Plant Omics Journal POJ

POJ 14(01):11-22 (2021) ISSN:1836-3644
doi: 10.21475/POJ.14.01.21.p3067

De novo transcriptome sequencing, assembly and characterization of *Heliopsis longipes* roots *vs.* leaves to discover putative genes involved in specialized metabolites biosynthesis

Génesis V. Buitimea-Cantúa[1, 2, ‡*] and Jorge Molina-Torres[1, ‡*]

[1]***Laboratorio de Fitobioquímica, CINVESTAV Unidad Irapuato, Departamento de Biotecnología y Bioquímica, Irapuato, Guanajuato, México***
[2]***Tecnologico de Monterrey, Centro de Biotecnología-FEMSA, Escuela de Ingeniería y Ciencias. Av. Eugenio Garza Sada 2501, Monterrey, N.L., C.P. 64849, México***

[‡] The authors contributed equally to this work

***Corresponding authors: genesis.vidal@tec.mx; jmolinat@cinvestav.mx**

Abstract

Heliopsis longipes is a valuable source of specialized metabolites (or secondary metabolites) with medicinal properties mainly in roots. However, little is known about genes involved in the biosynthesis of these metabolites, primarily due to the lack of genome or transcriptome resources. In this work, the genes of the biosynthetic pathway of the specialized metabolism from *H. longipes* roots and leaves through *de novo* RNA sequencing (RNA-Seq) using the platform of Illumina paired-end sequencing were studied. After *de novo* transcriptome assembly using the software Newbler, a total of 172,342 non-redundant transcripts with an N50 value of 816 bp was obtained. Further functional classification and annotation with Gene Ontology (GO), BLAST2GO, Kyoto Encyclopedia of Genes and Genome (KEGG), and KEGG automatic annotation server (KAAS), revealed that active genes in tissues are predominately involved in the metabolic process and biosynthesis of specialized metabolite pathways. Differential expression analysis of roots *vs.* leaves using Cuffdiff software (*p*-value ≤0.05 and *log*-fold change ratio (*log*2) ≥1) revealed that differentially expressed genes (DEGs) were in an organ-specific manner, such as in leaf, DEGs were significantly enriched in photosynthesis, while in roots, were a higher enriched function of plant hormone signal transduction. A total of 63 transcripts DEGs were related to 9 specialized metabolites pathways, in roots the most abundant was the phenylpropanoid biosynthesis, and in leaves was the carotenoids biosynthesis. Several regulatory genes including the basic-helix-loop-helix and basic leucine zipper domain, transcriptions factor families involved in the regulation of phenylpropanoids and carotenoid biosynthesis, respectively, were discovered. This study established a global transcriptome dataset for *H. longipes*. Data shall be useful to study the functional genomics or genetic engineering of this specie. These results will promote the understanding of the genetic mechanism involved in the biosynthesis of specialized metabolites *in H. longipes.*

Keywords: *Heliopsis longipes*, RNA-Seq**,** specialized metabolism**,** transcriptome.
Abbreviations: BP_biological process; MF_molecular function; CC_cellular component; DEGs_differentially expressed genes; KEGG_Kyoto Encyclopedia of Genes and Genomes; TFs_Transcription factors; NGS_Next-generation sequencing; FPKM_fragments per kilobase of transcript per million fragments sequenced; GO_Gene ontology; RNA-Seq _RNA sequencing.

Introduction

Heliopsis longipes (A. Gray) S.F. Blake is a member of the tribe Heliantheae of the family Asteraceae (Fisher, 1957). It is a Mexican medicinal plant with distribution restricted to the Sierra Gorda and Sierra de Alvarez, which are part of the Sierra Madre Oriental in the state of Guanajuato, San Luis Potosi, and Queretaro (Rzedowski and Calderon, 2008). *H. longipes* roots have long been used in traditional medicine for the treatment of colds, sore throats, and as an analgesic in toothache. Today, many studies have explored the biological activity of *H. longipes* on the various organism (García-Chávez et al., 2004; Ramírez-Chávez et al., 2004; Méndez-Bravo et al., 2010; Méndez-Bravo et al., 2011; Buitimea-Cantúa et al., 2020a). These medicinal and biological properties have been attributed to its high content of alkamides, especially to the affinin/spilanthol, the main specialized metabolites that are synthesized and accumulated in *H. longipes* roots (García-Chávez et al., 2004 Barbosa et al., 2016). Recently, a set of candidate genes

involved in the biosynthesis of the acyl chain of the alkamides by transcriptomic data analysis in *H. longipes* roots was reported (Buitimea-Cantúa et al., 2020b). However, other specialized compounds have been poorly studied in *H. longipes* roots and no reports have been performed on their specialized metabolism at the transcriptomic level.

Significant progress has been made in the identification of the genes and enzymes of specialized metabolism pathways in plant species without genome sequence, using the next generation sequencing (NGS) especially with the RNA-Sequencing (RNA-Seq) technology (Xiao et al., 2013). RNA-Seq providing genomic resources for unraveling genes and biosynthetic pathways involved in metabolite biosynthesis in various medicinal plants, but, a vast majority of medicinal plants are yet to be studied (Kotwal et al., 2016; Loke et al., 2016; Bae et al., 2018; Rai et al., 2018; Eum et al., 2019). The Illumina is a highly utilized RNA-Seq platform employed for transcriptome analyses of various model and non-model organisms, including medicinal plants, due to its potential for a high sequence yield (Nakasugi et al., 2013; Lehnert and Walbot 2014; Rastogi et al., 2014). Illumin is based on sequencing by synthesis, achieving high sequencing coverage at a lower cost, and has been used extensively for *de novo* transcriptome studies (Sedano and Carrascal, 2012). And it is highly efficient to explore and characterizing genes that are involved in biologically active phytocompounds and related pathways (Schliesky et al., 2012; Johnson et al., 2012; Facchini et al., 2012; Xiao et al., 2013).

Therefore, the aim of this study was *de novo* RNA-Seq and the transcriptome analysis of *H. longipes* to identify genes involved in specialized metabolism. In this study, we established the transcript databases for this species and provided genetic information for further genome-wide research and analyses in this plant. Therefore, it presents an important resource for future studies on *H. longipes* and will assist in the functional characterization of candidate genes involved in the biosynthesis of pharmacologically active compounds.

Results

Sequence quality and transcriptome de novo assembly

A total of 9,829,676 base pair (bp) paired-end raw reads of high-quality (HQ) were generated. After the quality filtration (mean quality score >30) and adaptor trimming using Trimmomatic, the high-quality reads were used for de novo assembly using Newbler RNA-Seq assembler. The raw reads were submitted to the NCBI database and assigned numbers SRS654537 for roots and SRS654538 for leaves. Assembled transcript contigs were validated using the CLC Bio Genomics workbench by mapping high quality reads back to the assembled transcript contigs. The assembly resulted in a total of 172,342 non-redundant transcripts with an N50 value of 816 bp, and the largest contig length of 13,358 bp. The size distribution of transcripts ranged from 100 to 13,358 bp, wherein the maximum number of transcripts (99,745) was in the range of 101-500 bp followed by transcripts in the range of 501-1000 bp (21,740) since the number decrease as the transcript length increases (Figure 1). The GC content of the assembled transcript was 42.40%, which corresponds to values reported in plants (Šmarda et al., 2014).

The transcripts were validated by mapping high quality reads back to the assembled transcript. We observed that 91.5% of reads were mapped to the transcript thereby suggesting that the assembly was highly valid. From the total assembled 172,342 transcripts, 79,165 present a significant hit with a known protein, the remaining 93,178 transcripts that did not present a significant hit with any known protein were, most likely, short sequences with less than 100 bp, that probably represented genes that have not yet been functionally characterized, untranslated regions or specific genes for H. longipes. In the functional annotation, we obtained BLAST hits of the 95.0% (163,725 transcripts) related to green plants. Specifically, a large number of *H. longipes* transcripts showed significant similarity with the *Arabidopsis thaliana* (31.02 %), Populus trichocarpa (22.05%), and Vitis vinifera (18.91%).

