Data in Brief

Gene expression profiling of ramie roots during hydroponic induction and adaption to aquatic environment

Gang Gao, Heping Xiong, Kunmei Chen, Jikang Chen, Ping Chen, Chunming Yu⁎, Aiguo Zhu⁎

Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha 410205, China

A B S T R A C T

Ramie (Boehmeria nivea (L.) Gaud.) is a traditionally terrestrial fiber crop. However, hydroponic technology can enhance the quantity and quality of disease free Ramie plant seedlings for field cultivation. To date, few studies have attempted to examine the hydroponic induction of ramie roots and the molecular responses of ramie roots to aquatic environment. In this study, ramie tender stems was grown in the soil or in a hydroponic water solution, and cultured in the same environmental conditions. Root samples of terrestrial ramie, and different developmental stages of hydroponic ramie (5 days, 30 days), were firstly pooled for reference transcriptome sequencing by Illumina HiSeq 2000. Gene expression levels of each samples were quantified using the BGISEQ500 platform to help understand the distribution of aquatic root development related genes at the macro level (GSE98903). Our data resources provided an opportunity to elucidate the adaptation mechanisms of ramie seedlings roots in aquatic environment.

Specifications [standardized info for the reader]

| Organism/cell line/tissue | Terrestrial or aquatic root tissues of Boehmeria nivea (L.) Gaud |
|---------------------------|---------------------------------------------------------------|
| Sex                       | N/A                                                           |
| Sequencer or array type   | Illumina HiSeq 2000 and BGISEQ500 platform                   |
| Data format               | Raw data                                                      |
| Experimental factors      | Hydroponic induction of ramie roots                           |
| Experimental features     | Gene expression profiling of terrestrial and aquatic roots. The root samples from hydroponic ramie were collected from 5-day-old seedlings (the early stage of root induction) and 30-day-old seedlings (the late stage of root induction). The roots of ramie seedlings in soil were collected for comparative analysis. |
| Consent                   | N/A                                                           |
| Sample source location    | Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha, Hunan, China (GPS coordination: 112.907991, 28.217469) |

1. Direct link to deposited data

Deposited data can be found here: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE98903.

2. Introduction

Ramie (Boehmeria nivea (L.) Gaud.) is a commercially important plant species that is predominantly used for fabric production [1,2]. Due to the low survival rate of seedlings in the early stages, propagation by conventional means has been met with many difficulties. Hydroponic culture [3] is an excellent alternative method for the large scale and high quality production of this valuable crop. Several physiological and morphological investigations have been conducted to improve our understanding of the hydroponic culture of ramie. Here we report a de novo transcriptome assembly of ramie roots and gene expression profiling of ramie roots in an aquatic environment. Our aim was to obtain a high quality reference transcriptome of ramie root, elucidate the molecular response of ramie to aquatic environment, and find candidate waterlogging tolerance genes.
3. Experimental design, materials and methods

3.1. Ramie cultivation and RNA sample preparation

The elite ramie cultivar “Zhongzhu 2”, which can be hydroponically cultivated was used in this study. Ramie seedlings were cultivated in either soil or in a hydroponic environment with using the shoot-cutting propagation method. The root samples from hydroponic ramie were collected from the early (5-day-old seedlings) and late (30-day-old seedlings) stages of aquatic roots induction. The roots of ramie seedlings in soil were also collected for comparative analysis. The sample tissues were immediately frozen in liquid nitrogen and stored at −80 °C until use. In order to increase the root transcriptome coverage, a mixture of the three samples were firstly pooled for RNA sequencing.

Total RNA was extracted using TRIzol Reagent (Invitrogen, LifeTechnologies, USA) following the manufacturer’s instructions, then treated with DNase I (Invitrogen, Life Technologies, USA). The RNA integrity was verified using an Agilent 2100 BioAnalyzer (Agilent, USA).

3.2. RNA sequencing

RNA-Seq libraries were constructed using the RNA Library Prep Kit for Illumina according to the manufacturer’s instructions (NEB, USA). Library quality was assessed on the Agilent Bioanalyzer 2100 system. The libraries were sequenced on the Illumina HiSeq 2000 platform (Illumina, USA) based on sequencing by synthesis with 100 bp paired-end reads (Biomarker Technologies, Beijing). Whole RNA-Seq data were submitted to NCBI Sequence Read Archive and Gene expression Omnibus (series accession number GSE98903).

3.3. Root transcriptome assembly and gene functional annotation

Prior to assembly, the raw data were cleaned by trimming adaptor sequences and removing low quality sequences (Q < 20) with > 10% uncertain (N) bases using in-house perl scripts. These clean reads were then de novo assembled into unigenes using the short reads assembling program Trinity with min kmer cov set to 2 and all other parameters set to default [4,5]. The following databases were used to annotate the gene function: Nr (NCBI non-redundant protein sequences, NCBI blast 2.2.28+, e-value = 1e−5); Nt (NCBI non-redundant nucleotide sequences, NCBI blast 2.2.28+, e-value = 1e−5); Pfam (Protein family, http://pfam.sanger.ac.uk/, HMMER 3.0 package, hmmscan, e-value = 0.01); KOG/COG (Clusters of Orthologous Groups of proteins, http://www.ncbi.nlm.nih.gov/COG/, NCBI blast 2.2.28+, e-value = 1e−3); Swiss Prot (a manually annotated and reviewed protein sequence database, http://www.ebi.ac.uk/uniprot/, NCBI blast 2.2.28+, e-value = 1e−10); and GO (Gene Ontology, http://www.geneontology.org/).

