

Brief Report

Transcriptome Assembly of the Bast Fiber Crop, Ramie, *Boehmeria nivea* (L.) Gaud. (Urticaceae)

Louay Al-Ani ^{1,2} and Michael K. Deyholos ^{3,*} 

¹ Department of Biology, University of Anbar, Ramadi 31001, Iraq; alani@ualberta.ca

² Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada

³ Department of Biology, University of British Columbia, Okanagan Campus, Kelowna, BC V1V 1V7, Canada

* Correspondence: michael.deyholos@ubc.ca; Tel.: +1-250-807-8541

Received: 31 December 2017; Accepted: 24 January 2018; Published: 1 February 2018

Abstract: Ramie (*Boehmeria nivea*) is a perennial crop valued for its strong bast fibers. Unlike other major bast fiber crops, ramie fiber processing does not include retting, but does require degumming, suggesting distinctive features in pectin and the development and composition of fibers. A comprehensive transcriptome assembly of ramie has not been made available, to date. We obtained the sequence of RNA transcripts (RNA Seq) from the apical region of developing ramie stems and combined these with reads from public databases for a total of 157,621,051 paired-end reads (30.3 billion base pairs Gbp) used as input for de novo assembly, resulting in 70,721 scaffolds (≥ 200 base pairs (bp); $N_{50} = 1798$ bp). As evidence of the quality of the assembly, 36,535 scaffolds aligned to at least one Arabidopsis protein (BLASTP e-value $\leq 10^{-10}$). The resource described here for *B. nivea* will facilitate an improved understanding of bast fibers, cell wall, and middle lamella development in this and other comparative species.

Keywords: bast; fiber; pectin; ramie; transcriptome; Urticaceae

1. Introduction

Ramie (*Boehmeria nivea* (L.) Gaud.) is a perennial, herbaceous, semi-tropical species in the Urticaceae family. It has been grown for at least 6000 years in Asia. Approximately 70,000 ha are now cultivated worldwide (primarily in China), with 124,000 tonnes harvested in 2013 (<http://faostat3.fao.org/>) [1]. Ramie is valued primarily for its long, strong phloem (i.e., bast) fibers, which are used in textiles, most often as a blend with other types of fibers. There is also growing interest in other uses of the fibers, such as in composite materials [2]. Ramie fibers have similar properties to other commercial bast fibers [3,4]. While linen and hemp harvesting rely on the microbial retting (degradation) of pectins, ramie is normally extracted by decorticating fresh stems to produce bark ribbons that are subsequently dried, scraped, and de-gummed. Extracts of various parts of the ramie plant have also been reported to have beneficial medicinal properties, including antiviral and anti-inflammatory activities [5–7].

Research on ramie has included descriptions of specific genes including transcription factors, expansins, and cellulose synthases, as well as transcriptome-level responses related to water deficit, heavy metals, nematodes, hormones, and developmental processes [8–14]. However, no assembled transcriptome of ramie has been reported to date. We present here an assembly of all available transcript sequence reads of the *B. nivea* cultivar Zhongzhu 1, including novel RNA sequencing reads generated in our laboratory from the apex of developing shoots.

2. Materials and Methods

Seeds of *B. nivea* cv. Zhongzhu 1 were obtained from the Institute of Bast Fiber Crops (Changsha, China), and grown under natural light in a glass house at the University of Alberta