Functional characterization and identification of transcription factors (TFs) families

Gene ontology (GO) assignments showed 15,139 *TFs* in roots and 14,750 in leaves. In both roots and leaves tissues, the biological processes comprised the majority of the functional terms 47%, followed by cellular component 23%, and molecular functions 30% (Figure 2). The regulation of biological quality (353 transcripts), cation binding (261 transcripts), and the extracellular region (357 transcripts) represented the most abundant term in the biological process, molecular function, and cellular component category, respectively, suggesting, that these transcripts might be involved in some important metabolic activities in *H. longipes*. As observed in the top-ten of the GO analysis in Figure 2, both libraries of tissues, roots, and leaves, showed a similar type of distribution pattern of the transcripts under different GO terms. The distribution of unigenes into the three different GO categories is an indicator of the wide diversity of genes that were present in roots and leaves transcriptomes. These results were consistent with the results of other non-model plant species producing specialized metabolites (Xiao et al., 2013).

Pathway-based analysis can help us further understand the biological significance of the genes. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database contains a systematic analysis of inner-cell metabolic pathways and functions of gene products, which aid in studying the complex biological behavior of genes. Ortholog assignment and mapping of the transcripts to the biological pathways were performed using KEGG automatic annotation server (KAAS). All the transcripts were compared against the KEGG database using BLASTX with a threshold bit-score value of 60 (default). A total of 6,633 transcripts were enriched in 391 different pathway-based analysis were identified, this can help us further understand the biological significance of these genes. The mapped transcripts that represented the metabolic pathways of major biomolecules such as carbohydrates, lipids, nucleotides, amino acids, glycans, cofactors, vitamins, terpenoids, polyketides, and other specialized metabolites were identified. The mapped transcripts also represented genes involved in genetic information processing, environmental information processing, cellular processes, and organizational systems. The pathways with the most representation by the transcripts were: the metabolic pathways PATH: ko01100:

768 transcripts; followed by the biosynthesis of specialized metabolites PATH: ko01110; 353 transcripts.

The major class of plant specialized metabolites identified by chemical groups is flavonoids and phenolic compounds, terpenoids, and nitrogen/sulfur-containing compounds (Xiao et al., 2013). Through the pathway-based analysis for the transcriptome of *H. longipes*, a total of 1,276 unigenes were assigned to 18 different pathways in the KEGG database. The most represented pathways were the terpenoid backbone biosynthesis PATH: ko00900, 24 transcripts; followed by carotenoid biosynthesis (PATH: ko00906, 20 transcripts; and phenylpropanoid biosynthesis PATH: ko00940, 20 transcripts. A minor number of genes were found involved in the biosynthesis of specialized metabolites such as diterpenoid biosynthesis, flavonoid biosynthesis, anthocyanin biosynthesis, and polyketides. These results indicated that a wide range of genes related to specialized metabolites was identified in H. *longipes*. All these genes involved in the enrichment of specialized metabolite biosynthesis would greatly enhance the potential utilization of *H. longipes* in traditional use.

Transcription factors (TFs) affect metabolic flux by regulating gene expression of particular gene' encoding enzymes involved in the biosynthetic pathway and their information would help manipulate metabolic pathways in plants. In this study, BLASTX with a threshold E-value of 1x10-5 was performed to search against the known plant transcription factor database (http://planttfdb.cbi.pku.edu.cn) and Arabidopsis Gene Regulatory Information Server (AGRIS) (https://agris-knowledgebase.org). The identification, annotation, and classification of TFs in the roots and leaves transcriptomes revealed high homology with 46 known TF families. Among them, 35,618 transcripts identified related to a TFs, the most abundant presented high homology to members of the HB-other family 8,738 transcripts, followed by the bHLH 8,448 transcripts, AP2 5,589 transcripts, and members of the MYB 2,818 transcripts. The MYB and bHLH transcription factor genes, regulate the gene expression of the flavonoids, anthocyanins, and condensed tannins (Vom Endt et al., 2002; Hichri et al., 2011). While, the AP2/ERF are involved in the regulation of the terpenoids, and the HB-other have been associated with stress induction and tissue specificity (Davies and Schwinn, 2003; Yang et al., 2012). This result reveals that the transcriptional regulation provides an important control point for the specialized metabolic pathways in *H. longipes*.

Screening and functional characterization differentially expressed genes (DEGs) of roots vs. leaves

The genes with a p-value ≤ of 0.05 and fold change ratio (log2) ≥ |1| were identified as DEGs (Figure 3). The DEGs were visualized as an MA plot (log Fc vs. average log CPM) of roots vs. leaves. The red dots represent transcripts with positive and negative log 2-fold change values, indicating the up-regulation and down-regulation of the DEGs in each comparison (Figure 3 A). The heatmap shows that a high number of genes that are differentially expressed are down-regulated in roots (Figure 3 B). A total of 1,145 genes are up-regulated and 4,303 genes down-regulated in the roots (Figure 3 C). The DEGs were classified with GO and KEGG analysis, this was performed again based on the NR database annotations. The DEGs were assigned to 90 GO categories in the up-regulated 7,073 transcripts, and down-regulated 21,075 transcripts. The top-ten of the GO assignment for the up-and down-regulated genes are shown in Figure 4. The up-regulated were distributed as follows: 2,147 genes were categorized in 30 functional groups for biological process (BP); 3,208 genes in 30 groups for molecular function (MF) category; and 1,718 genes in 30 groups for cellular component (CC) category. The highest abundance of genes was represented in the MF category. While, the down-regulated were categorized as follows: 5,414 genes were categorized in 30 functional groups for BP, 6,422 genes in 30 groups for MF category, and 9,239 genes in 30 groups for the CC category. The highest abundance of genes was represented in the CC category. The enrichment of DEGs in GO terms was tested to gain insights into the biological implications. The GO terms response to stress for BP, transferase activity for MF, and cell periphery for CC, showed a high abundance of up-regulated into these categories. The enrichment of DEGs down-regulated in GO terms was tested to gain insights into the biological implications. The GO terms response to stress for BP, cation binding for MF, and plastid for CC showed a high abundance of down-regulated into these categories.

The metabolic analysis of the DEGs using the KEGG database classified 444 up-regulated transcripts into 27 metabolic pathways and 1,209 down-regulated transcripts were assigned to 30 pathways (Figure 5 A). The top five metabolic pathways were identified. The up-regulated transcripts are related to metabolic pathways, biosynthesis of specialized metabolites, phenylpropanoid biosynthesis, carbon metabolism, and biosynthesis of amino acids; while the down-regulated transcripts are related to metabolic pathways, biosynthesis of specialized metabolites, carbon metabolism, biosynthesis of amino acids, and photosynthesis (Figure 5 B). Regarding transcripts associated with the specialized metabolism, we find in total 63 transcripts DEGs assigned to 9 pathways. Of these, 42 transcripts up-regulated were assigned to 7 pathways and 21 transcripts down-regulated were assigned to 4 pathways (Figure 6). Among these KEGG pathways, were the biosynthesis of phenylpropanoid, terpenoid backbone, flavonoid, isoquinoline-alkaloid, stilbenoid, diarylheptanoid, gingerol, ubiquinone, and another terpenoid-quinone, monoterpenoid, glucosinolate, and carotenoid. The most abundant pathway detected in the up-regulated transcripts was the phenylpropanoid biosynthesis (Figure 6 B-a), while in the down-regulated transcripts was the carotenoid biosynthesis (Figure 6 B-b). These pathways may be playing an important role in H. longipes. In plants, the production of the specialized metabolites would be induced and is mediated by signaling molecules such as reactive oxygen species (ROS), ethylene (ET), and jasmonic acid (JA) (Jacobo-Velázquez et al., 2015). In our RNA-Seq data, 8 genes related to auxins (AUX) were identified up-regulated in roots, while, 22 genes related to JA and 6 genes AUX were identified as highly expressed in the leaves (Figure 7 A). These suggest that the high expression of genes related to the specialized metabolites in roots and leaves would be related to the JA and AUX signals molecules. Thus, the fact that the phenylpropanoid and carotenoid biosynthesis is highly expressed in roots and leaves, indicate that these pathways may be playing an important role in H. longipes tissues.

The identification, annotation, and classification of transcription factors (TFs) in the DEGs revealed a total of 34

transcripts with high homology with 6 known TF´s families. Among these, 30 transcripts highly expressed in roots were assigned to 6 TFs families and 4 transcripts highly expressed in leaves were assigned to 1 TFs family (Figure 7 B). The most abundant TFs family detected in roots transcripts is the bHLH, while in leaves was the bZIP. The bHLH transcription factor genes regulate the gene expression of the flavonoids, anthocyanins, and condensed tannins (Vom Endt et al., 2002; Hichri et al., 2011). While the bZIPs perform a plethora of functions in developmental, environmental, and stress signaling (Dröge-Laser et al., 2018). In plants, bZIP transcription factors regulate many processes, including ABA and stress signaling, as well as, an important player in light signaling and have been related to activating carotenoid biosynthesis genes (Zhang et al., 2008; Stanley and Yuan,2019). This result reveals that the transcriptional regulation provides an important control point for the specialized metabolic pathways in H. longipes.

