Draft genome of the herbaceous bamboo *Raddia distichophylla*

Wei Li,*¹ Cong Shi,†¹ Kui Li,†¹ Qun-jie Zhang,* Yan Tong,† Yun Zhang,† Jun Wang,‡
Lynn Clark,§² and Li-zhi Gao,*†²

*Institution of Genomics and Bioinformatics, South China Agricultural University, Guangzhou 510642, China, †Plant Germplasm and Genomics Center, Germplasm Bank of Wild Species in Southwestern China, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650204, China, ‡Southwest China Forestry University, Kunming 650224, China, and §Ecology, Evolution and Organismal Biology, Iowa State University, 251 Bessey Hall, Ames, 50011-1020, IA, USA.

¹These authors contributed equally to this work.

²Correspond author: E-mail address: Lgaogenomics@163.com & Lgclark@iastate.edu

© The Author(s) (2021). Published by Oxford University Press on behalf of the Genetics Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-Non-Commercial-NoDerivs licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
ABSTRACT

Bamboos are important non-timber forest plants widely distributed in the tropical and subtropical regions of Asia, Africa, America, and Pacific islands. They comprise the Bambusoideae in the grass family (Poaceae), including approximately 1,700 described species in 127 genera. In spite of the widespread uses of bamboo for food, construction and bioenergy, the gene repertoire of bamboo still remains largely unexplored. *Raddia distichophylla* (Schrad. ex Nees) Chase, belonging to the tribe Olyrae (Bambusoideae, Poaceae), is a diploid herbaceous bamboo with only slightly lignified stems. In this study, we report a draft genome assembly of the ~589 Mb whole-genome sequence of *R. distichophylla* with a contig N50 length of 86.36 Kb. Repeat sequences account for ~49.08% of the genome assembly, of which LTR retrotransposons occupy ~35.99% of the whole genome. A total of 30,763 protein-coding genes were annotated in the *R. distichophylla* genome with an average transcript size of 2,887 bp. Access to this herbaceous bamboo genome sequence will provide novel insights into biochemistry, molecular marker-assisted breeding programs and germplasm conservation for bamboo species world-wide.

KEYWORDS

Bamboos, *Raddia distichophylla*, whole-genome sequencing
INTRODUCTION

Bamboos are important non-timber forest plants with a wide native geographic distribution in tropical, subtropical and temperate regions except Europe and Antarctica (BambooPhylogenyGroup, 2012). Bamboos are of notable economic and cultural significance worldwide, and can be broadly used as food, bioenergy and building materials. Bamboos comprise the Bambusoideae in the grass family (Poaceae), including approximately 1,700 described species in 127 genera (Clark and Oliveira, 2018; Soreng et al. 2017; Vorontsova et al. 2016). Molecular phylogenetic analysis suggested that Bambusoideae falls into the Bambusoideae-Oryzoideae-Pooideae (BOP) clade, which is phylogenetically sister to Pooideae (Saarela et al. 2018).

Bambusoideae may be divided into two morphologically distinct growth forms: woody bamboos and herbaceous bamboos (tribe Olyreae); woody bamboos can be further divided into two lineages: temperate woody (Arundinarieae) and tropical woody (Bambuseae) (BambooPhylogenyGroup, 2012; Kelchner and Group, 2013; Sungkaew et al. 2009). The tribe Olyreae comprises 22 genera and 124 described species native to South America except the genus Buergersiochloa and Olyra latifolia (BambooPhylogenyGroup, 2012; Clark et al. 2015). Herbaceous bamboos are characterized by usually weakly developed rhizomes and less lignification in the culms. Culm leaves and foliage leaves with the outer ligule are absent in herbaceous bamboos. In contrast to woody bamboos, herbaceous bamboos have at least functionally unisexual flowers and they flower annually or seasonally for extended periods (Clark et al. 2015; Gaut et al. 1997; Kelchner and Group, 2013; Oliveira et al. 2014; Wysocki et al. 2015). The tribe is fundamentally diploid, but chromosome counts indicating tetraploidy or hexaploidy or possibly even octoploidy (in Eremitis genus) are available (Judziewicz et al. 1999; Soderstrom, 1981). The genus Raddia belongs to the Olyrinae, in which R. distichophylla (Schrad. ex Nees) Chase is a representative species and almost completely restricted to the forests of eastern Brazil (Oliveira et al. 2014). The species is a perennial plant growing in dense tufts. R.
*distichophylla* is a monoecious plant with male and female spikelets in different inflorescences. It is delicate and much smaller than woody bamboos in height (usually 12-35 cm long) (Plants of the World online, [http://powo.science.kew.org/](http://powo.science.kew.org/)) (Figure 1). In the last twenty-five years, *R. distichophylla* has seriously been threatened by a rapid deforestation and the conversion of cacao plantations (Giulietti *et al.* 2005). Besides, *R. distichophylla* is narrow endemic that has further led to a reduced effective population size.

