data

MDPI

Data Descriptor

Transcriptome Dataset of Leaf Tissue in *Agave* H11648

Xing Huang [1], Li Xie [2], Thomas Gbokie Jr. [3], Jingen Xi [1] and Kexian Yi [1,*]

[1] Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China; hxalong@gmail.com (X.H.); xijingen@163.com (J.X.)
[2] College of Forestry, Hainan University, Haikou 570228, China; m13178981326@163.com
[3] College of Plant Protection, Nanjing Agricultural University, Nanjing 210095, China; gbokiejr@gmail.com
* Correspondence: yikexian@126.com

Received: 14 April 2019; Accepted: 6 May 2019; Published: 6 May 2019

Abstract: Sisal is widely cultivated in tropical areas for fiber production. The main sisal cultivar, *Agave* H11648 ((*A. amaniensis* × *A. angustifolia*) × *A. amaniensis*) has a relatively scarce molecular basis and no genomic information. Next-generation sequencing technology has offered a great opportunity for functional gene mining in *Agave* species. Several published *Agave* transcriptomes have already been reused for gene cloning and selection pressure analysis. There are also other potential uses of the published transcriptomes, such as meta-analysis, molecular marker detection, alternative splicing analysis, multi-omics analysis, genome assembly, weighted gene co-expression network analysis, expression quantitative trait loci analysis, miRNA target site prediction, etc. In order to make the best of our published transcriptome of *A.* H11648 leaf, we here represent a data descriptor, with the aim to expand *Agave* bio information and benefit *Agave* genetic researches.

Dataset: https://www.ncbi.nlm.nih.gov/sra/SRP132128

Dataset License: CC-BY

Keywords: *Agave* H11648; leaf; transcriptome

1. Introduction

Sisal is an important fiber crop in tropical areas around the world [1]. The main sisal cultivar is *Agave* H11648 ((*A. amaniensis* × *A. angustifolia*) × *A. amaniensis*), which has been widely cultivated in American, African and Asian countries [2]. The long life-cycle of *A.* H11648 has significantly restricted genetic improvement by traditional breeding, which makes plant biotechnology an efficient way to improve its fiber quality and yield [3]. The molecular basis of *Agave* species is lacking compared with model plants [4]. The large genomes also challenge researchers to reveal the *Agave* secret [5]. In recent years, the fast development of next-generation sequencing has brought an efficient method for gene mining in minor crops [6]. Till now, next-generation sequencing has been successfully carried out in several *Agave* species, which revealed the transcriptome dynamics of different tissues in *A. deserti* and *A. tequilana*, crassulacean acid metabolism (CAM) photosynthesis in *A. americana*, shoot organogenesis in *A. salmiana* and drought stress response in *A. sisalana* [7–10]. As the leaf is the main vegetative part above ground and used for fiber production, we have conducted the transcriptome analysis of *A.* H11648 leaf as a reference for gene mining in our previous study [11]. Several full-length cellulose synthase genes were cloned in *A. deserti*, *A. tequilana*, *A. americana* and *A.* H11648 according to their transcriptomes, which provides an approach for the reuse of published *Agave* transcriptomes. Besides, these datasets can also be used for selection pressure analysis to estimate the domestication patterns of *Agave* species [12]. Here, we present a data descriptor of *A.* H11648 leaf transcriptome dataset, in

order to make the best of it. The dataset aimed to expand *Agave* bio information and benefit *Agave* genetic researches.

2. Results

2.1. Illumina Sequencing and De Novo Assembly

A. H11648 leaf samples were collected for RNA isolation and library construction. Illumina paired-end sequencing generated 60,791,648 raw reads, from which 49,252,060 clean reads were filtered. 98.97% and 96.10% clean bases had quality scores above 20 and 30, respectively (Figure 1). The GC content was 48.86% (Figure 2) and the error rate was 0.0117%. *De novo* assembly generated 148,046 unigenes and the total length was 76,779,911 base pairs (bp). The mean length, median length and N50 length were 518.63 bp, 330 bp and 591 bp, respectively. Fragments per kilobase of exon per million reads mapped (FPKM) values were calculated to estimate the expression patterns of each unigenes (Table S1). Among these, there were 41405 (27.97%), 44598 (30.12%), 46351 (31.31%), 12016 (8.12%) and 3676 (2.48%) unigenes within the FPKM values ranging from 0–1, 2–3, 4–15, 16–60 and >60, respectively (Figure 3).