Table 1

| Index                        | Ramie roots |
|------------------------------|-------------|
| Number of assembled reads    | 66,209      |
| Number of unigenes           | 43,541      |
| Average unigene length (bp)  | 966         |
| Unigenes annotated in Nr     | 31,066      |
| Unigenes annotated in Nt     | 27,378      |
| Unigenes annotated in Swiss Prot | 22,887    |
| Unigenes annotated in KEEG   | 24,522      |
| Unigenes annotated in COG    | 15,513      |
| Unigenes annotated in Interpro | 24,071     |
| Unigenes annotated in GO     | 18,486      |

Fig. 1. Heatmap of correlation coefficient values across ramie root samples by growth condition and developmental stages. Gradient color barcode at the right top indicates the minimum value in white and the maximum in blue. If one sample is in highly similar with another one, the correlation value between them is very close to 1.
In this study, we obtained a total of 6.76 Gb raw data for root transcriptome sequencing. When all samples were assembled, 43,541 Unigenes were acquired. The total length, average length, N50, and GC content of Unigenes were 42,081,259 bp, 966 bp, 1667 bp, and 45.86%, respectively. The Unigenes were then annotated with 7 functional databases, finally, 31,066 (NR: 71.35%), 27,378 (NT: 62.88%), 22,887 (Swiss Prot: 52.56%), 15,513 (COG: 35.63%), 24,522 (KEGG: 56.32%), 18,486 (GO:42.46%), and 24,071 (Interpro: 55.28%) Unigenes were annotated. The information on the transcriptome sequencing and assembly is summarised in Table 1.

3.4. Comparative expression analysis of aquatic roots development related genes

In total, 9 RNA samples of ramie root (early stage 1,2,3; late stage 1,2,3, and terrestrial 1,2,3) were sequenced using the BGIseq500 platform (BGI, Wuhan, China, http://www.seq500.com/en/) with 3 biological replicates, and averagely generating 24,136,832 raw sequencing reads. After filtering, the clean reads were mapped to reference using HISAT/Bowtie2 tool [6,7]. Gene expression level was quantified by RSEM [8], and the common genes among samples were displayed by Venn Charts (Fig. 1). NOISeq method was used to screen differentially expressed genes (Fig. 2). Cluster and java Treeview [9] were used to perform clustering analysis of the DEGs (Fig. 3).

The 9 samples were firstly grouped as early stage (labeled 1), late stage (labeled 2) and terrestrial (label 3) so that pairwise comparison between every two groups could be done later. Gene expression levels were calculated and expressed as Reads Per Kilobase per Million reads mapped. Expression data were evaluated considering RPKM values. The log fold changes were considered as significant when the weight of a sample was at least one-fold higher or lower than another, with an FDR corrected p-value ≤ 0.05. The analysis was limited to genes showing RPKM ≥ 1 in at least one sample, as used in other studies. Annotation analysis of Gene Ontology (GO) and GO functional classification were done to help understand the distribution of gene functions of the DEGs from macro view.

Conflit of interest

Authors declare no conflict of interest.

Acknowledgements

This research work was supported by the “Agricultural Science and Technology Innovation Project of Chinese (CAAS-ASTIP-2017)” and “China Agricultural Research System (no. CARS-16).”

References

[1] A.K. Mohanty, M. Misra, G. Hinrichsen, Biobfibres, biodegradable polymers and bio- composites: an overview, Macromol. Mater. Eng. 276–277 (1) (2000) 1–24.
[2] K. Goda, M.S. Sreekala, A. Gomez, T. Kaji, J. Ohgi, Improvement of plant based natural fibers for toughening green composites—effect of load application during mercerization of ramie fibers, Compos. A: Appl. Sci. Manuf. 37 (12) (2006) 2213–2220.
[3] T.J. Crees, B. Hocking, M. Dayod, B. Xu, A. Ahman, S. Henderson, et al., Protocoal: optimising hydroponic growth systems for nutritional and physiological analysis of Arabidopsis thaliana and other plants, Plant Methods 9 (1) (2013) 4.
[4] M.G. Grabherr, B.I. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, et al., Full-length transcriptome assembly from RNA-Seq data without a reference genome, Nat. Biotechnol. 29 (7) (2011) 644–652.
[5] G. Pertea, X. Huang, P. Lang, V. Antonescu, R. Sultana, S. Karamycheva, et al., TGIF Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets, Bioinformatics (Oxford, England) 19 (5) (2003) 651–652.
[6] D. Kim, R. Langmead, S.L. Salzberg, HISAT: a fast spliced aligner with low memory re- quirements, Nat. Methods 12 (4) (2015) 357–360.
[7] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol. 10 (3) (2009) R25.
[8] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, BMC Bioinf. 12 (2011) 323.

[9] S. Tarazona, F. Garcia-Alcalde, J. Dopazo, A. Ferrer, A. Conesa, Differential expression in RNA-seq: a matter of depth, Genome Res. 21 (12) (2011) 2213–2223.