the Platycladus orientalis (L.) Franco transcriptome provides insight into the development and pollination mechanism of female cone based on RNA-Seq data. *Scientific reports*, *9*(1), p.10191.
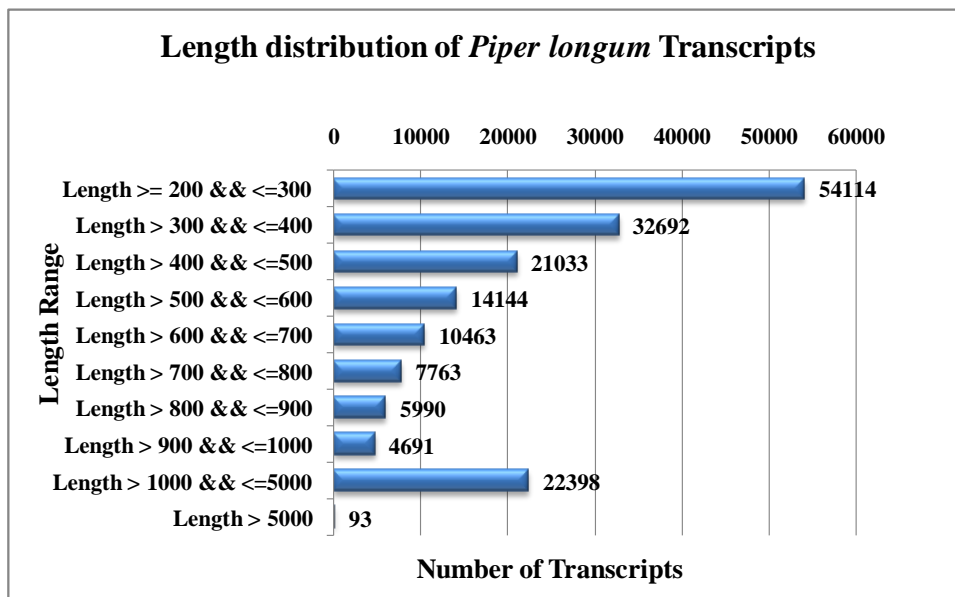
Figures and Tables