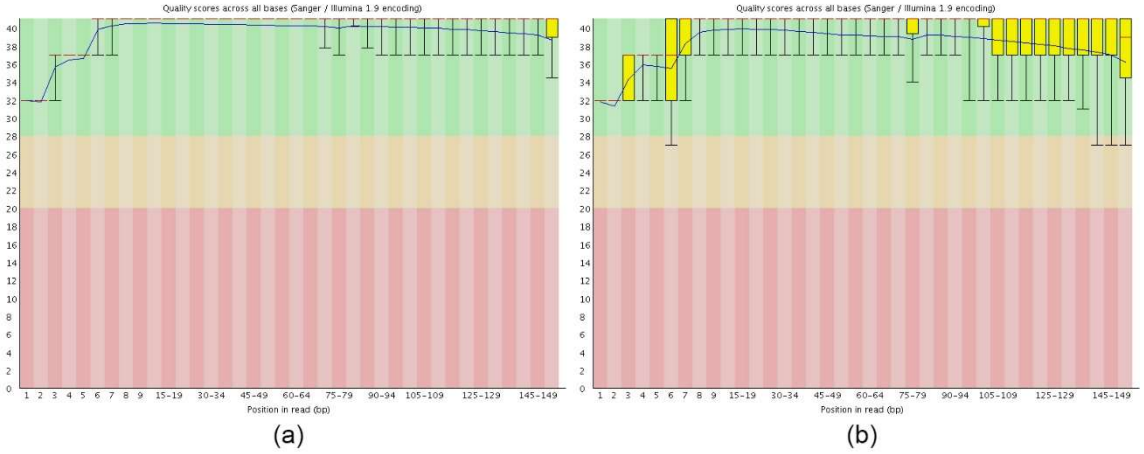

Figure 1. Quality scores of the left-end (**a**) and the right-end (**b**) raw reads. Blue and red lines represent median and mean scores, respectively.

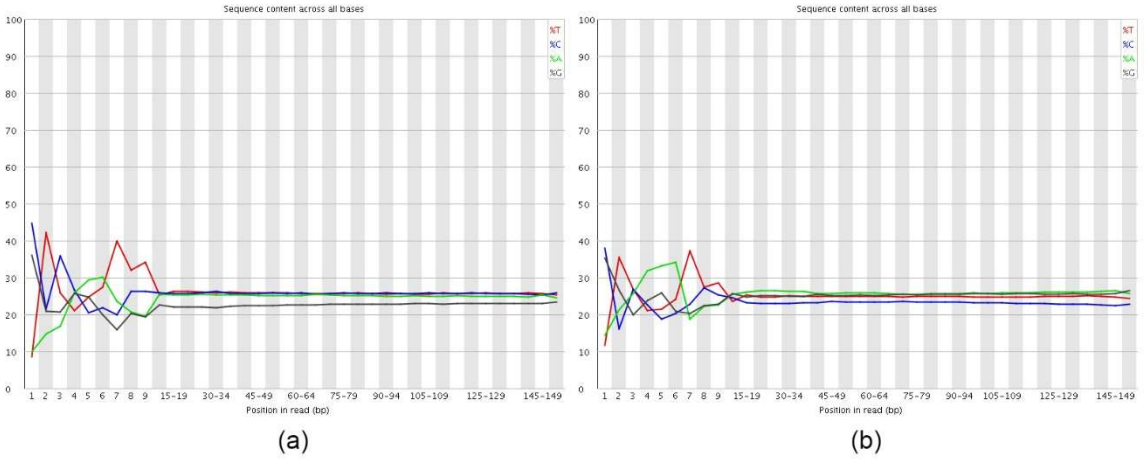

Figure 2. Nucleotide bases distribution of left-end (**a**) and right-end (**b**) raw reads.

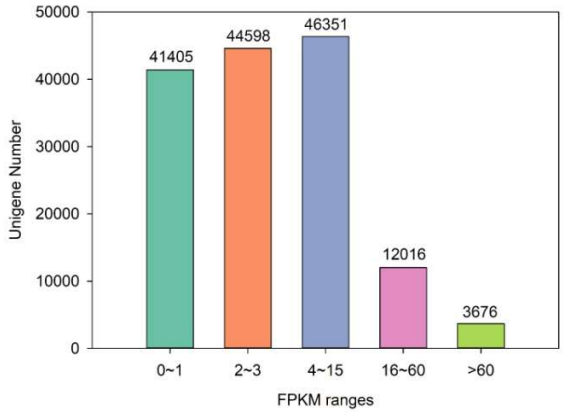

Figure 3. Count numbers of unigenes at different FPKM ranges.

2.2. Value of the Data

This dataset was primarily established as a reference transcriptome for gene mining in *A*. H11648, which also provided an important resource for molecular biology and genetic studies in *Agave* species. There were a series of potential uses with the dataset, such as meta-analysis, gene cloning, selection pressure analysis, molecular marker detection, alternative splicing analysis, multi-omics analysis, genome assembly, weighted gene co-expression network analysis, expression quantitative trait loci analysis, miRNA target site prediction, etc. [8,11–17].