Since the first comparative DNA sequence analysis of bamboos by Kelchner and Clark (1997), a number of studies have been further carried out through combining various biotechnology platforms (Das *et al.* 2005; Gui *et al.* 2010; Oliveira *et al.* 2014; Peng *et al.* 2013; Sharma *et al.* 2008; Sungkaew *et al.* 2009; Wysocki *et al.* 2016; Zhang *et al.* 2011). Peng *et al.* (2013) reported the first draft genome of tetraploid moso bamboo (*Phyllostachys edulis*, 2n = 4x = 48). Recently, Guo *et al.* (2019) have released four draft bamboo genomes, including *Olyra latifolia* (2n = 2x = 22), *Raddia guianensis* (2n = 2x = 22), *Guadua angustifolia* (2n = 4x = 46), and *Bonia amplexicaulis* (2n = 6x = 72). Whole-genome sequencing of herbaceous bamboo is limited to *R. guianensis* with a relatively fragmented assembly (contig N50 = 11.45 Kb, and scaffold N50 = 12.09 Kb) and a relatively low BUSCO completeness rate (~72.0%). Thus, the lack of a high-quality genome sequence for the diploid bamboo has been an impediment to our understanding of the bamboo diversification and evolution. It is recognized that genomics allows novel insights into the evolutionary history of species and offers basic information for taking efficient conservation strategies (Silva-Junior *et al.* 2018). In this study, we generated a draft genome assembly of *R. distichophylla* through sequencing on an Illumina HiSeq 2000 platform. The availability of the fully sequenced and annotated genome assembly will provide functional, ecological and evolutionary insights into the bamboo species and greatly enhance conservation genetics of this endangered species.
METHODS AND MATERIALS

Sample collection, total DNA and RNA extraction and sequencing

The source plant was an individual of *R. distichophylla* grown in cultivation at the R.W. Pohl Conservatory, Iowa State University. Fresh and healthy leaves were harvested and immediately frozen in liquid nitrogen, followed by storage at -80°C in the laboratory prior to DNA extraction. A modified CTAB method (Porebski *et al.* 1997) was used to extract high-quality genomic DNA. The quantity and quality of the extracted DNA were examined using a NanoDrop D-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE) and electrophoresis on a 0.8% agarose gel, respectively. A total of 3 paired-end and 6 mate-pair sequencing libraries, spanning 180, 300, 500, 2000, 5000, 10000 and 20000 bp, were prepared using Illumina’s paired-end and mate-pair kits, respectively (Illumina, San Diego, CA). Over 5 µg genomic DNA was fragmented by nebulization with compressed nitrogen gas for paired-end libraries. About 10-30 µg of high-quality genomic DNA was required for the long-insert mate-pair libraries. The DNA fragments were circularized by self-ligation, while after the digestion of linear DNA, circularized DNA was again fragmented. The fragmented DNA was purified using streptavidin-coated magnetic beads before adapter ligation. After quality control and concentration estimation of DNA samples with an Agilent 2100 bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), libraries were sequenced on Illumina HiSeq 2000 platform.

Total RNA was extracted from four tissues (root, stem, young leaf and female inflorescence), using a Water Saturated Phenol method. RNA libraries were built using the Illumina RNA-Seq kit (mRNA-Seq Sample Prep Kit P/N 1004814). The extracted RNA was quantified using NanoDrop-1000 UV-VIS spectrophotometer (NanoDrop), and RNA integrity was checked using Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). For each tissue, only total RNAs with a total amount ≥ 15 µg with a concentration ≥ 400 ng/µl, RNA integrity number (RIN) ≥ 7, and rRNA ratio ≥ 1.4 were used for constructing cDNA library according to the manufacturer’s instruction (Illumina, USA). The libraries were then sequenced (100
nt, paired-end) with Illumina HiSeq 2000 platform.

**De novo assembly of the *R. distichophylla* genome**

Two orthogonal methods were used to estimate the genome size of *R. distichophylla*, including *k*-mer frequency distribution and flow cytometric analysis. Firstly, we generated the 17-mer occurrence distribution of sequencing reads using GCE v1.0.0 (settings: -m 1 -D 8 -H 1) (Liu *et al.* 2013), and the genome size was then calculated with the equation

\[ G = \frac{N \times (L - K + 1)}{L \times D} \]

where *G* represents the genome size; *N* is the total number of bases from sequencing data; *L* is the average length of reads; *K* is set to 17; and *D* indicates expected coverage depth for bases. Secondly, the genome size was further estimated and validated using flow cytometry analysis. We employed the rice cultivar *Nipponbare* as an inner standard with an estimated genome size of ~389 Mb (IRGSP 2005).