Figure 1: Length distribution of *Piper longum* transcripts
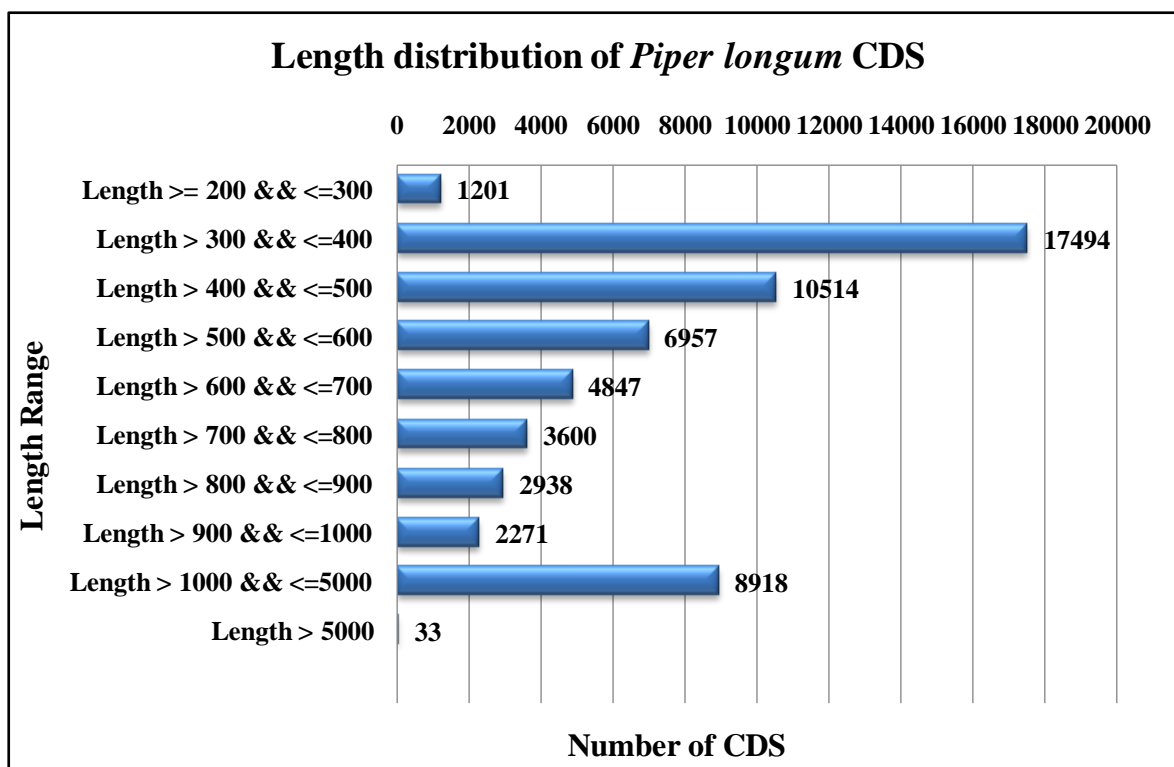
Figure 2: Length distribution of CDS of *Piper longum* sample
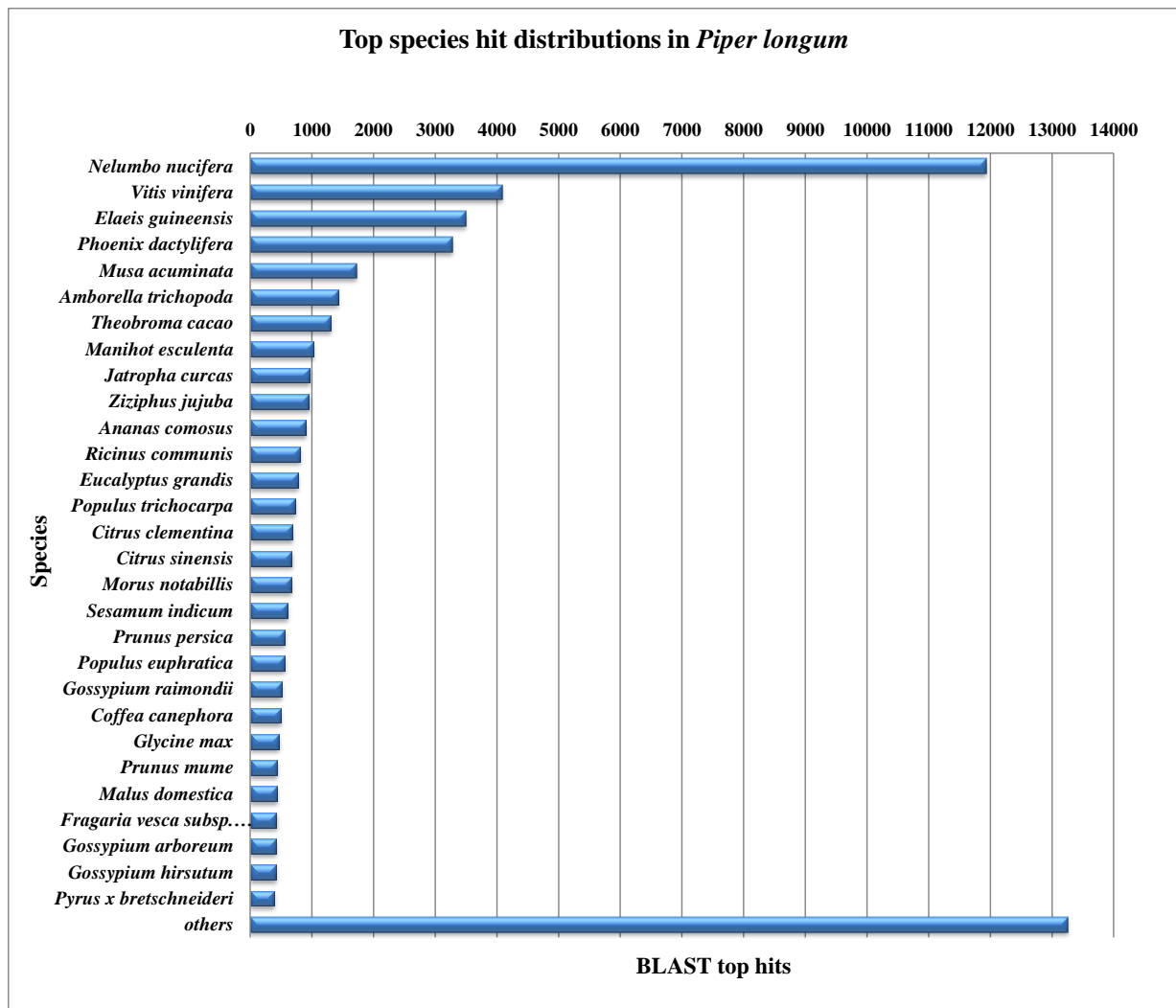
