**MANUSCRIPT**

**Applications in Plant Sciences**

**GENOMIC RESOURCES ARTICLE**

**PREMISE OF THE STUDY:** Maca (*Lepidium meyenii*; Brassicaceae) has been cultivated by Andeans for thousands of years as a food source and has been used for medicinal purposes. However, little is known about the mechanism underlying material accumulation during plant growth.

**METHODS:** RNA-Seq technology was used to compare the transcriptome of black maca root at three developmental stages. Gene Ontology term enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were applied for the identification of pathways in which differentially expressed genes were significantly enriched.

**RESULTS:** Trinity was used to de novo assemble the reads, and 120,664 unigenes were assembled. Of these, 71.53% of the unigenes were annotated based on BLAST. A total of 18,321 differentially expressed genes were observed. Gene Ontology term enrichment analysis found that the most highly represented pathway among the differentially expressed genes was for genes involved in starch and sucrose metabolism. We also found that genes involved in secondary metabolite biosynthesis, such as glucosinolate biosynthesis, were significantly enriched.

**DISCUSSION:** The genes that were differentially expressed between developmental time points likely reflect both developmental pathways and responses to changes in the environment. As such, the transcriptome data in this study serve as a reference for subsequent mining of genes that are involved in the synthesis of important bioactive components in maca.

**KEY WORDS:** Brassicaceae; differentially expressed genes; *Lepidium meyenii*; maca; RNA-Seq; secondary metabolite biosynthesis; transcriptome.
processes that are active or suppressed under certain conditions, such as metabolic or signaling pathways (Yang et al., 2015). In this study, we used genome-wide transcript profiling to predict the accumulation of compounds and metabolites during the development of maca root. Because the biosynthesis and accumulation of certain secondary metabolites are often stage- and tissue-specific (Salam et al., 2014; Yu et al., 2015), we analyzed several development stages of maca root, which we believe provides an important foundation for the identification of candidate genes acting in the mechanism underlying material accumulation during root growth.

MATERIALS AND METHODS

Black maca roots grown in Dahai, Huize County, Yunnan Province, China, were used in this study. From August 2012 to February 2013, we harvested 100 maca root specimens (from 100 individuals) every month to investigate the growth dynamics of root diameter and root fresh weight. The results showed that the growth curve of maca root was S-shaped (Appendix S1). Based on these findings, we divided the development of maca root into three stages: an early stage with slow growth (stage I), a middle stage with fast growth (stage II), and a late stage when growth slows and then stops (stage III). We then harvested maca roots in September 2015 (stage I), November 2015 (stage II), and January 2016 (stage III). For the RNA-Seq samples, we harvested 15 specimens (from 15 individuals) from each time point, which were randomly divided into three replicate groups, and the epidermis was quickly removed. The remaining tissue of each specimen was then cut into small pieces and stored in liquid nitrogen until later usage. Each group was considered as one biological repetition. Total RNA was isolated using TRIzol Reagent (Invitrogen, Carlsbad, California, USA) according to the manufacturer's instructions. DNase I (TaKaRa Bio Inc., Otsu, Shiga, Japan) was added to remove any genomic DNA contamination in the total RNA. Finally, the integrity and concentration of the RNA were measured using the Agilent 2100 Bioanalyzer system (Agilent Technologies, Santa Clara, California, USA).

We then isolated mRNA via oligo(dT) magnetic beads (Thermo Fisher Scientific, Waltham, Massachusetts, USA), followed by fragmentation using the Covaris M220 Focused-ultrasonicator (Covaris, Woburn, Massachusetts, USA), allowing for ~200 bp average length. The mRNA fragments were used as a template to generate first-strand cDNAs via random hexamer-primers, followed by second-strand cDNA synthesis. cDNA preparations were purified via the Agencourt AMPure XP kit (Beckman Coulter, Brea, California, USA). Adapters were ligated to the cDNA fragments after generating blunt-ends and 3'-addition of polyA tails. The fragments were then amplified by PCR. Strand-specific library preparation and paired-end sequencing (read length 125 nucleotides) were performed on the Illumina HiSeq 2500 platform (Illumina, San Diego, California, USA) by Berry Genomics (Beijing, China). Raw sequence reads were filtered using Trimmomatic version 0.30 (Bolger et al., 2014) to remove nonsense sequences including adapters, low-quality reads, reads containing poly-N, and other very short sequences. Clean reads were pooled to assemble a reference transcriptome with the Trinity software package (version 2.1.1) using the default parameters (Grabherr et al., 2011).