(Edmonton, AB, Canada). When plants were approximately 1 m high, the top 20 cm of stems of three plants were harvested. Leaves that were larger than 10 mm (including petioles) were removed, the stem segments were homogenized in liquid nitrogen, and RNA was extracted using a modified CTAB extraction combined with an RNeasy Kit (Qiagen, Valencia, CA, USA) [15]. A total of 10 µg of RNA was sent to the Beijing Genome Institute (BGI) (BGI Inc., Shenzhen, China) for library preparation and paired-end transcriptome sequencing, as previously described [16]. Briefly, poly-A⁺ mRNA was isolated using oligodT coupled magnetic beads, and this mRNA was used as a template for random hexamer-primed first strand cDNA synthesis, followed by second strand synthesis using *E. coli* DNA Pol I. Double-stranded cDNA was sheared with a nebulizer, end repaired, and ligated to Illumina PE adapter oligos. Subsequently, 200 base pair (bp) fragments were selected by gel purification, and these were PCR-amplified for 15 cycles sequencing using an Illumina Genome Analyzer II (San Diego, CA, USA) with 75 bp, paired-end reads. Raw reads from all runs were filtered to remove adapter sequences, contamination, and low-quality reads, and the filtered raw reads were deposited in the SRA archive as ERR364387.

The length of sequence read substrings used in assembly is defined as the k-mer. Reads were used as input for de novo assembly using a k-mer sweep ($k = 35, 39, 45, 49, 55, 59, 65, 69$), and all of the reads except ERR364387 were additionally used with $k = 75, 79, 85, 89$ [17,18]. The scaffold file output for each k-mer was merged with Trans-Abyss, which is a software package that uses de Bruijn graphs to assemble sequence reads [19]. The resulting merged scaffolds were further assembled using CAP3 [19] (minimum overlap < 100 bp), which joins together fragments based on shared sequences. This was followed by redundancy reduction with CD-HIT-EST, which clusters together similar sequences (0.98% identity, [20]). Reads from each of the original libraries were mapped to the assembled scaffolds using default parameters with Bowtie 2, which uses an implementation of the Burrows-Wheeler Transform to align reads to a reference sequence.

3. Results

3.1. Assembly

RNA was obtained from the apical-most 20 cm of stems of *B. nivea*. This segment included xylem and phloem fibers in a gradient of developmental stages, ranging from specification to fiber cell wall thickening. The RNA molecules were reverse transcribed and the resulting DNA was sequenced using paired-end technology, which allows sequences to be obtained from both ends of each DNA molecule. In total, 813 Mbp (mega base pairs) of sequence were obtained from 5,571,079 PE (paired-end) reads of 75 bp (base pairs) each, from which low-quality reads and adaptors had been removed, and the reads were submitted to the NCBI Sequence Read Archive (SRA) as accession ERR364387. These reads were used along with other publicly available reads to provide a total of 30.3 Gbp (giga base pairs) of input for the assembly (Table 1).

Using ABySS, all of the reads in Table 1 were used as input for de novo assembly using a k-mer sweep, with the output merged using Trans-Abyss [17], followed by further assembly in CAP3 [19] (minimum overlaps <100 bp). Redundant assembly units were removed using CD-HIT-EST, [20], resulting in a final assembly that contained 70,721 scaffolds with a minimum of 200 bp in length, with a scaffold N₅₀ of 1798 bp, and maximum scaffold length of 22,363 bp (Table 2, Supplemental File 1).

To evaluate the quality of the assembly, reads from each of the original libraries (Table 1) were mapped to the assembled scaffolds using Bowtie 2, resulting in overall alignment rates ranging between 78% (for ERR364387) and 98% (for SRR546782) [21]. The distribution of scaffolds lengths in the assembly is shown in Figure 1.

Table 1. Libraries used in transcriptome assembly.