Phenylpropanoids biosynthesis: the main specialized metabolism pathway expressed in H. longipes roots

The main specialized metabolism pathway highly expressed in roots was related to the biosynthesis of the phenylpropanoids. These are a family of organic compounds with an aromatic ring and a three-carbon propene tail and are synthesized by plants from the phenylalanine and tyrosine amino acids (Zhang and Stephanopoulos, 2016). In this work, 8 genes related to the phenylpropanoid pathway were identified up-regulated in roots (Figure 8). The most differentially expressed gene was codifying the enzyme peroxidase (POD), followed by the cinnamate-4-hydroxylase (ATC4H), and S-adenosyl-L-methionine-dependent O-methyltransferases (CCoAOMT). The POD is responsible for the oxidative polymerization of monolignols to produce lignin (Sánchez et al., 1996). The ATC4H is a cytochrome P450 that catalyzes the second step of the main phenylpropanoid pathway leading to the synthesis of lignin, pigments, and many defense molecules as salicylic acid (SA), an essential trigger of plant disease resistance (Schoch et al., 2002; Vogt, 2010). The CCoAOMT is involved in the biosynthesis of lignin and many plant secondary products, and the methylations are often essential in determining specific physiological functions of the resultant molecules (Coiner et al., 2006). Phenylpropanoids act as preformed and inducible antimicrobial compounds, as well as signal molecules in plant-microbe interactions. Phenylpropanoid plant defense ranges from preformed or inducible physical and chemical barriers against infection to signal molecules involved in local and systemic signaling for defense gene induction (Naoumkina et al., 2010). Among the genes related to phenylpropanoids biosynthesis, also was identified as the phenylalanine ammonia-lyase (PAL), the 4-coumarate: CoA ligase (4CL), and the cinnamate-4-hydroxylase (CCR). Combined enhanced expression of PAL1, 4CL, C4H, and the lignin-specific biosynthesis gene, CCR, has been reported in plant resistance against secondary infection in roots (Lahlali et al., 2017). It has also been suggested that the accumulation of phenolic compounds in root cell walls suggests that cell wall lignification is a defense response of plants (Fuchs and Sacristán, 1996). These data demonstrated that high expression of phenylpropanoid pathway genes in *H. longipes* roots provides an insight into the role of these specialized metabolites in this tissue.

Carotenoids biosynthesis: the main specialized metabolism pathway expressed in H. longipes leaves

The main specialized metabolism pathway highly expressed in leaves was related to the biosynthesis of the carotenoids: isoprenoid compounds, mostly C40 with polyene chains that may contain up to 15 conjugated double bonds. These metabolites act as plant pigments that function as antioxidants, hormone precursors, and colorants (Howitt and Pogson, 2006). Carotenoids can function as regulators of plant growth and development, as accessory pigments in photosynthesis, as photoprotection preventing photo-oxidative damage (Hirschberg 2001). The genes encoding nearly all the enzymes for carotenoid biosynthesis in plants have been identified, and their enzymatic activities have been characterized (Cunningham and Gantt, 1998; Hirschberg, 2001; Howitt and Pogson, 2006). In this work, 9 genes related to the carotenoids biosynthesis pathway was identified as highly expressed in leaves (Figure 9). The most differentially expressed genes were codifying the enzyme nine-cis-epoxycarotenoid dioxygenase (NCED4), followed by the zeta-carotene desaturase (ZDS), and phytoene synthase (PSY). The NCED4 belongs to the carotenoid cleavage dioxygenase (CCD) family and regulated the biosynthesis of the abscisic acid (ABA). These enzymes catalyze the 11, 12 double bond cleavage of 9-cis-violaxanthin and 9-cis-neoxanthin, leading to the production of xanthoxin (Priya and Siva, 2015). The ZDS and PSY, act on the first step of the carotenoid metabolic pathway during the biosynthesis of the first true carotenoid C40 molecule, phytoene (Araya-Garay et al., 2014). The carotenoids act as precursors to the hormone abscisic acid (ABA) and perhaps other hormones as well (Hirschberg 2001; Auldridge et al., 2006). Thus, in addition to the NCED also was identified the abscisic aldehyde oxidase 3 (AAO3), is an enzyme responsible for the conversion of ABA- aldehyde to ABA, the final step in ABA biosynthesis (González-Guzman et al., 2004; Szepesi et al., 2009). The highly expressed genes of the carotenoids pathway in *H. longipes* leaves provide an insight into the role of these specialized metabolites in this photosynthetic tissue.

Discussion

Heliopsis longipes is a valuable source of bioactive metabolites with medicinal and industrial esteem. However, little is known about genes involved in the biosynthesis of these specialized metabolites, primarily due to the lack of genome or transcriptome resources. In previous studies, RNA-Seq technology has been used to study the genes related to specialized metabolism (Xiao et al., 2013; Zhang et al., 2015; Rai et al., 2018; Guo et al., 2019). In this study, the global expression patterns of genes involved in metabolism, particularly specialized (secondary) metabolism of H. longipes roots and leaves, were identified through RNA sequencing. A total of 172,342 unigenes were obtained from transcriptome sequencing, where 79,165 (46%) presented a significant hit with a known protein. In the functional annotation, we obtained BLAST hits of the 95% related to green plants, especially, a significant similarity with Arabidopsis thaliana. Differential gene expression analysiswas performed to detect potential differentially expressed genes (DEGs) in *H. longipes* roots vs. leaves.

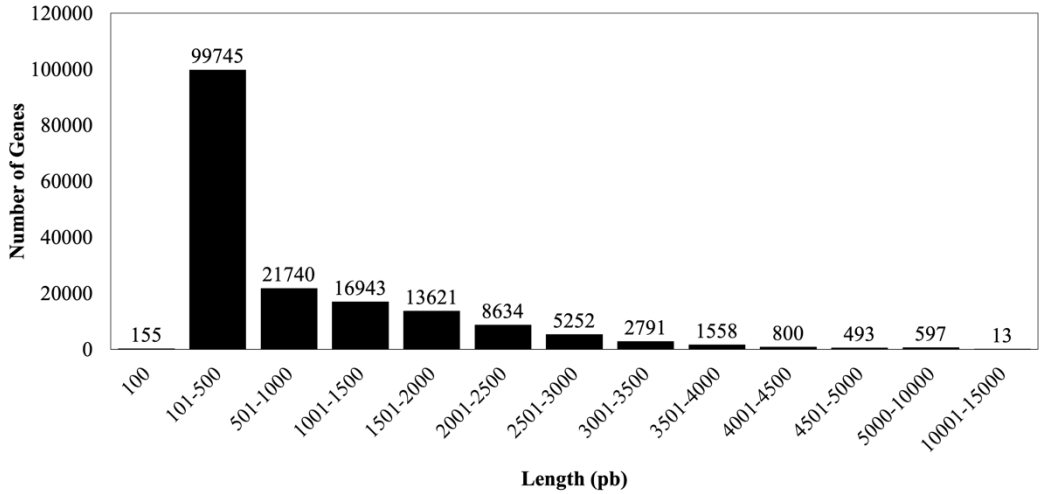

Fig 1. Unigenes length distribution obtained by the *de novo* assembly of the transcriptome sequencing in *Heliopsis longipes* roots and leaves transcriptome.