Figure 3: Top species hit distributions of *Piper longum* CDS
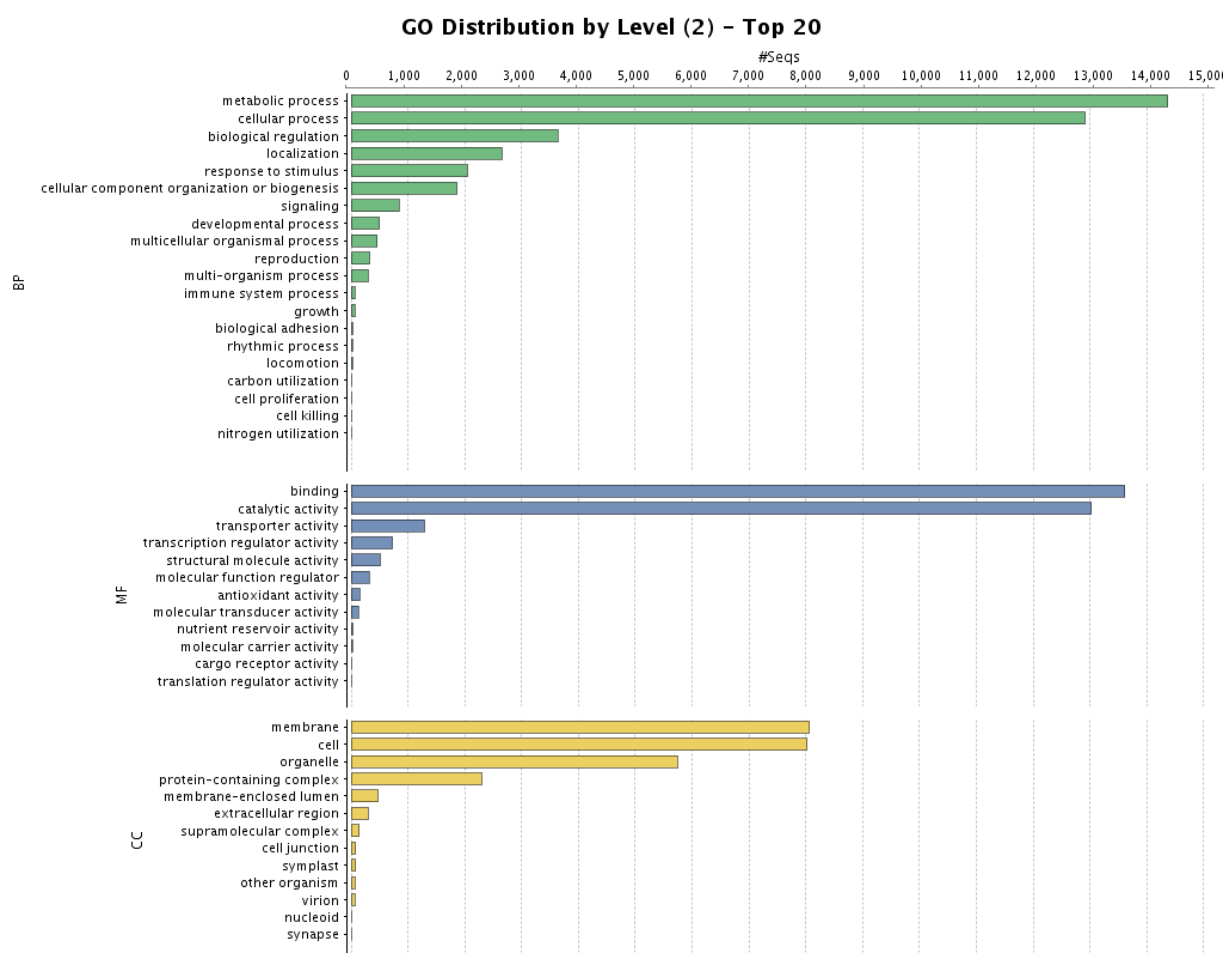
Figure 4: Gene ontology term (Biological process, Cellular Component and Molecular Function) distribution in *Piper longum* CDS