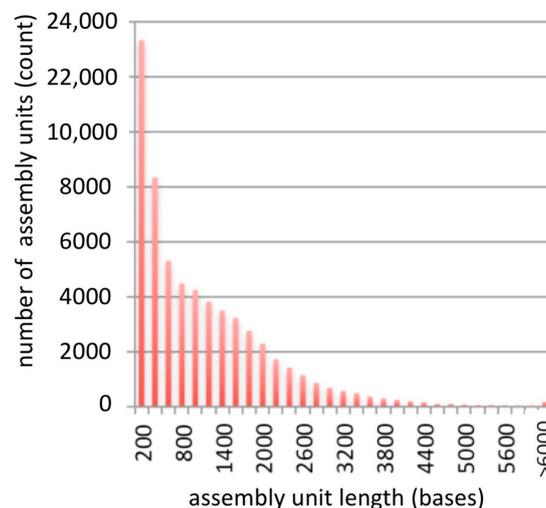
Tissue	SRA ¹ Accession	Number of Paired- End Reads	Read Length (bp ²)	Total Reads (Gbp ³)	Reference
shoot apex with young leaves and woody stem	ERR364387	5,571,079	73	0.813	this study
whole plant excluding root	SRR546782	26,457,905	90	4.76	[22]
phloem	SRR1380980	13,513,706	90	2.43	[10]
xylem	SRR1380981	13,865,114	90	2.50	[10]
whole plant excluding root	SRR1021565	28,165,784	101	5.69	[7]
whole plant excluding root	SRR1021566	25,846,434	101	5.22	[7]
whole plant excluding root; roots were infected by nematodes	SRR1021567	21,701,137	101	4.38	[7]
whole plant excluding root; roots were infected by nematodes	SRR1021568	22,499,892	101	4.54	[7]

¹ SRA (short read archive); ² bp (base pairs); ³ Gbp (giga base pairs).

Table 2. Transcriptome de novo assembly statistical summary.

Descriptor	Value
Total clean paired reads	157,621,051
Assembled scaffolds ¹	70,271
Longest scaffold (bp)	22,363
Smallest scaffold (bp)	200
N ₅₀ length (bp)	1798
N ₅₀ count	14,016
N ₉₀ length (bp)	459
N ₉₀ count	43,828
Total scaffolds length (bp) ¹	77,022,208

¹ Only scaffolds with a length of ≥ 200 bp are counted here.

**Figure 1.** Length distribution of assembled scaffolds.

3.2. Analysis of Assembly Quality

To further evaluate the assembly, BLASTX was used to align 70,721 scaffolds to the complete set of predicted Arabidopsis proteins [23,24]. As a result, 36,535 scaffolds aligned to at least one Arabidopsis protein, with a minimum significance of e-value 10^{-10} . Conversely, at the same significance threshold, 20,835 out of 27,416 Arabidopsis proteins aligned to at least one scaffold from the *B. nivea* assembly. The top Arabidopsis match for each of the BLASTX alignments was used with the Gene Ontology (GO) Annotation Search, Functional Categorization, and Download Tool to categorize the scaffolds

according to their cellular component, molecular function, and biological process (<https://www.arabidopsis.org/tools/bulk/go/index.jsp>). The distribution of GO categories assigned to the *B. nivea* scaffolds was generally similar to the distribution within the Arabidopsis proteins, indicating good coverage of the *B. nivea* genome by this assembly (Figure 2).