Paired-end sequencing reads were processed to remove adaptors and low-quality sequences using Trimmomatic v0.33 (Bolger *et al.* 2014). Reads were retained only if both paired reads passed quality control filtering. We assembled the clean reads using Platanus v1.2.1 software (Rei Kajitani, 2014), which is optimized for highly heterozygous diploid genomes. First, the high-quality paired-end Illumina reads from short-insert size libraries (\(\leq 500 \text{ bp}\)) were assembled into contigs using Platanus. The initial *K*-mer size was set 37, *K*-mer coverage cutoff was 2, and the step size was 10. The assembled contigs were then scaffolded using SSPACE v3.0 (settings: -k 5 -x 0 -g 3 -a 0.7) (Boetzer *et al.* 2011) using Illumina mate pair data. GapCloser v1.12 (settings: -a scaffolds.fasta -b reads.lib -o gapcloser_scaffolds.fasta -l 149 -t 4) (Li *et al.* 2010) was finally used to fill gaps within the scaffolds with the paired-end sequencing data.

Three methods were employed to assess the quality and completeness of the *R. distichophylla* genome. First, high-quality reads from short-insert size libraries were mapped to the genome assembly using BWA v0.7.15 with default parameters for paired-end reads (Li and Durbin, 2009). Second, the genome assembly was checked with Benchmarking Universal Single-Copy Orthologs (BUSCO) (Simao *et al.* 2015).
Third, the RNA sequencing reads generated in this study were assembled using Trinity v r20131110 with default parameters (Grabherr et al. 2011), the assembled transcripts were then aligned back to the assembled genome using GMAP v2014-10-2 (Wu and Watanabe, 2005) at 30% coverage and 80% identity thresholds.

**Annotation of repetitive sequences and non-coding RNA genes**

We used a combination of homology-based and de novo approaches to identify the repetitive sequences in the *R. distichophylla* genome. RepeatModeler v1.0.10 (Tarailo-Graovac and Chen, 2009), which included two de novo repeat finding programs, RECON (Bao and Eddy, 2002) and RepeatScout (Price et al. 2005), was used for the construction of the repeat library. This produced library, along with the Poaceae repeat library, were used as the reference database for RepeatMasker (Tarailo-Graovac and Chen, 2009). Simple sequence repeats (SSRs) were identified in the genome sequence using the MISA perl script (Thiel et al. 2003) with the default settings: monomer (one nucleotide, \( n \geq 12 \)), dimer (two nucleotides, \( n \geq 6 \)), trimer (three nucleotides, \( n \geq 4 \)), tetramer (four nucleotides, \( n \geq 3 \)), pentamer (five nucleotides, \( n \geq 3 \)), and hexamer (six nucleotides, \( n \geq 3 \)).

Non-coding RNA genes play important roles in many cellular processes. The five different types of non-coding RNA genes, namely transfer RNA (tRNA) genes, ribosomal RNA (rRNA) genes, small nucleolar RNA (snoRNAs) genes, small nuclear RNA (snRNAs) genes and microRNA (miRNAs) genes, were predicted using various de novo and homology search methods. We used tRNAscan-SE algorithms (version 1.23) (Lowe and Eddy, 1997) with default parameters to identify the tRNA genes. The rRNA genes (8S, 18S, and 28S), which is the RNA component of the ribosome, were predicted by using RNAmmer algorithms (v1.2) (Lagesen et al. 2007) with default parameters. The snoRNA genes were annotated using snoScan v1.0 (Lowe and Eddy, 1999) with the yeast rRNA methylation sites and yeast rRNA sequences provided by the snoScan distribution. The snRNA genes were identified by INFERNAL software (v1.1.2) (Nawrocki et al. 2009) against the Rfam database (release 9.1) with default parameters.
**Genome annotation**