Unigenes were annotated via BLAST (Götz et al., 2008) by mining the following databases using an E-value cutoff of 10^-5: the National Center for Biotechnology Information (NCBI) non-redundant protein sequence database (nr), the NCBI non-redundant nucleotide (nt) database, the Gene Ontology (GO) database, the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, the euKaryotic Orthologous Groups (KOG) database, the Pfam database (a large collection of protein families), and SwissProt. Gene expression levels are based on FPKM (fragments per kilobase of transcript per million mapped reads). The quantification of fold changes and analysis of differential gene expression was carried out with Cufflink (Trapnell et al., 2012). The software edgeR was used to analyze the differential expression genes (Robinson et al., 2010). Differentially expressed genes between two samples were defined as having a log_2 ratio ≥1 and a false discovery rate (FDR) ≤0.001. Term enrichment analyses were carried out for GO and KEGG terms.

RESULTS

De novo transcriptome sequencing of maca root

To characterize the transcriptome of maca root and to generate expression profiles, we collected samples representing the three developmental stages of the plant as defined above, and sequenced corresponding cDNA samples using the Illumina HiSeq 2500. More than 162 million raw reads were generated from all three developmental stages, amounting to ~23.6 Gbp of data. The raw reads were submitted to the NCBI Sequence Read Archive (BioProject ID PRJNA450202, SRA accession SRP139986). The Transcriptome Shotgun Assembly project has been deposited at GenBank (accession GGN100000000), and the version described in this paper is the first version (GGN100000000). The de novo assembled transcriptome had 182,270 transcripts. The average GC content of the maca root transcriptome was 39.7%. The longest transcript of each gene was used as a unigene, and de novo assembly of the filtered reads was assembled into 120,664 unigenes. Transcript and unigene length ranged from 200 to >5000 bp.

Annotation and classification of gene function

Through analysis of GO databases, 50,303 unigenes were assigned to 56 categories (Fig. 1). A total of 46,237 unigenes were categorized into 25 groups by the KOG database (Fig. 2). “General functional prediction only” (“R” in Fig. 2) represented the major group, followed by “posttranslational modification, protein turnover, chaperones” (“O” in Fig. 2), and “signal transduction mechanisms” (“T” in Fig. 2). In this study, 30,539 unigenes were classified with the KEGG database. Among these, 9058 unigenes (7.5% of the total) were assigned to “organismal systems,” 18,657 (15.46%) to “metabolism,” 9034 (7.49%) to “genetic information processing,” 5995 (4.97%) to “environmental information processing,” and 5243 (4.35%) to “cellular processes” (Fig. 3). To optimize our efforts to functionally annotate all unigenes, 72,181, 70,191, 19,238, and 47,396 consensus sequences were annotated using the nr, nt, Pfam, and SwissProt databases, respectively (Table 1). Generally, a total of 71.53% L. meyenii unigenes were annotated, and the remaining 28.47% unigenes could not be annotated because of low or absent similarity to the sequences in the seven public databases.

We screened genes for significant differential expression between the three defined root stages. There were 6975 differentially expressed genes identified between stage III and stage I, with 4804 genes that were up-regulated and 2171 genes that were
Between stage III and stage II, we identified 12,508 differentially expressed genes, of which 6669 genes were up-regulated and 5839 genes were down-regulated. Finally, 12,682 differentially expressed genes were identified between stage III and stage I, with 7717 genes that were up-regulated and 4965 genes that were down-regulated. A comparison between the differentially expressed genes among the three stages is shown in Figure 4.