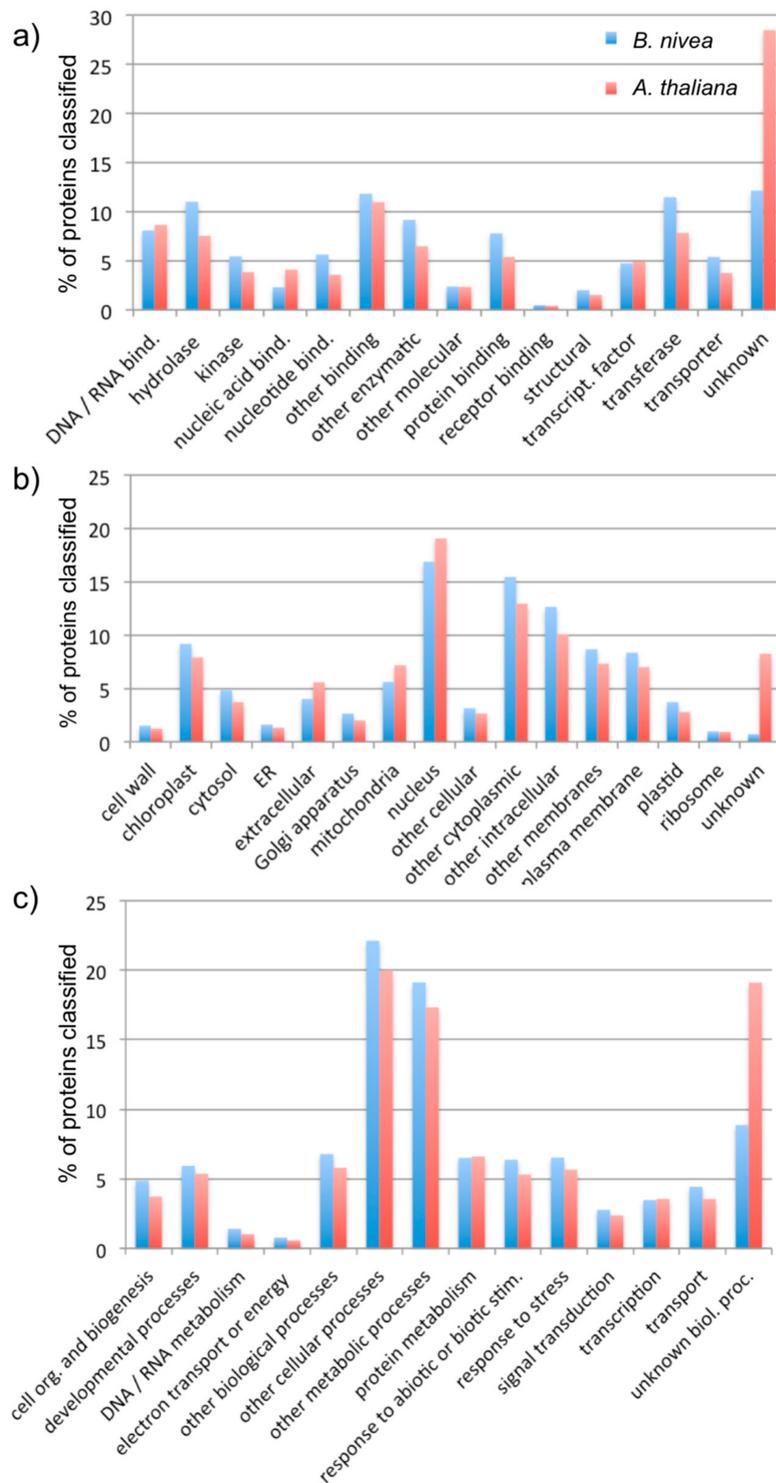


Figure 2. Gene Ontology (GO) categorization of predicted proteins among *B. nivea* scaffolds. The figure represents the proportion of proteins assigned to each category within the three aspects of (a) molecular function, (b) cellular compartment, and (c) biological process for both *B. nivea* and *A. thaliana*.

4. Discussion

In evaluating the quality of this assembly, we considered first the number and length of scaffolds, where each scaffold presumably represents a transcript of a gene (Figure 1, Table 1). The N_{50} value of this assembly is 1798 bp, meaning that 50% of the assembly is made of transcripts with 1798 bp or larger. This is longer than the median 1089 bp gene coding sequence length reported for land plants, suggesting that many of the assembled scaffolds represent full-length transcripts [25]. Furthermore, there were 43,828 transcripts in the assembly that were more than 459 bp long (Table 1), which is within the range of the number of proteins typically found in plant genomes [26]. Finally, we considered the ability to assign sequences to functional categories, and found that these generally matched the distribution of functions in the model plant species, *A. thaliana* (Figure 2), suggesting that the *B. nivea* assembly we presented accurately reflects the genes encoded in these species.