The gene prediction pipeline combined the *de novo* method, the homology-based method and the transcriptome-based method. Augustus v2.5.5 (Stanke *et al.* 2004) and Fgenesh (Salamov and Solovyev, 2000) were used to perform the *de novo* prediction. To improve the quality of gene prediction, we performed self-training with Augustus. RNA-seq reads were *de novo* assembled using Trinity and refined with PASA (Haas *et al.* 2008) to produce additional genome-guided transcriptome assemblies. Manual curation was performed with the training set, genes were retained if: (1) they have the complete gene structure without inner stop codons; (2) they have multiple exons and the CDS length exceed 800 bp. CD-Hit (Li and Godzik, 2006) was used to remove the training set with over 70% sequence similarity. The protein sequences of moso bamboo ([http://server.ncgr.ac.cn/bamboo/](http://server.ncgr.ac.cn/bamboo/)) (Peng *et al.* 2013), stiff brome (GenBank, assembly accession GCA_000005505.4) (TheInternationalBrachypodiumInitiative, 2010), barley (https://webblast.ipk-gatersleben.de/barley_ibsc/) (Mascher *et al.* 2017), maize (http://ensembl.gramene.org/Zea_mays/Info/Index) (Jiao *et al.* 2017), *Oropetium thomaeum* (GenBank, assembly accession LFJQ00000000) (VanBuren *et al.* 2015), foxtail millet (GenBank, assembly accession AGTC01000000) (Zhang *et al.* 2012), rice (http://rice.plantbiology.msu.edu/index.shtml) (IRGSP 2005) and sorghum (GenBank, assembly accession QWKM00000000) (Deschamps *et al.* 2018) were mapped to the genome using Exonerate (settings: genome2protein option) (Slater and Birney, 2005). To further aid the gene annotation, Illumina RNA-seq reads were assembled using the Trinity software (v20131110) with default parameters (Grabherr *et al.* 2011). The resulting transcripts were then aligned to the soft-masked genome assembly using GMAP v2014-10-2 (Wu and Watanabe, 2005) and BLAT v35 (Kent, 2002). The potential gene structures were derived using PASA v20130907 (Program to Assemble Spliced Alignments) (Haas *et al.* 2003). All gene models produced by the *de novo*, homology-based and transcriptome-based methods were integrated using GLEAN (Elsik *et al.* 2007).

The predicted genes were searched against Swiss-Prot database (Boeckmann *et al.* 2004).
2003) using BLASTP (e-value cutoff of $10^{-5}$). The motifs and domains within gene models were identified by InterProScan (Jones et al. 2014). Gene Ontology terms and KEGG pathway for each gene were retrieved from the corresponding InterPro entry. Gene functions were also assigned with TrEMBL database (Boeckmann et al. 2003) using BLASP with an e-value threshold of $10^{-5}$.

**Data availability**

All sequencing reads have been deposited in the NCBI Sequence Read Archive SRR8759078 to SRR8759084 (2019) (under accession number PRJNA528150) and BIG Genome Sequence Archive CRR049770 to CRR049776 (2019) (under accession number PRJCA001330). The assembled genome sequence is available at the NCBI and BIG Genome Warehouse under accession number SPJY00000000 and GWHAAKD00000000, respectively. Gene prediction and peptide fasta of *R. distichophylla* may also be accessed through the BIG Genome Warehouse under accession number GWHAAKD00000000. Figure S1 presents evolutionary history of TE super-families in the *R. distichophylla* and moso bamboo genomes. Figure S2 shows comparisons of gene features among *R. distichophylla* and three other plant species. Table S1 shows the whole genome sequencing (WGS) reads used to assemble the *R. distichophylla* genome. Table S2 shows the summary of RNA sequencing (RNA-Seq) of *R. distichophylla*. Table S3 shows the summary of genome assembly. Table S4 shows the validation of the *R. distichophylla* genome assembly using reads mapping and transcript alignments. Table S5 shows the assessment of the *R. distichophylla* genome assembly using BUSCO. Table S6 shows the statistics of repeat sequences in the *R. distichophylla* and moso bamboo genomes. Table S7 shows the statistics of typical transposable elements between *R. distichophylla* and *R. guianensis*. Table S8 shows the summary of types and number of simple sequence repeats in the *R. distichophylla* and moso bamboo genomes. Table S9 shows the non-coding RNA genes in the *R. distichophylla* genome. Table S10 shows the statistics of predicted protein-coding genes in the *R. distichophylla* genome. Table S11 shows the functional annotation of the *R. distichophylla* protein-coding genes.
Supplemental material available at figshare.