Figure 5: KEGG functional pathway of *Piper longum*
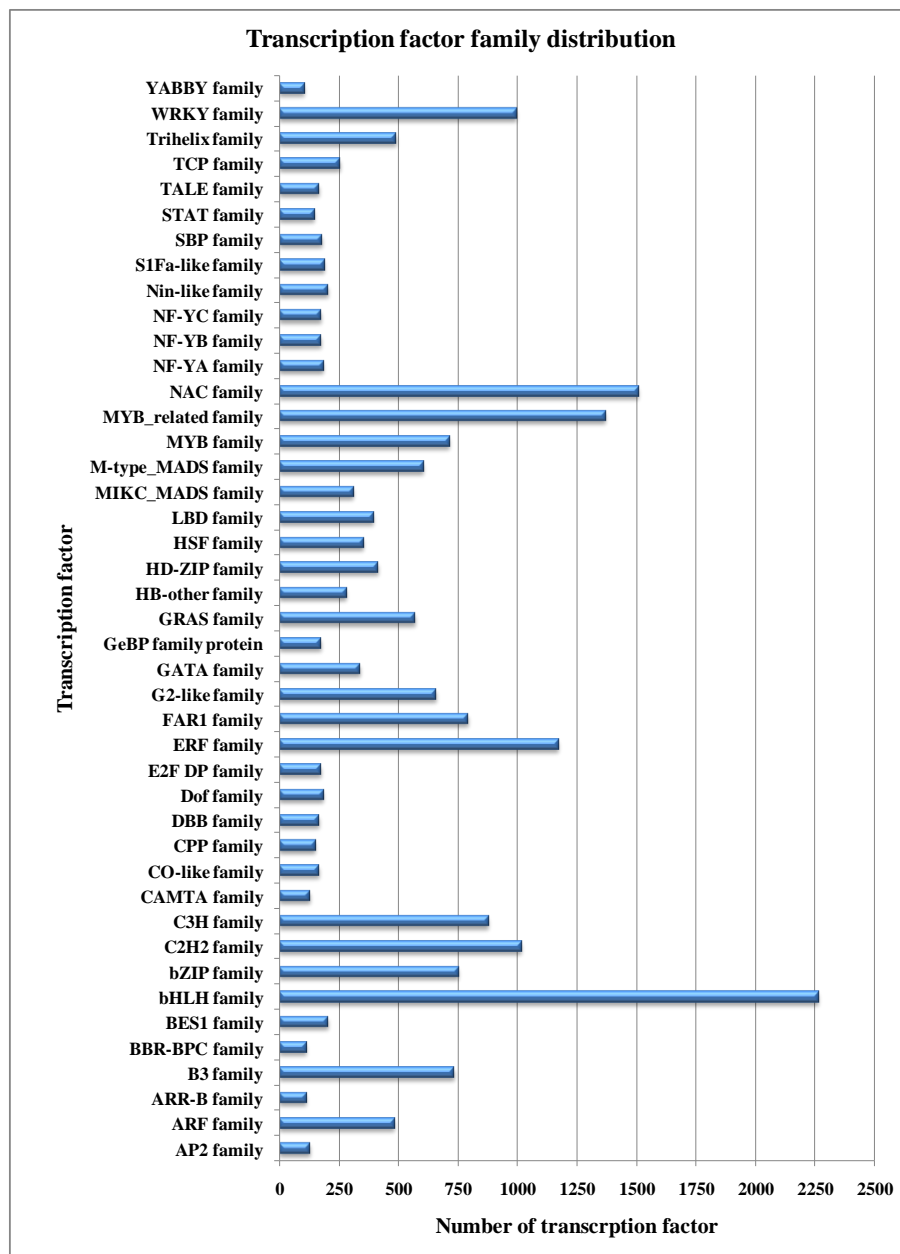
Figure 6: Distribution of different transcription factors in CDS of *Piper longum*