2.3. Data Records

The raw data have been deposited to Sequence Read Archive (SRA) under the accession of SRP132128. The BioProject, BioSample and SRA ID are PRJNA432160, SAMN08435960 and SRR6668799, respectively.

3. Materials and Methods

3.1. Plant Material and RNA Isolation

The *A*. H11648 plants have been planted in Wenchang experimental field (19°32′19″ N 110°46′08″ E) of Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences since 2013. Developing leaves were selected for sampling from the upper portion of 3-year-old plants. The distal parts of leaves (leaf length < 30 cm) were cut for sampling. Leaf samples were collected from three individuals for RNA isolation. Total RNAs were isolated with a Tiangen RNA Prep Pure Plant Kit (Tiangen Biomart, Beijing, China) according to the instruction of the manufacturer's protocol.

3.2. Library Construction and Illumina Sequencing

Equal mass of the three RNA samples were mixed together and sent to Genoseq Technology Co. Ltd (Wuhan, Hubei, China) for next-generation sequencing. Ten micrograms of RNA were used for cDNA library construction [11,18]. The mRNAs were obtained after purification with poly-T oligo-attached magnetic beads, which were subjected for fragmentation with TruSeq RNA Sample Prep Kit (Illumina, San Diego, CA, USA). Random hexamer primer and M-MuLV Reverse Transcriptase (RNase H) were used for first-strand cDNA synthesis. The second-strand cDNA was synthesized by DNA Polymerase I and RNase H. The ends of these double-stranded cDNA fragments were modified with a single A base and adaptors. After gel purification, adaptor-attached fragments were utilized for PCR amplification, with the aim to construct a cDNA library. Illumina sequencing was carried out with the Illumina HiSeq platform to generate 150 bp paired-end raw reads.

3.3. Data Processing

The quality of all reads was evaluated by FastQC software [19]. Adaptor sequences were removed by Cutadapt [20]. Low-quality sequences were filtered by Trimmomatic [21]. The clean data were subjected to Trinity for *de novo* transcriptome assembly [22]. Gene expression levels were estimated by RNA-Seq by Expectation-Maximization (RSEM) and normalized to FPKM [23,24].

Supplementary Materials: The following are available online at http://www.mdpi.com/2306-5729/4/2/62/s1, Table S1: FPKM values of unigenes in *Agave* H11648.

Author Contributions: Conceptualization: X.H., J.X. and K.Y.; Formal analysis: X.H.; Funding acquisition: X.H., J.X. and K.Y.; Investigation: X.H., L.X. and T.G.; Supervision: X.H. and K.Y.; Writing—original draft: X.H.; Writing—review and revise: T.G.

Funding: This research was funded by National Key R&D Program of China (2018YFD0201100), the earmarked fund for China Agriculture Research System (CARS-16-E16), Central Public-interest Scientific Institution Basal Research Fund for Chinese Academy of Tropical Agricultural Sciences (1630042019012, 1630042019041) and Hainan Provincial Natural Science Foundation of China (319QN275).