Analysis of differentially expressed genes

We performed term enrichment analyses for both GO and KEGG terms to characterize differentially expressed genes that may have potential roles in metabolic pathways. GO enrichment classification was carried out using the GOseq package (Guo et al., 2014). We considered GO terms to be significantly enriched when the corrected P value was <0.05. In summary, 9967 differentially expressed genes were classified into “biological process,” 4296 genes were classified into “cellular component,” and 4058 genes were classified into “molecular function.” The majority of genes involved in root growth and development were strongly enriched, such as cell proliferation (GO:0008283), cell wall organization (GO:0071555), and xyloglucan:xyloglucosyl transferase activity (GO:0016762). For term enrichment analysis of differentially expressed genes based on KEGG pathways, we used the KOBASE software package (Guo et al., 2014). We mapped differentially expressed genes to 288 KEGG pathways, and the most highly represented pathway was starch and
sucrose metabolism (Ko00500) (Fig. 5). Starch is the major carbohydrate of tuber and root crops (Hoover, 2001); sucrose and other sugars serve as substrates for starch production, and they also play a regulatory role in root development and various metabolic processes (Ravi et al., 2009). The larger part of maca root is formed by parenchyma, which is rich in starch and sugar (León, 1964). Dried maca root contains 59% carbohydrates and 8.5% fiber (Dini et al., 1994). The results from the term enrichment analysis suggest that during development maca root undergoes rapid cell division and proliferation followed by mass production of carbohydrates.

It is interesting to note that we also found that GO terms corresponding to stress responses, such as cold (GO:0009409) and wounding (GO:0009641), were significantly enriched. This may be caused by a gradual decrease in temperature starting in October, which explains the enrichment of cold-response GO terms as the maca root adapts to increasingly colder temperatures. It is also worth noting that some genes involved in secondary metabolite biosynthesis were significantly enriched, as they were associated with the terms “positive regulation of flavonoid biosynthesis” (GO:0009963) and “glucosinolate biosynthetic process” (GO:0019761). Additionally, glucosinolate biosynthesis (PATHWAY: ko00966) was one of the top 20 KEGG pathway enrichment terms (Fig. 5). This result is consistent with the analysis of the maca genome (Zhang et al., 2016), which showed that gene families involved in secondary metabolite biosynthesis (i.e., glucosinolate biosynthesis) underwent expansion in this species. Glucosinolates and their derivatives have attracted much attention because of their biological activities. For example, benzyl isothiocyanate is a potent inhibitor of breast cancer, stomach cancer, and liver cancer (Wattenberg, 1981; Rosa et al., 2010).

Maca root has a comparatively high concentration of glucosinolates, and the glucosinolate content of fresh maca root is approximately 100 times higher than that in other cruciferous crops such as cauliflower and cabbage (Li et al., 2001). Studies in animals suggest that maca root has a range of pharmacological functions, such as enhancing fatigue resistance (Ikeuchi et al., 2009), anticancer activity (Večeřa et al., 2007), and reducing prostate hyperplasia (Gonzales et al., 2008). These reported pharmacological activities were attributed to the high glucosinolate concentrations present in maca root. Therefore, it is necessary to further study the metabolism of glucosinolate biosynthesis and to use genomics to identify some of the key genes involved in its regulation.

FIGURE 3. Kyoto Encyclopedia of Genes and Genomes (KEGG) classification of maca root all-unigenes. A = cellular processes; B = environmental information processing; C = genetic information processing; D = metabolism; E = organismal systems.

FIGURE 4. Number of differentially expressed genes between three developmental stages in maca root. The statistical threshold was log2 ratio ≥ 1.