26
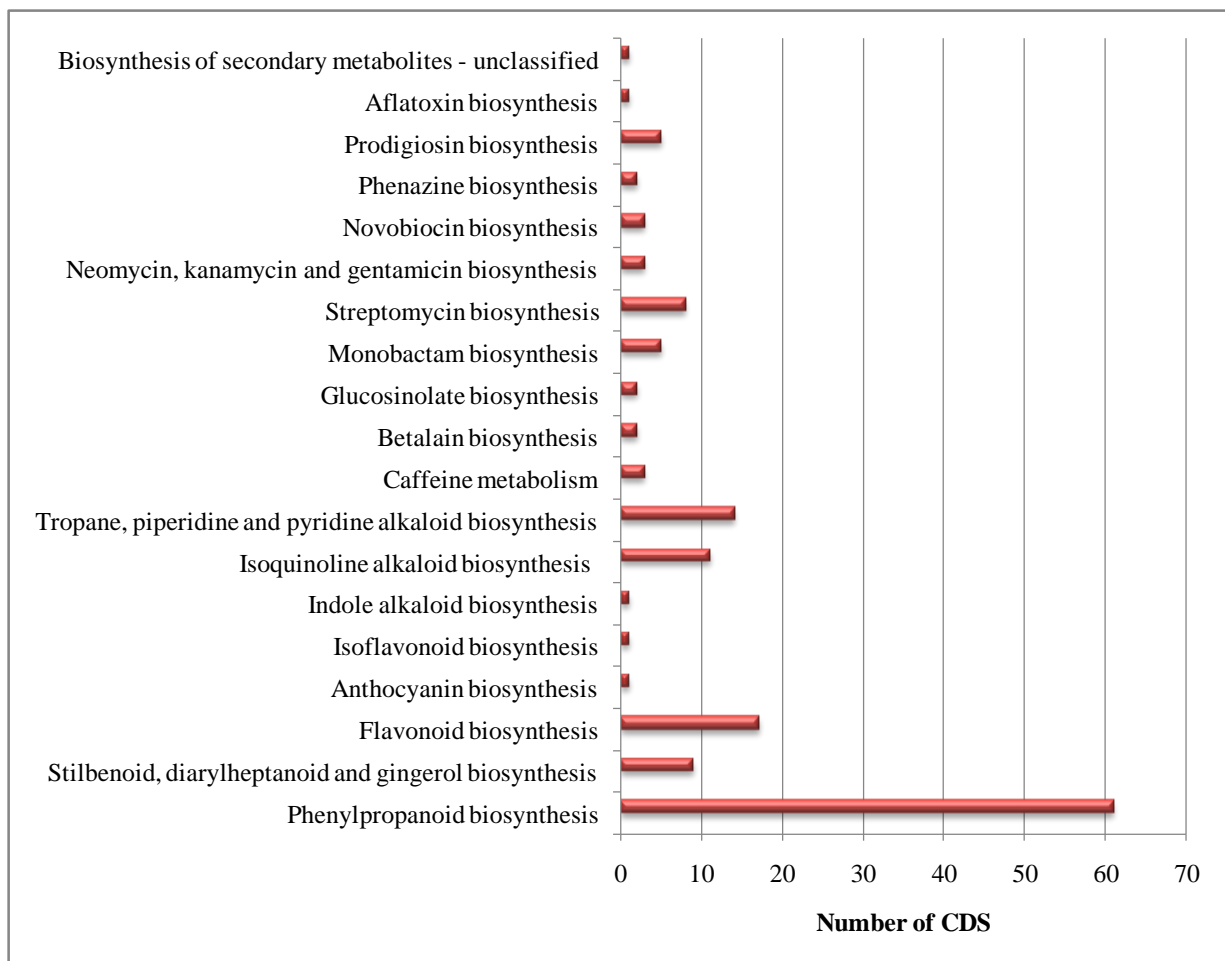
Figure 7: Classification based on categories of secondary metabolite biosynthesis
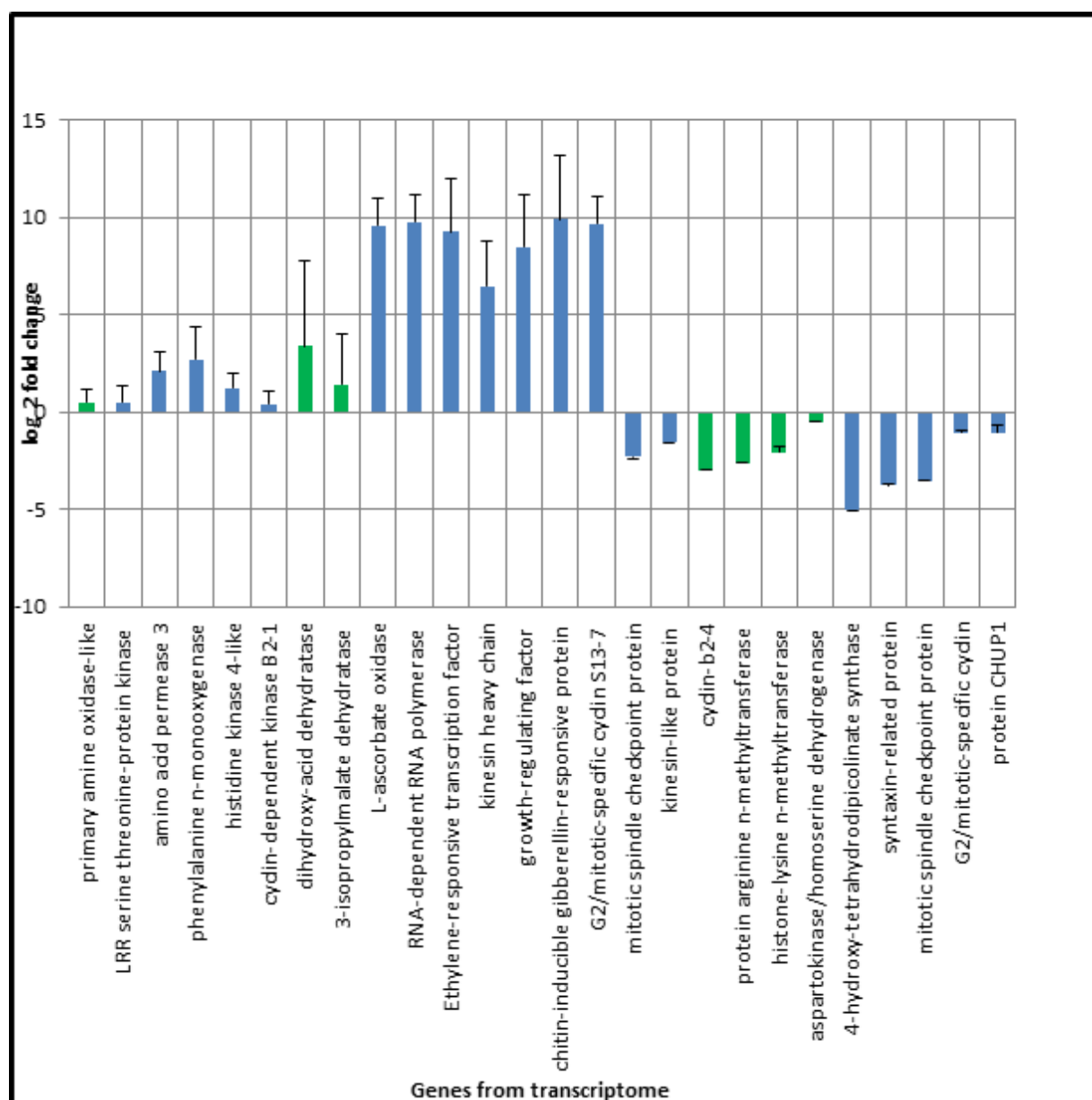
Figure 8: shows log2 fold change of selected differentially Expressed gene
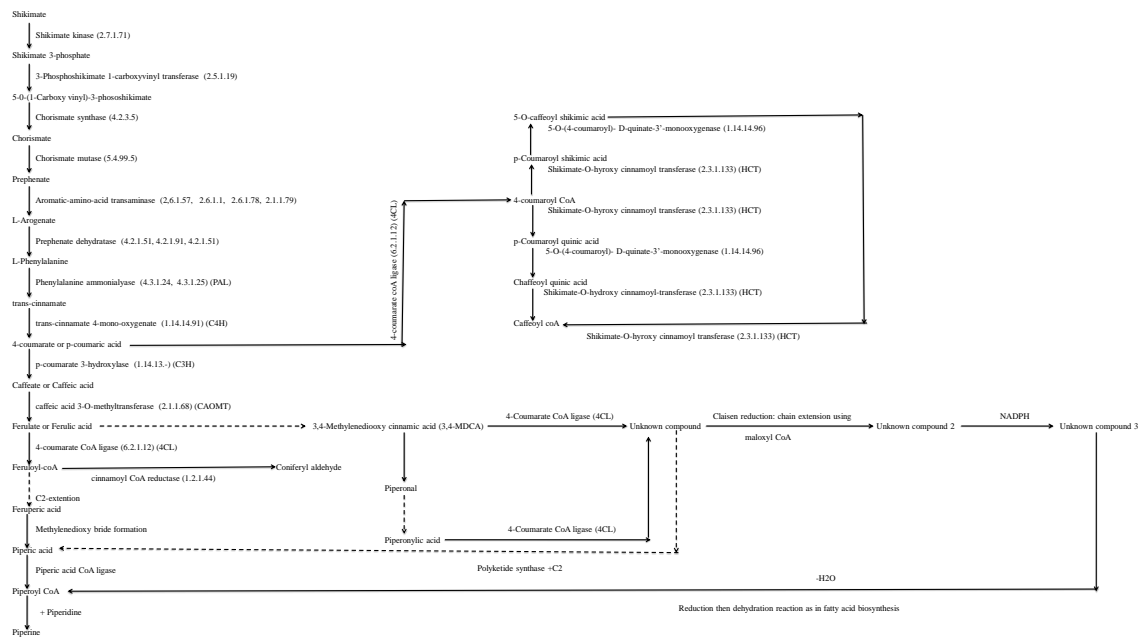
Figure 9: Piperine biosynthetic pathway