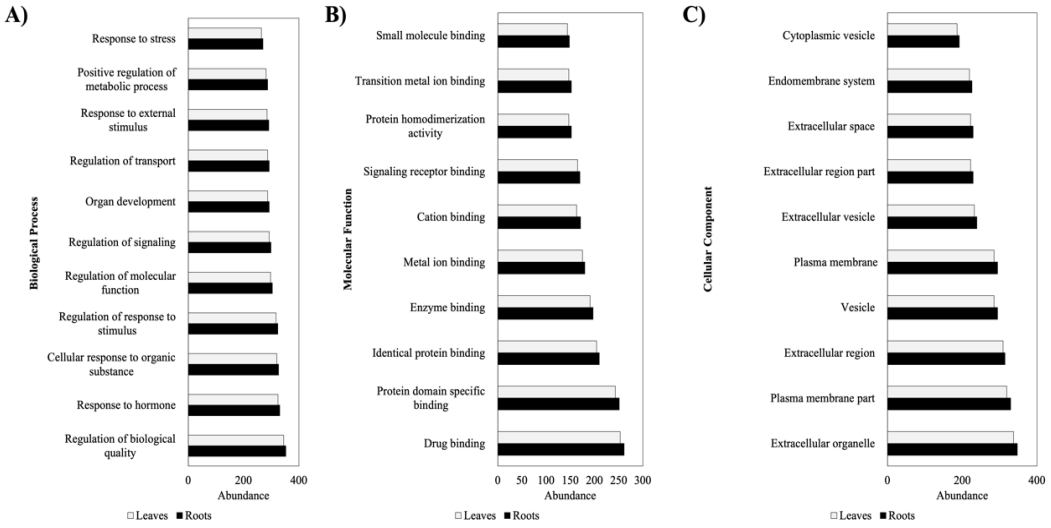

Fig 2. Top-ten of the gene ontology **(**GO) term assignment to the *de novo* assembled transcripts of *Heliopsis longipes* roots and leaves. Distribution in the categories of GO: biological process, molecular function, and cellular component.

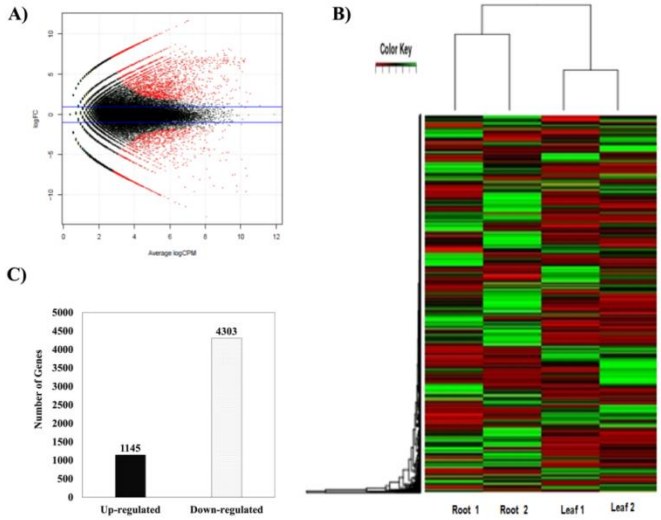

Fig 3. Differential expression genes (DEGs) analysis. **A)** MA plot (*log*FC *vs.* average *log*CPM) of roots *vs.* leaves. The red dots represent transcripts with positive and negative log2fold change values, indicating the up-regulation and down-regulation of the DEGs, respectively. **B)** Heat map of the differently expressed genes. Red and green colors indicate up-and down-regulated gene expression, respectively. **C)** Total of genes up-regulated and down-regulated.

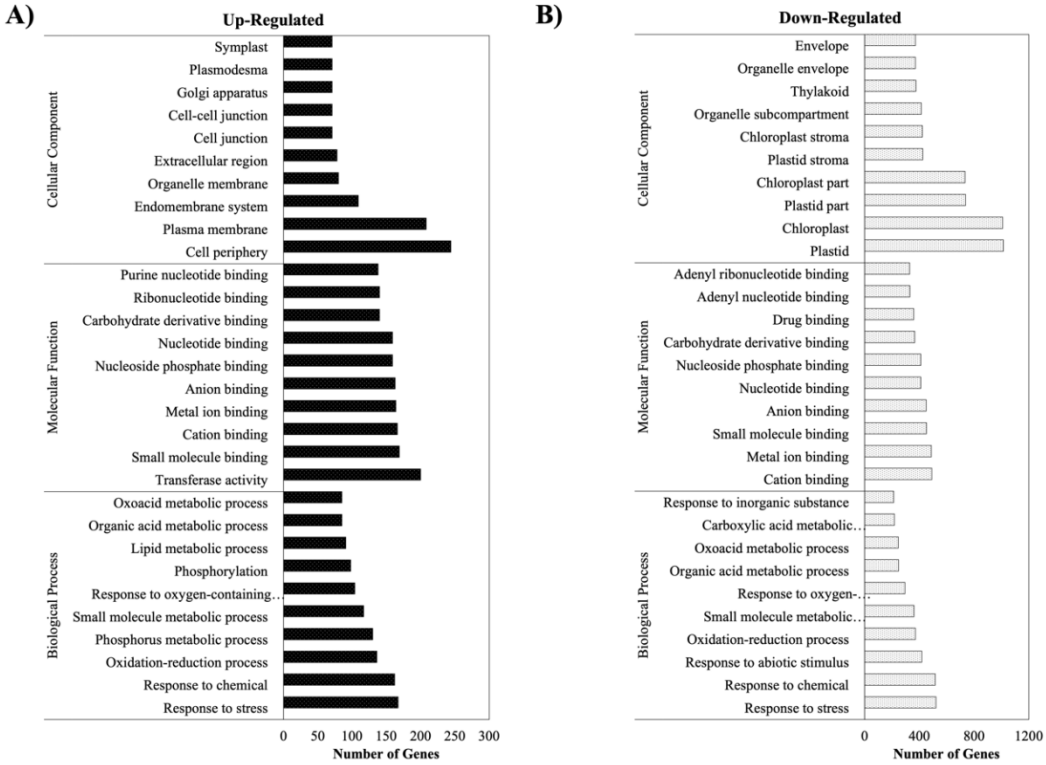

Fig 4. Top-ten of the GO term assignment to the differential expression genes in *Heliopsis longipes* roots and leaves transcriptomes. Distribution in the categories of GO: biological process, molecular function, and cellular component.

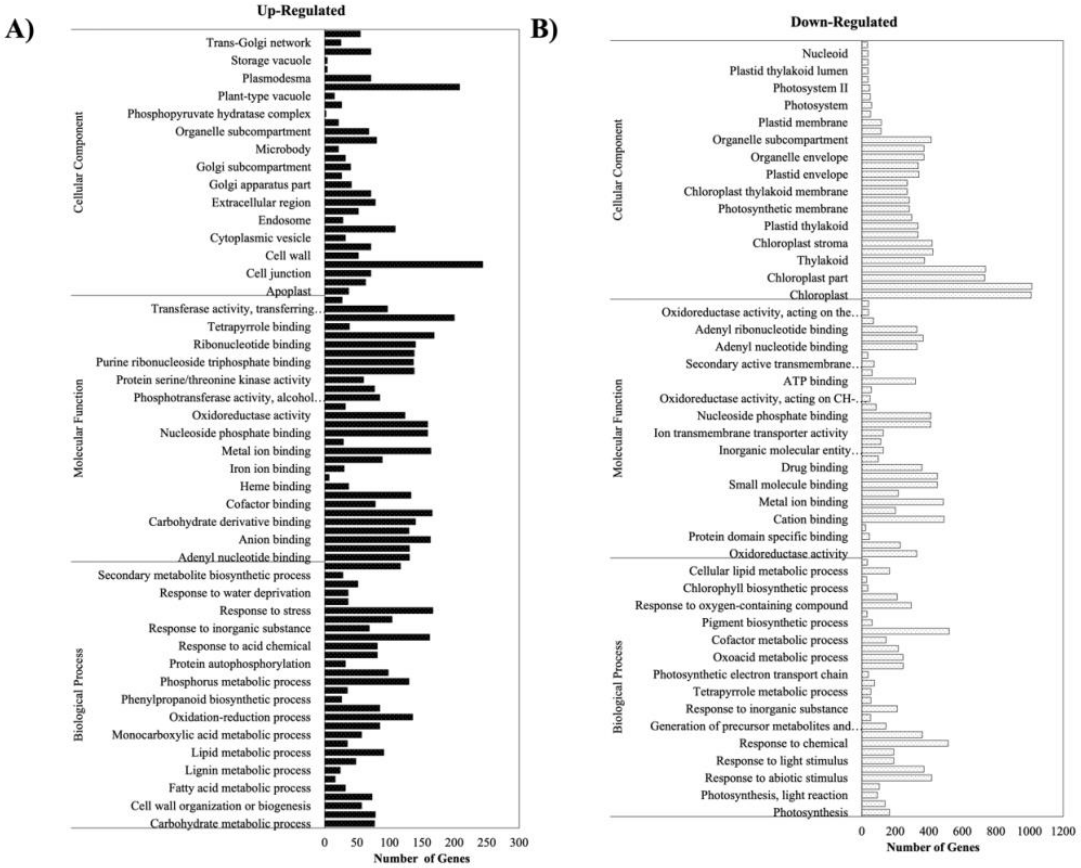

Fig 5. KEGG pathway analysis of the genes with differential expression in *Heliopsis longipes* roots and leaves transcriptomes. **A)** Abundance comparison of the KEGG pathways of the up and down-regulated genes. **B)** Network of the pathway in the up and down-regulated genes.