5. Conclusions

We have described the first public transcriptome assembly of the bast fiber crop *B. nivea* (ramie) from 30.3 Gbp of paired-end reads resulting in 70,721 scaffolds (≥ 200 bp; $N_{50} = 1798$ bp), and have demonstrated the quality of this assembly through a comparison of its predicted proteins to the Arabidopsis reference proteome, using both BLASTX alignment and GO categorization. These data will facilitate further research in bast fiber crops including ramie.

Supplementary Materials: The following are available online at www.mdpi.com/2079-6439/6/1/8/s1, File S1: Bniv200.fna.txt.

Acknowledgments: Funding was provided by the Natural Sciences and Engineering Research Council (Canada) and Genome Alberta (One Thousand Plants 1KP Project).

Author Contributions: L.A.-A. and M.K.D. conceived and designed the experiments and analyzed the data. M.K.D. wrote the paper.

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

References

- Liao, L.; Li, T.J.; Zhang, J.; Xu, L.L.; Deng, H.S.; Han, X.J. The domestication and dispersal of the cultivated ramie (*Boehmeria nivea* (L.) Gaud. in Freyc.) determined by nuclear SSR marker analysis. *Genet. Resour. Crop Evol.* **2014**, *61*, 55–67. [[CrossRef](#)]
- Wang, H.; Xian, G.; Li, H.; Sui, L. Durability study of a ramie-fiber reinforced phenolic composite subjected to water immersion. *Fibers Polym.* **2014**, *15*, 1029–1034. [[CrossRef](#)]
- Li, J.; Feng, J.H.; Zhang, H.; Zhang, J.C. Wear properties of hemp, ramie and linen fabrics after liquid ammonia/crosslinking treatment. *Fibres Text. East. Eur.* **2010**, *18*, 81–85.
- Summerscales, J.; Gwinnett, C. Forensic Identification of Bast Fibres. In *Biocomposites for High-Performance Applications*; Ray, D., Ed.; Elsevier B.V.: Amsterdam, The Netherlands, 2017; pp. 125–164.
- Huang, K.-L.; Lai, Y.-K.; Lin, C.-C.; Chang, J.-M. Inhibition of hepatitis B virus production by *Boehmeria nivea* root extract in HepG2 2.2.15 cells. *World J. Gastroenterol.* **2006**, *12*, 5721–5725. [[CrossRef](#)] [[PubMed](#)]
- Paiva, P.; Nonato, F.; Zafred, R.; Sousa, I.; Ruiz, A.L.; Foglio, M.A.; Carvalho, J.E. Anti-inflammatory activity of *Boehmeria caudata* sw. (Urticaceae) crude ethanol extract on croton oil mouse ear oedema. *Planta Medica* **2014**, *80*, 1528. [[CrossRef](#)]
- Sung, M.J.; Davaatseren, M.; Kim, S.H.; Kim, M.J.; Hwang, J.-T. *Boehmeria nivea* attenuates LPS-induced inflammatory markers by inhibiting p38 and jnk phosphorylations in raw264.7 macrophages. *Pharm. Biol.* **2013**, *51*, 1131–1136. [[CrossRef](#)] [[PubMed](#)]
- Zhu, S.; Tang, S.; Tang, Q.; Liu, T. Genome-wide transcriptional changes of ramie (*Boehmeria nivea* L. Gaud) in response to root-lesion nematode infection. *Gene* **2014**, *552*, 67–74. [[CrossRef](#)] [[PubMed](#)]
- Liu, T.; Zhu, S.; Tang, Q.; Tang, S. Genome-wide transcriptomic profiling of ramie (*Boehmeria nivea* L. Gaud) in response to cadmium stress. *Gene* **2015**, *558*, 131–137. [[CrossRef](#)] [[PubMed](#)]