Table 1: Statistics of assembled transcript of *P. longum* sample

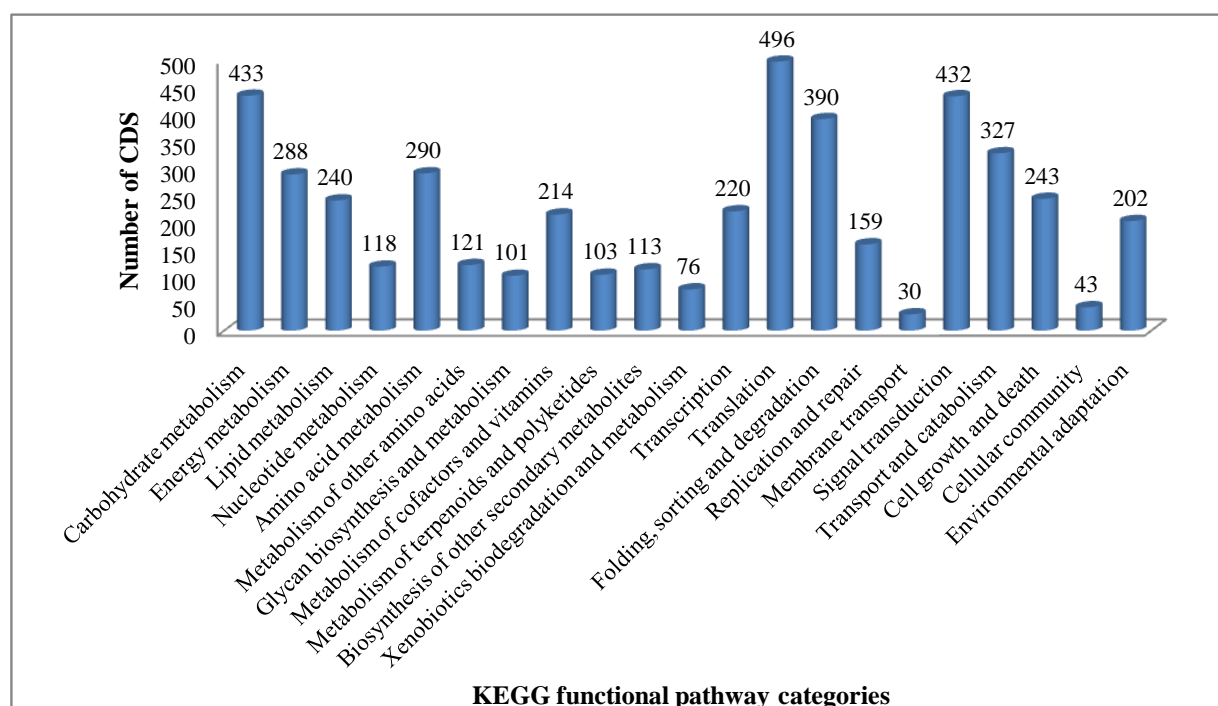| Descriptions	*P. longum* Assembly (bp)
Number of transcripts	1,73,381
Total transcriptome length	9,96,12,836
Maximum transcriptlength	12,994
Transcript N50	722

Table 2: Statistics of predicted CDS of *Piper longum* sample

Descriptions	*Piper longum* CDS (bp)
Number of CDS	58773
Size of CDS	38734722
Maximum CDS size	12990
CDS N50	747

Table 3: GO distribution statistics of the annotated CDS of *Piper longum*

Description	*Piper longum*
Biological Process	18508
Cellular Component	14429
Molecular Function	21103

Table 4: Distribution of genes expressed between *P. longum* leaf and root

29

Description	Count
Commonly expressed genes	30296
Up-regulated genes	15148
Down-regulated genes	15148
Significantly Up-regulated genes	581
Significantly Down-regulated genes	1307
Exclusive in *Piper longum* leaf	2741
Exclusive in *Piper longum* root	1595

Table 5: Distribution of genes expressed between *P. longum* leaf and spike

Description	Count
Commonly expressed genes	32216
Up-regulated genes	16108
Down-regulated genes	16108
Significantly Up-regulated genes	625
Significantly Down-regulated genes	1246
Exclusive in *Piper longum* leaf	821
Exclusive in *Piper longum* root	1033

Table 6: Distribution of genes expressed between *P. longum* root and spike

Description	Count
Commonly expressed genes	30225
Up-regulated genes	15113
Down-regulated genes	15112
Significantly Up-regulated genes	993
Significantly Down-regulated genes	878
Exclusive in *Piper longum* leaf	1667
Exclusive in *Piper longum* root	3024

Table 7: SSRs identification statistics in *Piper longum*

Description	*P. longum*
Total number of sequences examined	173381
Total size of examined sequences (bp)	99612836
Total number of identified SSRs	8041
Number of SSR containing sequences	7139
Number of sequences containing more than 1 SSR	737
Number of SSRs present in compound formation	570

Table 8: Distribution to different repeat type classes

Unit size	Number of SSRs
Dinucleotide	3209
Trinucleotide	4506
Tetranucleotide	216
Pentanucleotide	79
Hexanucleotide	31

Table 9: CDS involved in Tropane, piperidine and pyridine alkaloid biosynthesis [PATH: ko00960]

| 00960 Tropane, piperidine and pyridine alkaloid biosynthesis [PATH: ko00960] |

S. No.	CDS IDs	KO IDs	EC IDs	Biosynthetic pathway
1	cds_18211_Transcript_23293	K00276 AOC3, AOC2, tynA	primary-amine oxidase [EC:1.4.3.21]	Pyridine biosynthesis pathway
2	cds_20102_Transcript_25925	K00276 AOC3, AOC2, tynA	primary-amine oxidase [EC:1.4.3.21]	Piperidine biosyhthetic pathway
3	cds_30149_Transcript_40912	K00276 AOC3, AOC2, tynA	primary-amine oxidase [EC:1.4.3.21]	Pyridine biosynthesis pathway
4	cds_36462_Transcript_51168	K00276 AOC3, AOC2, tynA	primary-amine oxidase [EC:1.4.3.21]	Piperidine biosyhthetic pathway
5	cds_16196_Transcript_2043	K08081 TR1	tropinone reductase I [EC:1.1.1.206]	Tropane biosynthetic pathway
6	cds_56318_Transcript_925	K08081 TR1	tropinone reductase I [EC:1.1.1.206]	Tropane biosynthetic pathway
7	cds_56356_Transcript_926	K08081 TR1	tropinone reductase I [EC:1.1.1.206]	Tropane biosynthetic pathway
8	cds_16021_Transcript_2020	K14454 GOT1	aspartate aminotransferase, cytoplasmic [EC:2.6.1.1]	Tropane biosynthetic pathway
9	cds_39099_Transcript_55714	K14454 GOT1	aspartate aminotransferase, cytoplasmic [EC:2.6.1.1]	Tropane biosynthetic pathway
10	cds_42796_Transcript_6228	K15849 PAT, AAT	bifunctional aspartate aminotransferase and glutamate/aspartate-prephenate aminotransferase [EC:2.6.1.1 2.6.1.78 2.6.1.79]	Tropane biosynthetic pathway
11	cds_42801_Transcript_6229	K15849 PAT, AAT	bifunctional aspartate aminotransferase and glutamate/aspartate-prephenate aminotransferase [EC:2.6.1.1 2.6.1.78 2.6.1.79]	Tropane biosynthetic pathway