TABLE 1. Unigene functional annotation statistical results.

| Database    | No. of annotated sequences | Percentage of all unigenes (%) |
|-------------|---------------------------|-------------------------------|
| GO          | 50,303                    | 41.69                         |
| KEGG        | 30,539                    | 25.31                         |
| KOG         | 46,237                    | 38.32                         |
| nr          | 72,181                    | 59.82                         |
| nt          | 70,191                    | 58.17                         |
| Pfam        | 19,238                    | 15.94                         |
| SwissProt   | 47,396                    | 39.28                         |
| Annotated   | 86,310                    | 71.53                         |
| Total unigenes | 120,664                   | 100.00                        |

Note: GO = Gene Ontology database; KEGG = Kyoto Encyclopedia of Genes and Genomes database; KOG = euKaryotic Orthologous Groups database; nr = NCBI non-redundant protein sequence database; nt = NCBI non-redundant nucleotide database.

*E-value < 10^-5.
CONCLUSIONS

A total of 182,270 transcripts and 120,664 unigenes were identified from three developmental stages of black maca root using RNA-Seq and de novo assembly via the Trinity software package. A total of 18,321 genes were identified as differentially expressed in these three developmental stages. These genes fall into distinct molecular pathways and are closely related to regulating root development and secondary metabolite biosynthesis. To the best of our knowledge, this is the first study that investigated tissue-specific transcript profiles of black maca root during different developmental stages. Our results will provide a foundation for future investigations aimed at elucidating the molecular mechanisms of compound accumulation during maca root development.

ACKNOWLEDGMENTS

This work was supported by Fundamental Research Funds of the Chinese Academy of Forestry (CAFYBB2014QA015) and Fundamental Research Funds of the Research Institute of Resources Insects, Chinese Academy of Forestry (riricaf2013001Z). The authors thank Liping Shu for his assistance with the data analysis and Prof. Kirst King-Jones (University of Alberta) for his helpful comments.

DATA ACCESSIBILITY

The raw reads were submitted to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (BioProject ID PRJNA450202, SRA accession SRP139986). The Transcriptome Shotgun Assembly project has been deposited at GenBank (accession GGN10000000), and the version described in this paper is the first version (GGN10100000).

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

APPENDIX S1. The growth dynamics of diameter and fresh weight of maca root.

LITERATURE CITED

Beharry, S., and M. Heinrich. 2018. Is the hype around the reproductive health claims of maca (Lepidium meyenii Walp) justified? Journal of Ethnopharmacology 211: 126–170.
Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.

Campos, D., R. Chirinos, O. Barreto, G. Noratto, and R. Pedreschi. 2013. Optimized methodology for the simultaneous extraction of glucosinolates, phenolic compounds and antioxidant capacity from maca (*Lepidium meyenii*). *Industrial Crops and Products* 49: 747–754.

Choi, E. H., J. I. Kang, J. Y. Cho, S. H. Lee, T. S. Kim, I. H. Yeo, and H. S. Chun. 2012. Supplementation of standardized lipid-soluble extract from maca (*Lepidium meyenii*) increases swimming endurance capacity in rats. *Journal of Functional Foods* 4: 568–573.

Clément, C., J. Kneubuehler, A. Urryweiler, U. Witschi, and M. Kreuzer. 2010. Effect of maca supplementation on bovine sperm quantity and quality followed over two spermatogenetic cycles. *Theriogenology* 74: 173–183.

Dini, A., G. Migliuolo, L. Rastrelli, P. Saturnino, and O. Schettino. 1994. Chemical composition of *Lepidium meyenii*. *Food Chemistry* 49: 347–349.

Guo, Q., X. Ma, S. Wei, D. Qiu, I. W. Wilson, P. Wu, Q. Tang, et al. 2014. De novo transcriptome sequencing and digital gene expression analysis predict biosynthetic pathway of rynchophylline and isorhynchophylline from *Uncaria rhynchophylla*, a non-model plant with potent anti-jaundice properties. *BMC Genomics* 15: 676.

Hoover, R. 2001. Composition, molecular structure, and physicochemical properties of tuber and root starches: A review. *Carbohydrate Polymers* 45: 253–267.