RESULTS

We performed whole-genome sequencing with the Illumina sequencing platform. A total of 253.94 Gb short sequencing reads were generated (~272.89-fold coverage) (Table S1). A total of 21.99 Gb RNA-seq data was obtained from root, stem, young leaf and female inflorescence (Table S2). Based on the K-mer analysis, we estimated the genome size of *R. distichophylla* to be ~608 Mb (Figure 2). Flow cytometry analysis estimated the genome size of *R. distichophylla* to be ~589 Mb, which is close to the obtained result from the k-mer analysis. The final assembly amounted to ~580.85 Mb, representing for 95.56% of the estimated genome size. The N50 lengths of the assembled contigs and scaffolds were ~86.36 Kb and ~1.81 Mb, respectively (Table 1; Table S3). The contig N50 and scaffold N50 sizes represent ~7.14-fold and ~150.83-fold improvement compared with the previously reported *R. guianensis* genome assembly (Guo et al. 2019), respectively.

To assess the genome assembly quality, we first mapped ~211 Mb of high-quality reads to the genome sequences. Our results revealed that nearly 89.25% Illumina reads were mapped to the genome assembly (Table S4); second, BUSCO was used to assess the completeness of the genome assembly. The percentage of completeness for our assembly was 92.08% in the Embryophyta lineage (Table S5); and finally, we mapped the assembled transcripts to the genome sequences. Approximately 78.04% of the transcripts could be mapped to the genome (Table S4).

The annotation of repeat sequences showed that approximately 49.08% of the *R. distichophylla* genome consists of transposable elements (TEs), lower than the amount (63.15%) annotated in the moso bamboo genome (Peng et al. 2013) with the same methods (Table S6; Figure S1). The *R. distichophylla* showed a similar repeat content (~285.10 Mb; ~49.08%) compared with *R. guianensis* in the same genus (~339.26 Mb; ~54.15%) (Table 1) (Guo et al. 2019). LTR retrotransposons were the most abundant type of TEs, occupying roughly 35.99% of the *R. distichophylla*
specifically, the *R. guianensis* genome showed significantly expansion of Ty3/Gypsy retrotransposons compared with the *R. distichophylla* genome (Table S7). In total, 220,737 and 496,819 SSRs were found in the *R. distichophylla* and moso bamboo genomes, respectively, with trimer and tetramer as the most abundant SSR types (Table S8). Among the trimer motifs, (CCG/GGC)n were the predominant repeat in *R. distichophylla*, whereas (CCG/GGC)n, (AGG/CCT)n and (AAG/CCT)n showed a similar proportions in the moso bamboo genome. The identified SSRs will provide valuable molecular resources for germplasm characterization and genomics-based breeding programs. In total, we identified 727 tRNA genes, 90 rRNA genes, 242 snoRNA genes, 127 snRNA genes, and 256 miRNA genes, respectively (Table 1; Table S9).

In combination with *ab initio* prediction, protein and transcript alignments, we obtained a gene set consisted of 30,763 protein-coding genes (Table 1; Table S10), with an average length of 2,887 bp and an average coding sequence length of 1,099 bp (Table S10; Figure S2). Among these genes, 88.85% had significant similarities to sequences in the public databases (Table S11).

DISCUSSION

In this study, we present a draft genome assembly of the herbaceous bamboo *R. distichophylla* to supplement the currently existing bamboo genomic resources. The assembled genome was ~580.85 Mb in size, with a contig N50 length of ~86.36 Kb, ~7.14 times longer than the previously reported genome assembly of *R. guianensis* (Guo et al. 2019). The genome assembly comprised 38,269 scaffolds with a N50 length of ~1.80 Mb, which is far more contiguous than that of *R. guianensis* (Guo et al. 2019). Validation of genome assembly using reads mapping, transcripts alignments and BUSCO assessment together showed that the assembled *R. distichophylla* draft genome is accurate and complete. Polyploidy has been proved to be an important evolutionary force for the speciation as well as trait specialization in flowering plants (Blanc and Wolfe 2004; Jiao et al. 2011; Jiang et al. 2014; Wolfe 2001), which is...
commonly present in bamboos. Thus, the availability of a much more contiguous
diploid genome will greatly facilitate the reconstruction of the evolutionary history of
polyploidy in different bamboo clades.

SSRs were commonly utilized to develop molecular markers (Kumari et al. 2013;
Pandey et al. 2013), which have been extensively applied to exploring the molecular
phylogeny and taxonomy of bamboo species (e.g., Zhao et al. 2015). In this study, we
identified the whole genome-based SSR loci. Our results suggested that about ~59.2%
of SSRs were the tri- and tetra-nucleotide repeats. Of these, tri-nucleotide repeats
were the predominant class of repeat type. It is well known that the taxonomy of
bamboos has long puzzled the researcher community because of bamboos’
reproductive characteristics. The SSRs identified in this study would particularly
benefit population genetics and phylogenetics studies on bamboos towards efficient
conservation of the bamboo germplasms.