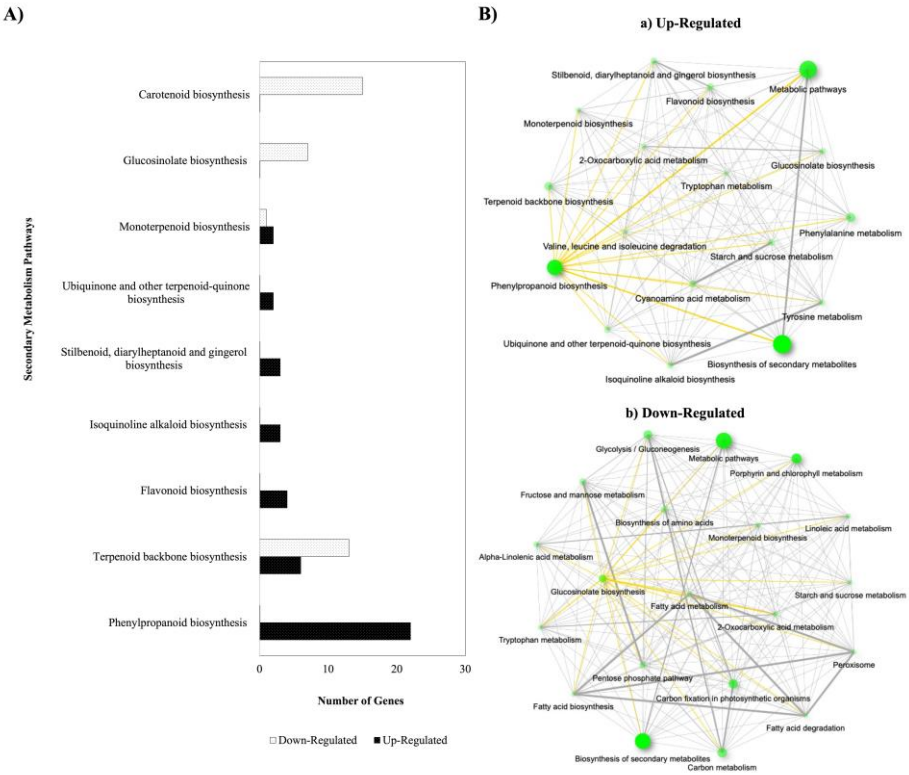

Fig 6. KEGG analysis of specialized metabolism pathways in the DEGs in *Heliopsis longipes* roots and leaves transcriptomes. **A)** Abundance comparison of the secondary metabolism pathways of the up and down-regulated genes. **B)** Network of the pathways in the up and down-regulated genes**. C)** Abundance comparison of the transcription factors of the up and down-regulated genes.

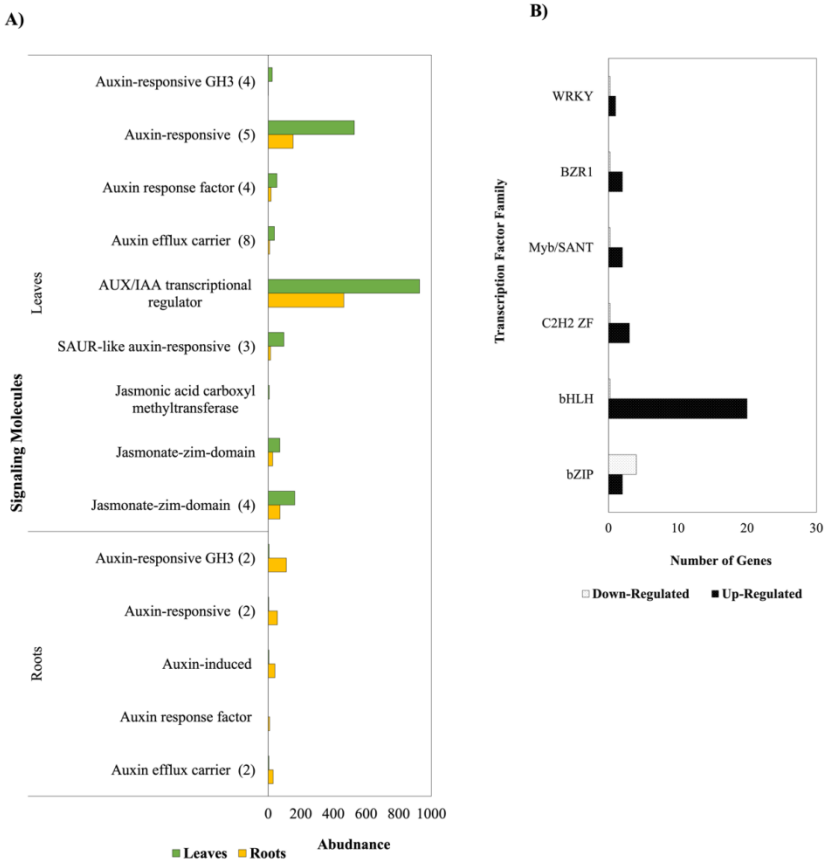

Fig 7. Analysis of up-regulated genes. **A)** Enzymes related to signaling molecules identified up-regulated genes in *Heliopsis longipes* roots and leaves transcriptomes. **B)** Abundance comparison of the transcription factors of the up-and down-regulated genes.

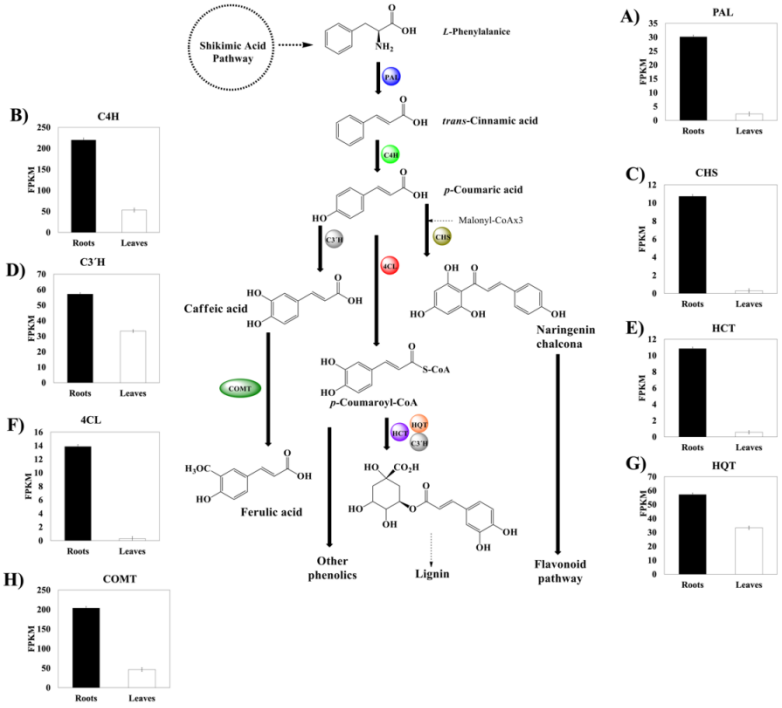

Fig 8. The main specialized metabolism pathway highly expressed in *Heliopsis longipes* roots. **A-H)** Relative expression of genes related to phenylpropanoids biosynthesis in plants. Phenylpropanoid biosynthetic pathway: 4CL, 4 coumarate CoA ligase; C3'H, *p*-coumaroyl ester 3 hydroxylase; C4H, cinnamate 4 hydroxylase; CHS, chalcone synthase; COMT I, caffeic/5-hydroxyferulic acid *O*-methyltransferase; HCT, hydroxycinnamoyl transferase; HQT, hydroxycinnamoyl CoA: quinate hydroxycinnamoyl transferase; PAL, phenylalanine ammonia-lyase.