Ikewuchi, M., T. Koyama, S. Takei, T. Kino, and K. Yuzawa. 2009. Effects of benzyl- glucosinolate on endurance capacity in mice. *Carbohydrate Polymers* 75: 255–262.

Maboreke, H. R., I. Feldhahn, M. Bönn, M. T. Tarkka, F. Buscot, S. Herrmann, R. Menzel, and L. Rues. 2016. Transcriptome analysis in oak uncovers a strong impact of endogenous rhythmic growth on the interaction with plant-parasitic nematodes. *BMC Genomics* 17: 627.

Ravi, V., S. K. Naskar, T. Makeshkumar, B. Babu, and B. S. P. Krishnan. 2009. Molecular physiology of storage root formation and development in sweet potato molecular physiology of storage root formation and development in sweet potato (*Ipomoea batatas* (L.) Lam.). *Journal of Root Crops* 35: 1–27.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140.

Rosa, E. A. S., R. K. Heaney, G. R. Fenwick, and C. A. M. Portas. 2010. Glucosinolates in crop plants. *Horticultural Reviews* 19: 99–215.

Rubio, J., W. Qiong, X. Liu, Z. Jiang, H. Dang, S. Chen, and G. F. Gonzales. 2011. Aqueous extract of black maca (*Lepidium meyenii*) on memory impairment induced by ovariectomy in mice. *Evidence-Based Complementary and Alternative Medicine* 2011: 239358.

Salam, J. S., F. Salam, K. S. Pothshamg, and D. B. Kumar. 2014. Effect of processing methods on secondary metabolites and enzyme inhibitors in different developmental stages of *Parkia rosbohghi* G. Don pods. *American Journal of Food Technology* 9: 89–96.

Somvanshi, V. S., S. Gahoi, P. Banakar, P. K. Thakur, M. Kumar, M. Sajinani, P. Pandey, and U. Rao. 2016. A transcriptomic insight into the infective juvenile stage of the insect parasitic nematode, *Heterorhabditis indica*. *BMC Genomics* 17: 166.

Trappnell, C., A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, et al. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* 7: 562–578.

Večevča, R., J. Orolin, N. Škottová, L. Kaxdová, O. Oliyarnik, J. Ulrichová, and V. Simánek. 2007. The influence of maca (*Lepidium meyenii*) on antioxidant status, lipid and glucose metabolism in rat. *Plant Foods for Human Nutrition* 62: 59–63.

Wang, Y., Y. Wang, B. McNeil, and L. M. Harvey. 2007. Maca: An Andean crop with multi-pharmacological functions. *Food Research International* 40: 783–792.

Wattenberg, L. W. 1981. Inhibition of carcinogen-induced neoplasia by sodium cyanate, tert-butyl isocyanate, and benzyl isothiocyanate administered sub- sequent to carcinogen exposure. *Cancer Research* 41: 2991–2994.

Wippler, J., M. Kleiner, C. Lott, A. Gruhl, P. E. Abraham, R. J. Giannone, J. C. Young, et al. 2016. Transcriptomic and proteomic insights into innate immunity and adaptations to a symbiotic lifestyle in the gutless marine worm *Olavius algarensis*. *BMC Genomics* 17: 942.

Yang, Y., G. Yao, W. Yue, S. Zhang, and J. Wu. 2015. Transcriptome profiling reveals differential gene expression in proanthocyanidin biosynthesis associated with red/green skin color mutant of pear (*Pyrus communis* L.). *Frontiers in Plant Science* 6: 795.

Yu, M., F. Liu, W. Zhu, M. Sun, J. Liu, and X. Li. 2015. New features of triacylglycerol biosynthetic pathways of peanut seeds in early developmental stages. *Functional and Integrative Genomics* 15: 707–716.

Zhang, J., Y. Tian, L. Yan, G. Zhang, X. Wang, Y. Zeng, J. Zhang, et al. 2016. Genome of plant maca (*Lepidium meyenii*) illuminates genomic basis for high-altitude adaptation in the Central Andes. *Molecular Plant* 9: 1066–1077.