ACKNOWLEDGEMENTS

This work was supported by Yunnan Innovation Team Project and the start-up grant
from South China Agricultural University (to L. G.), Guangdong Special Support
Program (to Q. Z.), and the Presidential Foundation of Guangdong Academy of
Agricultural Sciences (201611) (to Q. Z.).
LITERATURE CITED

BambooPhylogenyGroup 2012 An updated tribal and subtribal classification for the Bambusoideae (Poaceae). In Proceedings of the 9th World Bamboo Congress, pp. 10-15.

Bao, Z., and S. R. Eddy, 2002 Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res 12: 1269-1276.

Blanc, G., and K. H. Wolfe, 2004 Widespread Paleopolyploidy in Model Plant Species Inferred from Age Distributions of Duplicate Genes. The Plant Cell 16: 1667-1678.

Boeckmann, B., A. Bairoch, R. Apweiler, M. C. Blatter, A. Estreicher et al., 2003 The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res 31: 365-370.

Boetzer, M., C. V. Henkel, H. J. Jansen, D. Butler, and W. Pirovano, 2011 Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27: 578-579.

Bolger, A. M., M. Lohse, and B. Usadel, 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120.

Clark, L., X. Londoño, and E. Ruiz-Sanchez, 2015 Bamboo taxonomy and habitat. In Bamboo (Springer), pp. 1-30.

Clark, L., and R. Oliveira, 2018 Diversity and evolution of the New World bamboos (Poaceae: Bambusoideae: Bambuseae, Olyreae). In Proceedings of the 11th World Bamboo Congress, Xalapa, Mexico, pp. 35-47.

Das, M., S. Bhattacharya, and A. Pal, 2005 Generation and characterization of SCARs by cloning and sequencing of RAPD products: a strategy for species-specific marker development in bamboo. Annals of botany 95: 835-841.

Deschamps, S., Y. Zhang, V. Llaca, L. Ye, A. Sanyal et al., 2018 A chromosome-scale assembly of the sorghum genome using nanopore sequencing and optical mapping. Nature Communications 9: 4844.

Gaut, B. S., L. G. Clark, J. F. Wendel, and S. V. Muse, 1997 Comparisons of the molecular evolutionary process at rbcL and ndhF in the grass family (Poaceae).
Molecular Biology and Evolution 14: 769-777.
Giulietti, A. M., R. M. Harley, L. P. De Queiroz, M. D. G. L. Wanderley, and C. Van Den Berg, 2005 Biodiversity and conservation of plants in Brazil. Conservation Biology 19: 632-639.
Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson et al., 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29: 644-652.
Gui, Y. J., Y. Zhou, Y. Wang, S. Wang, S. Y. Wang et al., 2010 Insights into the bamboo genome: syntenic relationships to rice and sorghum. Journal of integrative plant biology 52: 1008-1015.
Guo, Z. H., P. F. Ma, G. Q. Yang, J. Y. Hu, Y. L. Liu et al., 2019 Genome sequences provide insights into the reticulate origin and unique traits of woody bamboos. Molecular plant 12: 1353-1365.
Haas, B. J., A. L. Delcher, S. M. Mount, J. R. Wortman, R. K. Smith et al., 2003 Improving the Arabidopsi genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res 31: 5654-5666.
Haas, B. J., S. L. Salzberg, W. Zhu, M. Pertea, J. E. Allen et al., 2008 Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol 9: R7.
IRGSP, 2005 The map-based sequence of the rice genome. Nature 436: 793-800.
Jiao, Y., P. Peluso, J. Shi, T. Liang, M. C. Stitzer et al., 2017 Improved maize reference genome with single-molecule technologies. Nature 546: 524-527.
Jiang, W. K., Y. L. Liu, E. H. Xia, and L. Z. Gao, 2013 Prevalent role of gene features in determining evolutionary fates of whole-genome duplication duplicated genes in flowering plants. Plant Physiology 161: 1844-1861.
Jiao, Y., N. J. Wickett, S. Ayyampalayam, A. S. Chanderbali, L. Landherr et al., 2011 Ancestral polyploidy in seed plants and angiosperms. Nature 473: 97-100.
Jones, P., D. Binns, H. Y. Chang, M. Fraser, W. Li et al., 2014 InterProScan 5: genome-scale protein function classification. Bioinformatics 30: 1236-1240.
Judziewicz, E. J., L. G. Clark, X. Londoño, and M. J. Stern, 1999 American bamboos
Kelchner, S. A., and L. G. Clark, 1997 Molecular Evolution and Phylogenetic Utility of the Chloroplast p16Intron in Chusquea and the Bambusoideae (Poaceae). Molecular Phylogenetics and Evolution 8: 385-397.
Kelchner, S. A., and Bamboo Phylogeny Group, 2013 Higher level phylogenetic relationships within the bamboos (Poaceae: Bambusoideae) based on five plastid markers. Molecular Phylogenetics and Evolution 67: 404-413.
Kent, W. J., 2002 BLAT--the BLAST-like alignment tool. Genome Res 12: 656-664.
Lagesen, K., P. Hallin, E. A. Rodland, H. H. Staerfeldt, T. Rognes et al., 2007 RNAmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35: 3100-3108.
Kumari, K., M. Muthamilarasan, G. Misra, S. Gupta, A. Subramanian et al., 2013 Development of eSSR-markers in Setaria italica and their applicability in studying genetic diversity, cross-transferability and comparative mapping in millet and non-millet species. PloS one 8: e67742.
Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754-1760.
Li, R., W. Fan, G. Tian, H. Zhu, L. He et al., 2010 The sequence and de novo assembly of the giant panda genome. Nature 463: 311-317.
Li, W., and A. Godzik, 2006 Cdh-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22: 1658-1659.
Liu, B., Y. Shi, J. Yuan, X. Hu, and F. Wei, 2013 Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. Quantitative Biology 35: 62-67.
Lowe, T. M., and S. R. Eddy, 1997 tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25: 955-964.
Lowe, T. M., and S. R. Eddy, 1999 A computational screen for methylation guide snoRNAs in yeast. Science 283: 1168-1171.
Mascher, M., H. Gundlach, A. Himmelbach, S. Beier, S. O. Twardziok et al., 2017 A chromosome conformation capture ordered sequence of the barley genome.
Nature 544: 427-433.