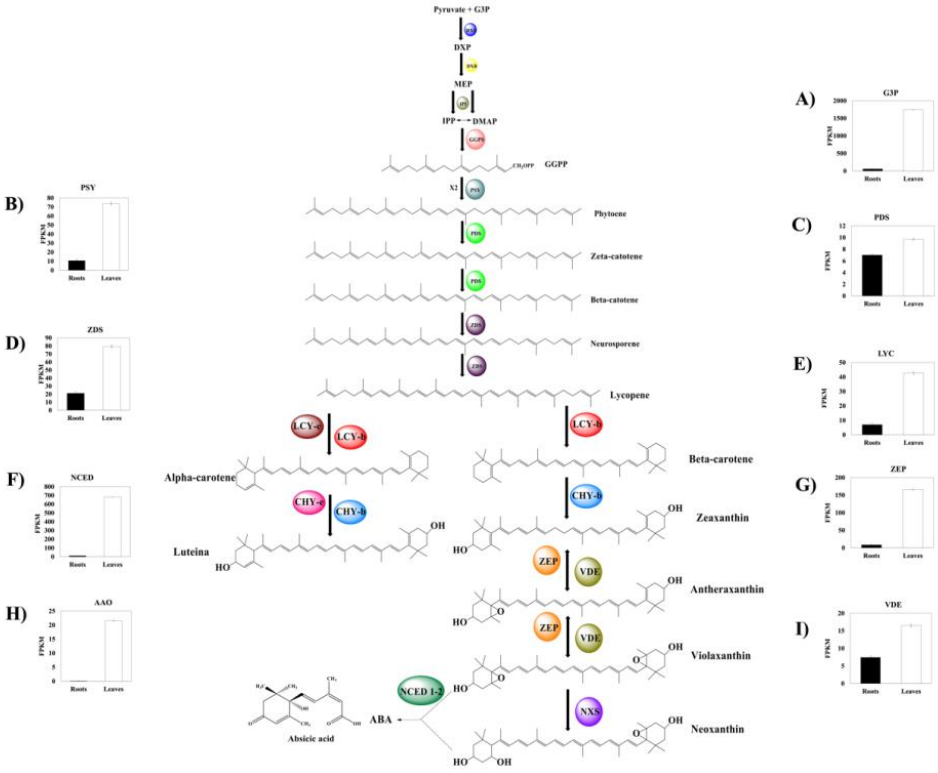

Fig 9. The main specialized metabolism pathway highly expressed in *Heliopsis longipes* leaves. **A-I)** Relative expression of genes related to carotenoids biosynthesis. Carotenoids biosynthetic pathway in plants: G3P, D-glyceraldehyde-3-phosphate; DXP, 1-deoxy-D-xylulose-5-phosphate; MEP, 2-C-methyl-D-erythritol-4-phosphate; IPP, isopentenyl diphosphate; DMAP, dimethylallyl diphosphate; GGPP, geranylgeranyl diphosphate; ABA, abscisic acid; DXS, 1-deoxy-D-xylulose-5-phosphate synthase; DXR, 1-deoxy-D-xylulose-5-phosphate reductase; IPI, isopentenyl diphosphate isomerase; GGPS, geranylgeranyl diphosphate synthase; PSY, phytoene synthase; PDS, phytoene desaturase; ZDS, ζ-carotene desaturase; CRTISO, carotene isomerase; LCY-e, lycopene ε-cyclase; LCY-b, lycopene *β*-cyclase; CHY-e, ε-carotene hydroxylase; CHY-b, *β*-carotene hydroxylase; VDE, violaxanthin de-epoxidase; ZEP, zeaxanthin epoxidase; NXS, neoxanthin synthase; NCED, 9-cis-epoxycarotenoid dioxygenase.

Based on the evidence, a total of 5,718 DEGs were identified, of which 1,415 were up-regulated and 4,303 genes down-regulated. Most DEGs are annotated in essential functions such as metabolic process and secondary metabolic process. It was observed that DEGs were assigned into different KEGG pathways according to the organ-specific function. For instance, in leaves, DEGs were significantly enriched in pathways like photosynthesis and carotenoid biosynthesis, whereas in the root tissue, DEGs were a higher enriched function of plant hormone signal transduction and phenylpropanoid biosynthesis. A functional analysis using the GO and KEGG classification system of DEGs can provide a new insight for studying organ-specific processes, functions, and pathways among the two different *H. longipes* tissues, such as reported in other plant tissues (Liang et al., 2019). DEGs were identified to analyze the genes involved in the coding of enzymes implied in the synthesis of specialized metabolites. A total of 42 transcripts highly expressed in roots were assigned to 7 pathways, while in leaves, 21 transcripts were assigned to 4 pathways. Several of the identified genes encode for the synthesis of specialized metabolites that are used in traditional medicine, as well as colorants, flavors, and antimicrobials (Xiao et al., 2013). It has been reported that, in plants, the production of the specialized metabolites would be induced and is mediated by signaling molecules such as reactive oxygen species (ROS), ethylene (ET), and jasmonic acid (JA) (Jacobo-Velázquez et al., 2015). It was found that the expression of genes related to the specialized metabolites identified in roots and leaves would be related to the JA and AUX signals molecules, identified in our transcriptomes. Previous studies revealed that the exogenous application of *H. longipes* roots extract on Arabidopsis thaliana promotes the plant-growth and the induction of defense-responsive transcriptional networks (Ramírez-Chávez et al., 2004; Méndez-Bravo et al., 2010; Méndez-Bravo et al., 2011). This effect has been related to the alkamides, a class of bioactive amides produced at high levels in H. longipes roots (García-Chávez et al., 2004). In particular, alkamides induce immunity in A. thaliana through JA-dependent signaling (Méndez-Bravo et al., 2011). These results indicate that AUX genes in roots, and the JA and AUX genes highly expressed in leaves, could induce the biosynthesis of the specialized metabolites by the indirect effect of the alkamides produced in the roots of the species in this study. The highly expressed pathway detected in the roots was the phenylpropanoid biosynthesis, while in leaves was carotenoids biosynthesis. As the results showed, the candidate genes involved in phenylpropanoid and carotenoid biosynthesis represent different expression patterns in the two different organs of H. longipes, where they have specialized functions, by example, phenylpropanoids in roots act as preformed plant defense, and carotenoids can function as accessory pigments in photosynthesis, augmenting light-harvesting or as precursors to the hormone abscisic acid (ABA) (Hirschberg 2001; Irani et al., 2018; 2019). Among the DEGs encoding several regulatory genes, including the basic-helix-loop-helix and basic Leucine Zipper Domain, transcriptions factor families involved in the regulation of phenylpropanoids and carotenoid biosynthesis, respectively, were discovered. All of them had been previously determined in transcriptional analysis in many plant species (Vom Endt et al., 2002; Hichri et al., 2011; Dröge-Laser et al., 2018). Overall results, the DEGs analysis revealed that the highest numbers of candidate genes that might be involved in the biosynthesis of specialized metabolites, mainly in the phenylpropanoids and carotenoids biosynthesis, therefore, these genes may help further functional genomic and transcriptomic analyses in *H. longipes*. This work will contribute to the comprehensive knowledge of plants for growers and consumers and provides additional characteristics and information regarding the pharmaceutical benefits associated with specialized metabolites biosynthesis.

Materials and Methods

Plant material

Twenty Heliopsis longipes specimens were collected at Puerto de Tablas, Xichú municipality, Sierra Gorda, state of Guanajuato, México (Lat. 21º14'20" N, Long. 100º05'19" W, Alt. 2,589 m above sea level). The collected tissue, roots, and leaves, were separately, washed with distilled water, immediately frozen in liquid nitrogen, and stored at -80°C until processed for RNA extraction. A voucher specimen was deposited at Instituto de Ecología, Centro Regional del Bajío. Herbarium IEB (Voucher 263787).

RNA extraction

RNA was extracted from roots and leaves of twenty individual plants of H. longipes roots and leaves separately as two replicates using the protocol established by "PureLinkTM Micro-to-Midi Total RNA Purification System (Invitrogen, Carlsbad, CA, USA). Yield obtained was on average 30 µg of RNA per mg of tissue. The quality of the RNA samples was evaluated using the RNA 6000 Nanochip on the BioAnalyzer 2100 (Agilent Technologies, Cedar Creek, TX, USA). The RNA was stored at -80°C until further use.

cDNA library construction and RNA sequencing

The TruSeq Stranded mRNA sample preparation kit (Illumina) was used to enrich samples for mRNA and construct complementary DNA (cDNA) libraries. Pair end datasets, with a read length of 150 nucleotides, were generated from each sample library on an Illumina MiSeq platform (Illumina, San 0Diego, CA, USA). cDNA preparation, library construction, and sequencing were performed according to Illumina manufacturer instructions at the Unidad de Servicios Genómicos of CINVESTAV, Irapuato, Guanajuato, México. Sequence reads with high-quality scores were written into Standard Flow gram Format (SFF) files. The sequences from the MiSeq runs have been deposited in the national center for biotechnology information (NCBI) and can be accessed through the BioProject ID PRJNA616161, BioSample: SAMN14483285, and the reviewer link: https://dataview.ncbi.nlm.nih.gov/object/PRJNA616161?reviewer=h693bllbujhlte4pqqc7d9ch0a.