10. 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*. [[CrossRef](#)] [[PubMed](#)]
11. Chen, J.; Liu, F.; Tang, Y.; Yuan, Y.; Guo, Q. Transcriptome sequencing and profiling of expressed genes in phloem and xylem of ramie (*Boehmeria nivea* L. Gaud). *PLoS ONE* **2014**, *9*. [[CrossRef](#)] [[PubMed](#)]
12. Liu, T.; Tang, S.; Zhu, S.; Tang, Q.; Zheng, X. Transcriptome comparison reveals the patterns of selection in domesticated and wild ramie (*Boehmeria nivea* L. Gaud). *Plant Mol. Biol.* **2014**, *86*, 85–92. [[CrossRef](#)] [[PubMed](#)]
13. Staff, P.O. Transcriptome sequencing and profiling of expressed genes in phloem and xylem of ramie (*Boehmeria nivea* L. Gaud). *PLoS ONE* **2015**, *10*. [[CrossRef](#)]
14. Chen, J.; Pei, Z.; Dai, L.; Wang, B.; Liu, L.; An, X.; Peng, D. Transcriptome profiling using pyrosequencing shows genes associated with bast fiber development in ramie (*Boehmeria nivea* L.). *BMC Genom.* **2014**, *15*. [[CrossRef](#)] [[PubMed](#)]
15. Guerriero, G.; Mangeot-Peter, L.; Hausman, J.-F.; Legay, S. Extraction of High Quality RNA from Cannabis sativa Bast Fibres: A Vademecum for Molecular Biologists. *Fibers* **2016**, *4*, 23. [[CrossRef](#)]
16. Zhang, N.; Deyholos, M.K. RNA-Seq analysis of the shoot apex of flax (*Linum usitatissimum*) to identify phloem fiber specification genes. *Front. Plant Sci.* **2016**, *7*. [[CrossRef](#)]
17. Robertson, G.; Schein, J.; Chiu, R.; Corbett, R.; Field, M.; Jackman, S.D.; Mungall, K.; Lee, S.; Okada, H.M.; Qian, J.Q.; et al. De novo assembly and analysis of RNA-seq data. *Nat. Methods* **2010**, *7*, 909–912. [[CrossRef](#)] [[PubMed](#)]
18. Simpson, J.T.; Wong, K.; Jackman, S.D.; Schein, J.E.; Jones, S.J.M.; Birol, I. ABYSS: A parallel assembler for short read sequence data. *Genome Res.* **2009**, *19*, 1117–1123. [[CrossRef](#)] [[PubMed](#)]
19. Huang, X.; Madan, A. CAP3: A DNA sequence assembly program. *Genome Res.* **1999**, *9*, 868–877. [[CrossRef](#)] [[PubMed](#)]
20. Li, W.; Godzik, A. CD-HIT: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [[CrossRef](#)] [[PubMed](#)]
21. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)] [[PubMed](#)]
22. Liu, T.; Zhu, S.; Tang, Q.; Chen, P.; Yu, Y.; Tang, S. De novo assembly and characterization of transcriptome using illumina paired-end sequencing and identification of cesa gene in ramie (*Boehmeria nivea* L. Gaud). *BMC Genom.* **2013**, *14*. [[CrossRef](#)] [[PubMed](#)]
23. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
24. Lamesch, P.; Berardini, T.Z.; Li, D.H.; Swarbreck, D.; Wilks, C.; Sasidharan, R.; Muller, R.; Dreher, K.; Alexander, D.L.; Garcia-Hernandez, M.; et al. The arabidopsis information resource (tair): Improved gene annotation and new tools. *Nucleic Acids Res.* **2012**, *40*, D1202–D1210. [[CrossRef](#)] [[PubMed](#)]
25. Ramírez-Sánchez, O.; Pérez-Rodríguez, P.; Delaye, L.; Tiessen, A. Plant Proteins Are Smaller Because They Are Encoded by Fewer Exons than Animal Proteins. *Genom. Proteom. Bioinform.* **2016**, *14*, 357–370.
26. Michael, T.P.; Jackson, S. The First 50 Plant Genomes. *Plant Genome* **2013**, *6*. [[CrossRef](#)]