Nawrocki, E. P., D. L. Kolbe, and S. R. Eddy, 2009 Infernal 1.0: inference of RNA alignments. Bioinformatics 25: 1335.

Oliveira, R. P., L. G. Clark, A. S. Schnadelbach, S. H. Monteiro, E. L. Borba et al., 2014 A molecular phylogeny of Raddia and its allies within the tribe Olyreae (Poaceae, Bambusoideae) based on noncoding plastid and nuclear spacers. Molecular Phylogenetics and Evolution 78: 105-117.

Pandey, G., G. Misra, K. Kumari, S. Gupta, S. K. Parida et al., 2013 Genome-wide development and use of microsatellite markers for large-scale genotyping applications in foxtail millet [Setaria italica (L.)]. DNA research 20: 197-207.

Peng, Z., Y. Lu, L. Li, Q. Zhao, Q. Feng et al., 2013 The draft genome of the fast-growing non-timber forest species moso bamboo (Phyllostachys heterocycla). Nat. Genet. 45: 456-461.

Porebski, S., L. G. Bailey, and B. R. Baum, 1997 Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Molecular Biology Reporter 15: 8-15.

Price, A. L., N. C. Jones, and P. A. Pevzner, 2005 De novo identification of repeat families in large genomes. Bioinformatics 21 Suppl 1: i351-358.

Kajitani, R., K. Toshimoto, H. Noguchi, A. Toyoda, Y. Ogura et al., 2014 Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome research 24: 1384-1395.

Saarela, J. M., S. V. Burke, W. P. Wysocki, M. D. Barrett, L. G. Clark et al., 2018 A 250 plastome phylogeny of the grass family (Poaceae): topological support under different data partitions. PeerJ 6: e4299.

Salamov, A. A., and V. V. Solovyev, 2000 Ab initio gene finding in Drosophila genomic DNA. Genome Res 10: 516-522.

Sharma, R., P. Gupta, V. Sharma, A. Sood, T. Mohapatra et al., 2008 Evaluation of rice and sugarcane SSR markers for phylogenetic and genetic diversity analyses in bamboo. Genome 51: 91-103.

Silva-Junior, O. B., D. Grattapaglia, E. Novaes, and R. G. Collevatti, 2018 Genome
assembly of the pink ipê (*Handroanthus impetiginosus*, Bignoniaceae), a highly valued, ecologically keystone neotropical timber forest tree. Gigascience 7: gix125.

Simao, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov, 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31: 3210-3212.

Soderstrom, T.R., 1981 Some Evolutionary Trends in the Bambusoideae (Poaceae). Annals of the Missouri Botanical Garden 68: 15-47.

Soreng, R. J., P. M. Peterson, K. Romaschenko, G. Davidse, J. K. Teisher *et al.*, 2017 A worldwide phylogenetic classification of the Poaceae (Gramineae) II: An update and a comparison of two 2015 classifications. Journal of Systematics and Evolution 55: 259-290.