12	cds_20398_Transcript_26340	K00815 TAT	tyrosine aminotransferase [EC:2.6.1.5]	Tropane biosynthetic pathway
13	cds_22551_Transcript_29424	K00815 TAT	tyrosine aminotransferase [EC:2.6.1.5]	Tropane biosynthetic pathway
14	cds_13069_Transcript_16650	K00817 hisC	histidinol-phosphate aminotransferase [EC:2.6.1.9]	Tropane biosynthetic pathway

Table 10: List of genes differentially expressed in leaf and spike selected for validation and its differential expression in terms of log 2 fold.

S. No.	Name of Genes	Function of Genes in Leaf	Function of Genes in Spike	log2fold
1	GAPDH	Internal control gene	Internal control gene	0
2	4-hydroxy-tetrahydrodipicolinate synthase	Absent	Lysine biosynthesis	-5.058892212
3	Amino acid permease 3	Absent	Lysine biosynthesis	2.114903455
4	Aspartokinase/homoserine dehydrogenase	Aspartate pathway	Absent	-0.505489973
5	LRR receptor-like serine threonine-protein kinase	Absent	Aspartate pathway	0.504469951
6	3-isopropylmalate dehydratase	Absent	biosynthesis of leucine	1.423867816
7	Phenylalanine n-monooxygenase	Absent	Valine, leucine and isoleucine pathway	2.715696386
8	Histidine kinase 4-like	Absent	biosynthesis of leucine	1.23091656
9	Dihydroxy-acid dehydratase	Absent	Valine, leucine and isoleucine pathway	3.408627507
10	Primary amine oxidase-like	Piperidine Pathway related enzyme	Absent	0.479662132
11	Protein arginine n-methyltransferase	Regulating flowering time	Absent	-2.610880587
12	Histone-lysine n-methyltransferase	Regulating flowering time	Absent	-2.040028228
13	L-ascorbate oxidase	Absent	Ascorbate metabolism	9.574864201
14	RNA-dependent RNA polymerase	Absent	Replication of RNA from an RNA template.	9.766782064
15	Ethylene-responsive transcription factor	Absent	Stress signal transduction pathways.	9.234066916
16	Kinesin heavy chain	Absent	Calmodulin-binding protein	6.475303198
17	Growth-regulating factor	Absent	Regulating flowering time	8.460050976
18	Chitin-inducible gibberellin-responsive protein	Absent	Regulating flowering time	9.902196359

19	Cyclin-dependent kinase B2-1	Absent	Ascorbate metabolism	0.389143966
20	Kinesin-like protein	Replication of RNA from an RNA template.	Absent	-1.568212769
21	Cyclin-b2-4-like	Stress signal transduction pathways.	Absent	-2.973046787
22	G2/mitotic-specific cyclin S13-7	Calmodulin-binding protein	Absent	9.662897235
23	Mitotic spindle checkpoint protein	Microtubule generation, bipolarity establishment, length control and chromosome alignment	Absent	-3.512551913
24	Syntaxin-related protein	Syntaxin localized at the plasma membrane and cell signaling	Absent	-3.749175868
25	Mitotic spindle checkpoint protein	Microtubule generation, bipolarity establishment, length control and chromosome alignment	Absent	-3.512551913
26	G2/mitotic-specific cyclin	Accumulate during G2 and destroyed at mitosis	Absent	-1.067365167
27	Protein CHUP1	For actin polymerization and chloroplast relocation movement	Absent	-1.058641104

Table 11: List of selected primers for validation of Expression of Differentially Expressed gene

S. No.	Gene	Primers (Both forward and reverse)	Amplicon length
1	primary amine oxidase-like	Forward: GTAGGGAAGGTGGAGGTAGAA	103
		Reverse: TTTGAAGACGATGGAGACGAAG	
2	LRR receptor-like serine threonine-protein kinase	Forward: CCCAAACTCAAGGAGAAGAAGA	124
		Reverse: TGTGAAGAAGACTGCCATACC	
3	amino acid permease 3	Forward: CCAGAGTCGCCTTCTTCATT	101
		Reverse: CATTGCCTTTGCCTACTCATATTC	
4	phenylalanine n-monooxygenase	Forward: AGTGTCTAACTGCTGCTCTAAC	85
		Reverse: GACAACCTCGTTCGATTCCT	
5	histidine kinase 4-like	Forward: CCAGCTGCAACTTTACGATTTAC	118
		Reverse:	