Sequence quality and de novo transcriptome assembly

The RNA-Seq reads quality was check-in Galaxy web-based platform (https://usegalaxy.org). The resulting FASTQ files were pre-processed by removing adapter sequences and low-quality bases (Q>30) using the software package Trimmomatic-0.32, as previously described (Kamitani et al., 2016). The reads that passed the quality check were assembled into unigenes using the Newbler software with

the default parameters. To reduce redundancy, Unigene Clustering was performed using the software package CD-HIT 4.0 with the 95% identity parameter to generate a set of non-redundant contig sequence files (Fu et al., 2012). To estimate expression abundance, reads were mapped to the de novo transcriptome assemblies using Bowtie 2.0 (Langmead et al., 2009). Expression abundance was calculated using the RNAseq by Expectation-Maximization (RSEM) software (Li and Dewey, 2011).

Transcriptome annotation and analysis

Unigenes were used for BLAST search and annotation against the non-redundant (NR) protein database at NCBI (http://www.ncbi.nlm.nih.gov) and the Swiss-Prot protein database (http://www.ebi.ac.uk/uniprot) with an E-value cutoff of 1x10e-5. Functional annotation by Gene Ontology (GO) (http://www.geneontology.org/) terms was analyzed using BLAST2GO software and the unigenes were assigned to biological functions on the macro levels of biological process, cellular component, and molecular function. The Kyoto Encyclopedia of Genes and Genome (KEGG) pathways database (http://www.genome.jp/kegg) was assigned to unigenes by KEGG automatic annotation server (KAAS). Using the basic local alignment search tool (BLAST), the Arabidopsis thaliana transcription factor (TF) database was queried to identify TFs among all unigenes (identity N80%) (Guo et al., 2005). The read counts were further normalized into FPKM values. The FPKM values from the four libraries were pairwise compared, and the fold changes were calculated by using RSEM software, and DEGs were identified by using EdgeR software package (v3.8.2), if $p< 0.05$ and FDR < 0.05, then the result would be considered statistically significant. Subsequently, the enrichment analysis of GO and KEGG pathways was performed based on these DEGs by using Goatools software (v0.4.7) and KOBAS 2.0 software (http://kobas.cbi.pku.edu.cn). All the heat maps for gene clustering in the present study were depicted by using the R program (http://www.R-project.org/).

Differential expression analysis of roots vs. leaves

The expression level of the genes was estimated and normalized as Reads Per Kilobase Million Mapped Reads (RPKM), which is a normalized measure of reading density that allows transcript levels to be compared both within and between samples (Trapnell et al., 2013). Cuffdiff (v2.2.1) software was used to identify the Differentially Expressed Genes (DEGs) between the two groups, and a p-value was assigned to each gene to evaluate its statistical significance (Trapnell et al., 2012). We determined the False Discovery Rate (FDR) of the test to account for Type I errors. Multiple-testing corrections were performed using the Benjamini and Hochberg step-up false-discovery rate (FDR)-controlling procedure to calculated adjusted p-values. Genes with an adjusted p-value ≤ of 0.05 and an expression log2 ratio ≥ 1 were identified as DEGs.

Conclusion

Our result indicated that H. longipes possess a wide range of genes involved in a complex metabolic grid because all major structural genes of specialized metabolism were found in this study. This work expands the resources available for investigating the gene expression profiles of the *H. longipes* species. These results aid our understanding of how the expression of specialized metabolite biosynthetic genes is regulated in plants. Thus, through this study, we not only generated a high-quality genomic resource for H. longipes but also propose candidate genes to be involved in the biosynthesis pathways of the specialized metabolites for further functional analysis.

Acknowledgments

The author Génesis VBC, with CVU 377834 gratefully acknowledges the Consejo Nacional de Ciencia y Tecnología (CONACYT, Mexico) for the financial support given to carry out her doctoral studies.

References

Araya-Garay JM, Feijoo-Siota L, Veiga-Crespo P, Sanchez-Perez A, González Villa, T (2014) Cloning and functional expression of zeta-carotene desaturase, a novel carotenoid biosynthesis gene from Ficus carica. Int J Microbiol Adv Immunol. 2:32-40.

Auldridge ME, McCarty DR, Klee HJ (2006) Plant carotenoid cleavage oxygenases and their apocarotenoid products. Curr Opin Plant Biol. 9:315-321.

Barbosa AF, de Carvalho MG, Smith RE, Sabaa-Srur AU (2016) Spilanthol: occurrence, extraction, chemistry, and biological activities. Rev Bras Farmacogn. 26:128-133.

Bae DY, Eum S M, Lee SW, Paik JH, Kim SY, Park M, Na JK (2018) Enrichment of genomic resources and identification of simple sequence repeats from medicinally important Clausena excavata. Biotech. 8:1-10.

Buitimea-Cantúa GV, Buitimea-Cantúa NE, del Refugio Rocha-Pizaña M, Rosas-Burgos EC, Hernández-Morales A, Molina-Torres J (2020a) Antifungal and anti-aflatoxigenic activity of Heliopsis longipes roots and affinin/spilanthol against Aspergillus parasiticus by downregulating the expression of alf D and afl R genes of the aflatoxins biosynthetic pathway. J Environ Sci Heal B. 55:210-219.

Buitimea-Cantúa GV, Marsch-Martinez N, Ríos-Chavez P, Méndez-Bravo A, Molina-Torres J (2020b) Global gene expression analyses of the alkamide-producing plant Heliopsis longipes supports a polyketide synthase-mediated biosynthesis pathway. PeerJ. 8:1-24.

Coiner H, Schröder G, Wehinger E, Liu CJ, Noel JP, Schwab W, Schröder J (2006) Methylation of sulfhydryl groups: a new function for a family of small molecule plant O-methyltransferases. Plant J. 46:193-205.

Cunningham JFX, Gantt E (1998) Genes and enzymes of carotenoid biosynthesis in plants. Annu Rev Plant Biol. 49:557-583.

Davies KM, Schwinn KE (2003) Transcriptional regulation of secondary metabolism. Funct Plant Biol. 30:913-925.

Dröge-Laser W, Snoek BL, Snel B, Weiste C (2018) The Arabidopsis bZIP transcription factor family-an update. Curr Opin Plant Biol. 45:36-49.

Eum SM, Kim SY, Hong JS, Roy NS, Choi S, Paik J, Na JK (2019) Transcriptome analysis and development of SSR markers of ethnobotanical plant Sterculia lanceolata. Tree Genet Genomes.15:31-37.

Facchini PJ, Bohlmann J, Covello PS, De Luca V, Mahadevan R, Page JE, Martin VJ (2012) Synthetic biosystems for the production of high-value plant metabolites. Trends Biotechnol. 30:127-131.

Fisher TR (1957) Taxonomy of the genus Heliopsis (Compositae). Ohio J Sci. 57:171-191.

Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinf. 28:3150-3152.

Fuchs H, Sacristán MD (1996) Identification of a gene in Arabidopsis thaliana controlling resistance to clubroot (Plasmodiophora brassicae) and characterization of the resistance response. Mol Plant Microbe Interact. 9:91-97.

García-Chávez A, Ramírez-Chávez E, Molina-Torre J (2004) El género Heliopsis (Heliantheae; Asteraceae) en México y las alcamidas presentes en sus raíces. Acta Bot Mex. 69:115-131.

González-Guzmán M, Abia D, Salinas J, Serrano R, Rodríguez PL (2004) Two new alleles of the abscisic aldehyde oxidase 3 gene reveal its role in abscisic acid biosynthesis in seeds. Plant Physiol.135:325-333.

Guo A, He K, Liu D, Bai S, Gu X, Wei L, Luo J (2005) DATF: a database of Arabidopsis transcription factors. Bioinf. 21:2568-2569.