Stanke, M., R. Steinkamp, S. Waack, and B. Morgenstern, 2004 AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. 32: W309-312.

Sugita, T., Y. Semi, H. Sawada, Y. Utoyama, Y. Hosomi *et al.*, 2013 Development of simple sequence repeat markers and construction of a high-density linkage map of *Capsicum annuum*. Molecular Breeding 31: 909-920.

Sungkaew, S., C. M. Stapleton, N. Salamin, and T. R. Hodkinson, 2009 Non-monophyly of the woody bamboos (Bambuseae; Poaceae): a multi-gene region phylogenetic analysis of Bambusoideae ss. Journal of plant research 122: 95.

Tarailo-Graovac, M., and N. Chen, 2009 Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics Chapter 4: Unit 4.10.

The International Brachypodium Initiative, 2010 Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463: 763-768.

Thiel, T., W. Michalek, R. K. Varshney, and A. Graner, 2003 Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl. Genet. 106: 411-422.

VanBuren, R., D. Bryant, P. P. Edger, H. Tang, D. Burgess *et al.*, 2015
Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*. Nature 527: 508-511.

Vorontsova, M. S., L. G. Clark, J. Dransfield, R. Govaerts, and W. J. Baker, 2016 World Checklist of Bamboos and Rattans: In Celebration of INBAR’s 20th Anniversary. Beijing.

Wolfe, K. H., 2001 Yesterday's polyploids and the mystery of diploidization. Nature Reviews Genetics, 2: 333-341.

Wu, T. D., and C. K. Watanabe, 2005 GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21: 1859-1875.

Wysocki, W. P., L. G. Clark, L. Attigala, E. Ruiz-Sanchez, and M. R. Duvall, 2015 Evolution of the bamboos (Bambusoideae; Poaceae): a full plastome phylogenomic analysis. BMC evolutionary biology 15: 50.

Wysocki, W. P., E. Ruiz-Sanchez, Y. Yin, and M. R. Duvall, 2016 The floral transcriptomes of four bamboo species (Bambusoideae; Poaceae): support for common ancestry among woody bamboos. BMC genomics 17: 384.

Zhang, G., X. Liu, Z. Quan, S. Cheng, X. Xu et al., 2012 Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. Nature Biotechnology 30: 549-554.

Zhang, Y. J., P. F. Ma, and D. Z. Li, 2011 High-throughput sequencing of six bamboo chloroplast genomes: phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). PloS one 6: e20596.

Zhao, H., L. Yang, Z. Peng, H. Sun, X. Yue et al., 2015 Developing genome-wide microsatellite markers of bamboo and their applications on molecular marker assisted taxonomy for accessions in the genus *Phyllostachys*. Scientific reports 5: 1-10.
Tables

Table 1. Summary of the genome assemblies and annotations of *R. distichophylla* and *R. guianensis*.

| Assembly                        | *R. distichophylla* | *R. guianensis* |
|--------------------------------|---------------------|-----------------|
| Estimated genome size (Mb)      | 608                 | 685             |
| Assembled sequence length (Mb)  | 581                 | 626             |
| Scaffold Number                 | 38,269              | 12,824          |
| Scaffold N50 (Mb)               | 1.81                | 0.012           |
| Contig Number                   | 45,206              | 13,300          |
| Contig N50 (Kb)                 | 86.36               | 12.09           |

**Annotation**

|                              | *R. distichophylla* | *R. guianensis* |
|------------------------------|---------------------|-----------------|
| Number of predicted protein-coding genes | 30,763              | 24,275          |
| Average gene length (bp)     | 2,886.93            | 2,635.83        |
| tRNAs                        | 727                 | 923             |
| rRNAs                        | 90                  | 743             |
| snoRNAs                      | 242                 | NA              |
| snRNAs                       | 127                 | 358             |
| miRNAs                       | 256                 | 387             |
| Transposable elements (Mb)   | 285.10              | 339.26          |
| Transposable elements (%)    | 49.08               | 54.15           |
Figure Legends

Figure 1. The sequenced *R. distichophylla* plant.

Figure 2. The 17-mer distribution of sequencing reads from *R. distichophylla*. The occurrence of 17-mer was calculated using GCE based on the sequencing data from short insert size libraries (insert size ≤ 500 bp) of *R. distichophylla*. The sharp peak on the left with low depths represents the essentially random sequencing errors. The middle and right peaks indicate the heterozygous and homozygous peaks, the depths of which are 52 and 103, respectively.