33

		ATGGAAAGGAGCAAGGTAAGG	
6	cyclin-dependent kinase B2-1	Forward: GATCAGCGCAAAGAAAGCTATG	105
		Reverse: GGAAAGCAGTAGGTATCCAAAGA	
7	kinesin-like protein	Forward: CTCCCGCTCTTTCTCCTTTC	139
		Reverse: CCGGAAGATCATGGAAGGTATT	
8	cyclin-b2-4-like	Forward: CTGAACTACCCACCTTCACTTT	100
		Reverse: TGTGCCATTCGCTAGTTCTC	
9	protein arginine n-methyltransferase	Forward: CAAGGAAATGGTGCTGGTTG	114
		Reverse: CCAACTACTTCTGCACCTACTC	
10	histone-lysine n-methyltransferase	Forward: GGACAGTGTTGGGAGAGATTAG	110
		Reverse: AACTTTGGCTCCTCCATACC	
11	bifunctionalaspartokinase/homoserine dehydrogenase	Forward: CAGAACCACAGCCTCACTAAC	108
		Reverse: GCAGCTATAATGGGTGCTCTT	
12	4-hydroxy-tetrahydrodipicolinate synthase	Forward: CACATTGTGCTTTCGCCTTAAT	98
		Reverse: TGATGATGGCGTCTTTGAGTAG	
13	dihydroxy-acid dehydratase	Forward: AAGTGACCTGGCTTGATTGT	106
		Reverse: CCATTCCGGGCTGTGATAAA	
14	3-isopropylmalate dehydratase	Forward: TCCCGGTCTTCCTAGCATAA	81
		Reverse: TGTACAAGCTCAAGCCCATC	
15	GAPDH	Forward: GTTTGCCGTCAACAGAGATTG	97
		Reverse: AATGCTTCTCACCTCCTCAAATA	
16	L-ascorbate oxidase-like	Forward CCCAGAATGTTCACGCTACT	97
		Reverse TTGGAAGTTGAGTGCAGAGAG	
17	RNA-dependent RNA polymerase	Forward: CTGCTTGGGTTCTCGAATCT	136
		Reverse: GAGAAGAATGGCGGGAAGAA	
18	Ethylene-responsive transcription factor	Forward: CTGAGGGTGATGGTTTGAGAAG	87
		Reverse: CCCTCTTTCTCCTCTTCCTATCT	

34

19	kinesin heavy chain	Forward: GCTCTCTATCAGGCAGGTAAAC	92
		Reverse GAACTCCTTGCCCTCAACTATT	
20	growth-regulating factor	Forward: ACATGCACAGAGGCAAGAA	98
		Reverse: AGGAAAGAGAGGATGAGAGGT	
21	chitin-inducible gibberellin-responsive protein	Forward: CATCTGGATAGAAGGTCCATGTC	106
		Reverse: GGGAGCTTCCTTCATCATAGTT	
22	G2/mitotic-specific cyclin S13-7	Forward: GCCAGCTCTGAGAAGAAGAAG	96
		Reverse: GGTGCCCACTCCATATGTTT	
23	mitotic spindle checkpoint protein	Forward: TGGAAGGCCATACTTCTTCAC	100
		Reverse: CGTTAGCGAGTTCTTTGGTTATTC	
24	syntaxin-related protein	Forward: TGAGCGTTATTTGTCGTCTCTTA	97
		Reverse: ACATTAGCCAGAAGCCACTC	
25	mitotic spindle checkpoint protein	Forward: CAACTGTTCTGGTAGCCATTTG	126
		Reverse: CCCATCGTCTTCTCCGTTATT	
26	G2/mitotic-specific cyclin	Forward: CCATCGCCAATCTCCAATCT	92
		Reverse: CTCCCACATACATCATCCTCTTC	
27	protein CHUP1	Forward: GCATCCAGTCAGTTGGAATTTG	96
		Reverse: CAGAGAAGATGGAACGAAGTGT	

Table 12: Transcripts associated with different enzymes and compounds

S. No.	Name	No. of Transcript
1	Shikimate	38
2	Shikimate kinase (2.7.1.71)	4
3	Shikimate 3-phosphate	-
4	3-Phoshoshikimate 1-carboxyvinyl transferase (2.5.1.19)	3
5	5-0-(1-Carboxy vinyl)-3-phososhikimate	3
6	Chorismate synthase (4.2.3.5)	7
7	Chorismate	1
8	Chorismate mutase (5.4.99.5)	4
9	Prephenate	24
10	Aromatic-amino-acid transaminase (2,6.1.57, 2.6.1.1,	10

	2.6.1.78, 2.1.1.79)	
11	L-Arogenate	-
12	Prephenate dehydratase (4.2.1.51, 4.2.1.91, 4.2.1.51)	3
13	L-Phenylalanine	16
14	phenylalanine ammonialyase (4.3.1.24, 4.3.1.25) (PAL)	18
15	trans-cinnamate	-
16	trans-cinnamate 4-mono-oxygenate (1.14.14.91) (C4H)	5
17	4-coumarate or p-coumaric acid	-
18	p-coumarate 3-hydroxylase (1.14.13.-) (C3H)	2
19	Caffeate / Caffeic acid	-
20	caffeic acid 3-O-methyltransferase (2.1.1.68) (CAOMT)	6
21	Ferulate or Ferulic acid	-
22	4-coumarate CoA ligase (6.2.1.12) (4CL)	12
23	Feruloyl-coA	4
24	cinnamoyl CoA reductase (1.2.1.44)	1
25	Coniferyl aldehyde	-
26	4-coumarate coA ligase (6.2.1.12) (4CL)	12
27	Shikimate-O-hyroxy cinnamoyl transferase (2.3.1.133) (HCT)	3
28	Caffeoyl coA	-
29	5-O-caffeoyl shikimic acid	15
30	Chaffeoyl quinic acid	-
31	5-O-(4-coumaroyl)- D-quinate-3'- monooxygenase (1.14.14.96)	3
32	p-Coumaroyl quinic acid	-
33	4-coumaroyl CoA	-