Guo Y, Zhu C, Zhao S, Zhang S, Wang W, Fu H, Lai Z (2019) De novo transcriptome and phytochemical analyses reveal differentially expressed genes and characteristic secondary metabolites in the original oolong tea (Camellia sinensis) cultivar 'Tieguanyin'compared with cultivar 'Benshan'. BMC Genet. 20:241-265.

Hichri I, Barrieu F, Bogs J, Kappel C, Delrot S, Lauvergeat V (2011) Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. J Exp Bot. 62:2465-2483.

Hirschberg J (2001) Carotenoid biosynthesis in flowering plants. Curr Opin Plant Biol. 4:210-218.

Howitt CA, Pogson B J (2006) Carotenoid accumulation and function in seeds and non-green tissues. Plant Cell Environ. 29:435-445.

Irani S, Trost B, Waldner M, Nayidu N, Tu J, Kusalik AJ, Bonham-Smith PC (2018). Transcriptome analysis of response to Plasmodiophora brassicae infection in the Arabidopsis shoot and root. BMC Genet. 19:1-23.

Irani S, Todd CD, Wei Y, Bonham-Smith PC (2019) Changes in phenylpropanoid pathway gene expression in roots and leaves of susceptible and resistant Brassica napus lines in response to Plasmodiophora brassicae inoculation. Physiol Mol Plant Path. 106:196-203.

Jacobo-Velázquez DA, González-Agüero M, Cisneros-Zevallos L (2015) Cross-talk between signaling pathways: the link between plant secondary metabolite production and wounding stress response. Sci Rep. 5:8598-8608.

Johnson MT, Carpenter EJ, Tian Z, Bruskiewich R, Burris JN, Carrigan CT, Edger PP (2012) Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. PloS one. 7:1-12.

Kamitani M, Nagano AJ, Honjo MN, Kudoh H (2016) RNA-Seq reveals virus–virus and virus–plant interactions in nature. FEMS Microbiol Ecol. 92:1-11.

Kotwal S, Kaul S, Sharma P, Gupta M, Shankar R, Jain M, Dhar MK (2016) De novo transcriptome analysis of medicinally important Plantago ovata using RNA-Seq. PloS one. 11:1-23

Lahlali R, Song T, Chu M, Yu F, Kumar S, Karunakaran C, Peng G (2017) Evaluating changes in cell-wall components associated with clubroot resistance using fourier transform infrared spectroscopy and RT-PCR. Int J Mol Sci.18:1-14

Lehnert EM, Walbot V (2014) Sequencing and de novo assembly of a Dahlia hybrid cultivar transcriptome. Front Plant Sci. 5:335-340.

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics. 12: 307-323.

Liang Y, Zhang X, Zou J, Shi Y, Wang Y, Tai J, Yang M (2019) Pharmacology mechanism of Flos magnoliae and Centipeda minima for treating allergic rhinitis based on pharmacology network. Drug Dev Ind Pharm. 45:1547-1555.

Loke KK, Rahnamaie-Tajadod R, Yeoh CC, Goh HH, Mohamed-Hussein ZA, Noor N M, Ismail I (2016) RNA-seq analysis for secondary metabolite pathway gene discovery in Polygonum minus. Genom Data. 7:12-13.

Méndez-Bravo A, Raya-González J, Herrera-Estrella L, López-Bucio J (2010) Nitric oxide is involved in alkamide-induced lateral root development in Arabidopsis. Plant Cell Physiol. 51:1612-1626.

Méndez-Bravo A, Calderón-Vázquez C, Ibarra-Laclette E, Raya-González J, Ramírez-Chávez E, Molina-Torres J, Herrera-Estrella L (2011) Alkamides activate jasmonic acid biosynthesis and signaling pathways and confer resistance to Botrytis cinerea in Arabidopsis thaliana. PloS one. 6(11), e27251.

Nakasugi K, Crowhurst RN, Bally J, Wood CC, Hellens RP, Waterhouse PM (2013) De novo transcriptome sequence assembly and analysis of RNA silencing genes of Nicotiana benthamiana. PloS one. 8:1-15.

Naoumkina MA, Zhao Q, Gallego-Giraldo LINA, Dai X, Zhao PX, Dixon RA (2010) Genome-wide analysis of phenylpropanoid defence pathways. Mol Plant Pathol. 11:829-846.

Priya R, Siva R (2015) Analysis of phylogenetic and functional diverge in plant nine-cis epoxycarotenoid dioxygenase gene family. J Plant Res.128:519-534.

Rai A, Nakaya T, Shimizu Y, Rai M, Nakamura M, Suzuki H, Yamazaki M (2018) De novo transcriptome assembly and characterization of Lithospermum officinale to discover putative genes involved in specialized metabolites biosynthesis. Planta Med. 84:920-934.

Ramírez-Chávez E, López-Bucio J, Herrera-Estrella L, Molina-Torres J (2004) Alkamides isolated from plants promote growth and alter root development in Arabidopsis. Plant Physiol.134:1058-1068.

Rastogi S, Meena S, Bhattacharya A, Ghosh S, Shukla RK, Sangwan NS, Nagegowda DA (2014) De novo sequencing and comparative analysis of holy and sweet basil transcriptomes. BMC Genet. 15:540-588.

Rzedowski J, Calderón de RG (2008) Familia Compositae, Tribu Heliantheae I (géneros Acmella-Jefea). Flora del Bajío y de Regiones Adyacentes. Pátzcuaro, Michoacán: Instituto de Ecología, AC, Centro Regional del Bajío, 157:1-344.

Sanchez M, Pena MJ, Revilla G, Zarra I (1996) Changes in dehydrodiferulic acids and peroxidase activity against ferulic acid associated with cell walls during growth of Pinus pinaster hypocotyl. Plant Physiol. 111: 941-946.

Schliesky S, Gowik U, Weber AP, Bräutigam A (2012) RNA-seq assembly-are we there yet?. Front Plant Sci.3:208-220.

Schoch GA, Nikov GN, Alworth WL, Werck-Reichhart D (2002) Chemical inactivation of the cinnamate 4-hydroxylase allows for the accumulation of salicylic acid in elicited cells. Plant Physiol. 130:1022-1031.

Sedano JCS, Carrascal CEL (2012) RNA-seq: herramienta transcriptómica útil para el estudio de interacciones planta-patógeno. Fitos.16:101-113.

Šmarda P, Bureš P, Horová L, Leitch IJ, Mucina L, Pacini E, Rotreklová O (2014) Ecological and evolutionary significance of genomic GC content diversity in monocots. Proc Natl Acad Sci USA. 111: E4096-E4102.

Szepesi Á, Csiszár J, Gémes K, Horváth E, Horváth F, Simon ML, Tari I (2009) Salicylic acid improves acclimation to salt stress by stimulating abscisic aldehyde oxidase activity and abscisic acid accumulation, and increases Na+ content in leaves without toxicity symptoms in Solanum lycopersicum L. J Plant Physiol. 166: 914-925.

Stanley L, Yuan YW (2019) Transcriptional regulation of carotenoid biosynthesis in plants: So many regulators, so little consensus. Front Plant Sci.10:1000-1017.

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 31:46-53.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 7:562-578.

Vogt T (2010) Phenylpropanoid biosynthesis. Mol Plant. 3:2-20.

Vom Endt D, Kijne JW, Memelink J (2002) Transcription factors controlling plant secondary metabolism: what regulates the regulators?. Pytochem. 61:107-114.

Wilhelm BT, Marguerat S, Goodhead I, Bähler J (2010) Defining transcribed regions using RNA-seq. Nat Protoc. 5:255-266.

Xiao M, Zhang Y, Chen X, Lee EJ, Barber CJ, Chakrabarty R, MacNevin G (2013) Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. J Biotechnol. 166:122-134.

Yang CQ, Fang X, Wu XM, Mao YB, Wang LJ, Chen XY (2012) Transcriptional Regulation of Plant Secondary Metabolism F. J Integr Plant Biol. 54:703-712.

Zhang YH, Zhang SD, Ling LZ (2015) De novo transcriptome analysis to identify flavonoid biosynthesis genes in Stellera chamaejasme. Plant Gene. 4:64-68.

Zhang X, Wollenweber B, Jiang D, Liu F, Zhao J (2008) Water deficits and heat shock effects on photosynthesis of a transgenic Arabidopsis thaliana constitutively expressing ABP9, a bZIP transcription factor. J Exp Bot. 59:839-848.

Zhang H, Stephanopoulos G (2016) Co-culture engineering for microbial biosynthesis of 3-amino-benzoic acid in Escherichia coli. Biotechnol J. 11:981-987.