Acknowledgments: We would like to thank Xiaohan Yang from Oak Ridge National Laboratory (Oak Ridge, TN 37831, USA) for his suggestions on the design of experiments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, Y.; Mai, Y.W.; Ye, L. Sisal fibre and its composites: A review of recent developments. *Compos. Sci. Tech.* **2000**, *60*, 2037–2055. [CrossRef]
2. Food and Agriculture Organization of the United Nations (FAO). Available online: http://www.fao.org/faostat (accessed on 30 March 2019).
3. Gao, J.; Yang, F.; Zhang, S.; Li, J.; Chen, H.; Liu, Q.; Zheng, J.; Xi, J.; Yi, K. Expression of a hevein-like gene in transgenic *Agave* hybrid No. 11648 enhances tolerance against zebra stripe disease. *Plant Cell Tissue Organ Culture* **2014**, *119*, 579–585. [CrossRef]
4. Nava-Cruz, N.Y.; Medina-Morales, M.A.; Martinez, J.L.; Rodriguez, R.; Aguilar, C.N. *Agave* biotechnology: An overview. *Crit. Rev. Biotechnol.* **2015**, *35*, 546–559. [CrossRef] [PubMed]
5. Robert, M.L.; Lim, K.Y.; Hanson, L.; Sanchez-Teyer, F.; Bennett, M.D.; Leitch, A.R.; Leitch, I.J. Wild and agronomically important *Agave* species (Asparagaceae) show proportional increases in chromosome number, genome size, and genetic markers with increasing ploidy. *Bot. J. Linn. Soc.* **2010**, *158*, 215–222. [CrossRef]
6. Schuster, S.C. Next-generation sequencing transforms today's biology. *Nat. Methods* **2008**, *5*, 16–18. [CrossRef] [PubMed]
7. Gross, S.M.; Martin, J.A.; Simpson, J.; Abraham-Juarez, M.J.; Wang, Z.; Visel, A. *De novo* transcriptome assembly of drought tolerant CAM plants, *Agave deserti* and *Agave tequilana*. *BMC Genom.* **2013**, *14*, 563. [CrossRef]
8. Abraham, P.E.; Yin, H.; Borland, A.M.; Weighill, D.; Lim, S.D.; De Paoli, H.C.; Engle, N.; Jones, P.C.; Agh, R.; Weston, D.J.; et al. Transcript, protein and metabolite temporal dynamics in the CAM plant *Agave*. *Nat. Plants* **2016**, *2*, 16178. [CrossRef] [PubMed]
9. Cervantes-Pérez, S.A.; Espinal-Centeno, A.; Oropeza-Aburto, A.; Caballero-Pérez, J.; Falcon, F.; Aragón-Raygoza, A.; Sánchez-Segura, L.; Herrera-Estrella, L.; Cruz-Hernández, A.; Cruz-Ramírez, A. Transcriptional profiling of the CAM plant *Agave salmiana* reveals conservation of a genetic program for regeneration. *Dev. Biol.* **2018**, *442*, 28–39. [CrossRef]
10. Sarwar, M.B.; Ahmad, Z.; Rashid, B.; Hassan, S.; Gregersen, P.L.; Leyva, M.O.; Nagy, I.; Asp, T.; Husnain, T. De novo assembly of Agave sisalana transcriptome in response to drought stress provides insight into the tolerance mechanisms. *Sci. Rep.* **2019**, *9*, 396.
11. Huang, X.; Xiao, M.; Xi, J.; He, C.; Zheng, J.; Chen, H.; Gao, J.; Zhang, S.; Wu, W.; Liang, Y.; et al. *De novo* transcriptome assembly of *Agave* H11648 by Illumina sequencing and identification of cellulose synthase genes in *Agave* species. *Genes* **2019**, *10*, 103. [CrossRef]

12. Huang, X.; Wang, B.; Xi, J.; Zhang, Y.; He, C.; Zheng, J.; Gao, J.; Chen, H.; Zhang, S.; Wu, W.; et al. Transcriptome comparison reveals distinct selection patterns in domesticated and wild *Agave* species, the important CAM plants. *Int. J. Genom.* **2018**, *2018*, 5716518. [CrossRef] [PubMed]
13. Smith, M.L.; Glass, G.V. Meta-analysis of psychotherapy outcome studies. *Am. Psychol.* **1977**, *32*, 752–760. [CrossRef] [PubMed]
14. Luo, X.; Xu, L.; Liang, D.; Wang, Y.; Zhang, W.; Zhu, X.; Zhu, Y.; Jiang, H.; Tang, M.; Liu, L. Comparative transcriptomics uncovers alternative splicing and molecular marker development in radish (*Raphanus sativus* L.). *BMC Genom.* **2017**, *18*, 505. [CrossRef] [PubMed]
15. Harkess, A.; Zhou, J.; Xu, C.; Bowers, J.E.; Van der Hulst, R.; Ayyampalayam, S.; Mercati, F.; Riccardi, P.; McKain, M.R.; Kakrana, A.; et al. The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nat. Commun.* **2017**, *8*, 1279. [CrossRef]
16. Verta, J.P.; Landry, C.R.; MacKay, J. Dissection of expression-quantitative trait locus and allele specificity using a haploid/diploid plant system—insights into compensatory evolution of transcriptional regulation within populations. *New Phytol.* **2016**, *211*, 159–171. [CrossRef] [PubMed]
17. Singh, N.K. miRNAs target databases: Developmental methods and target identification techniques with functional annotations. *Cell Mol. Life Sci.* **2017**, *74*, 2239–2261. [CrossRef] [PubMed]
18. Huang, X.; Chen, J.; Bao, Y.; Liu, L.; Jiang, H.; An, X.; Dai, L.; Wang, B.; Peng, D. Transcript profiling reveals auxin and cytokinin signaling pathways and transcription regulation during in vitro organogenesis of ramie (*Boehmeria nivea* L. Gaud). *PLoS ONE* **2014**, *9*, e113768. [CrossRef] [PubMed]
19. FastQC. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 30 March 2019).
20. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **2011**, *17*, 10–12. [CrossRef]
21. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
22. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [CrossRef]
23. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [CrossRef] [PubMed]
24. Mortazavi, A.; Williams, B.A.; Mccue, K.; Schaeffer, L.; Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **2008**, *5*, 621–628. [CrossRef] [PubMed]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).