Applications in Plant Sciences

Applications in Plant Sciences 2016 4(7): 1600017

Genomic Resources Note

Low-coverage, whole-genome sequencing of *Artocarpus camansi* (Moraceae) for phylogenetic marker development and gene discovery[1]

Elliot M. Gardner[2,3,5], Matthew G. Johnson[2], Diane Ragone[4], Norman J. Wickett[2,3], and Nyree J. C. Zerega[2,3]

[2]Department of Plant Science, Chicago Botanic Garden, 1000 Lake Cook Road, Glencoe, Illinois 60022 USA; [3]Plant Biology and Conservation, Northwestern University, 2205 Tech Drive, Hogan 2-144, Evanston, Illinois 60208 USA; and [4]Breadfruit Institute, National Tropical Botanical Garden, Kalaheo, Hawaii 96741 USA

- *Premise of the study:* We used moderately low-coverage (17×) whole-genome sequencing of *Artocarpus camansi* (Moraceae) to develop genomic resources for *Artocarpus* and Moraceae.
- *Methods and Results:* A de novo assembly of Illumina short reads (251,378,536 pairs, 2 × 100 bp) accounted for 93% of the predicted genome size. Predicted coding regions were used in a three-way orthology search with published genomes of *Morus notabilis* and *Cannabis sativa*. Phylogenetic markers for Moraceae were developed from 333 inferred single-copy exons. Ninety-eight putative MADS-box genes were identified. Analysis of all predicted coding regions resulted in preliminary annotation of 49,089 genes. An analysis of synonymous substitutions for pairs of orthologs (Ks analysis) in *M. notabilis* and *A. camansi* strongly suggested a lineage-specific whole-genome duplication in *Artocarpus*.
- *Conclusions:* This study substantially increases the genomic resources available for *Artocarpus* and Moraceae and demonstrates the value of low-coverage de novo assemblies for nonmodel organisms with moderately large genomes.

Key words: *Artocarpus camansi*; breadfruit; breadnut; genome; MADS box; Moraceae.

Artocarpus J. R. Forst. & G. Forst. (Moraceae) contains approximately 70 species of monoecious trees with a center of diversity in Malesia (Zerega et al., 2010). The genus includes several underutilized crops that can improve food security, most notably breadfruit (*A. altilis* (Parkinson) Fosberg), a long-lived perennial crop that is low input but high yielding (Jones et al., 2011). Other *Artocarpus* crops include the pantropically cultivated jackfruit (*A. heterophyllus* Lam.), crops of regional importance like cempedak (*A. integer* (Thunb.) Merr.) and terap (*A. odoratissimus* Blanco), and more than a dozen other species with edible fruits whose potential remains largely unexplored (Zerega et al., 2010). *Artocarpus camansi* Blanco (breadnut), native to New Guinea, is the diploid wild progenitor of breadfruit and is cultivated throughout the tropics for its large edible seeds (Zerega et al., 2005) (Fig. 1).

Existing genomic resources for breadfruit include nuclear (Witherup et al., 2013) and chloroplast (Gardner et al., 2015) microsatellites as well as transcriptomes of breadfruit and two wild relatives (Laricchia, 2014). In this study, we augment these resources with a low-coverage shotgun assembly of the *A. camansi* genome.

Recent studies focused on different taxonomic groups have used ultra-shallow sequencing (or “genome skimming”)—with coverage of less than 1×—to assemble portions of genomes for annotation and marker discovery, particularly high-copy sequences such as organellar genomes and repetitive elements (Straub et al., 2011; Blischak et al., 2014). Here, we explore the utility of somewhat deeper but still shallow (17×) genome sequencing for de novo genome assembly and annotation with the goal of phylogenomic marker development, gene discovery, and the detection of whole-genome duplications.

METHODS AND RESULTS

The individual for sequencing was selected for the absence of heterozygosity at 19 nuclear microsatellite loci (Zerega et al., 2015), possibly due to centuries of inbreeding. The individual is the offspring of a tree planted in the Lancetilla Botanical Garden, Honduras. The Honduran tree likely descended from Caribbean material, which in turn was likely descended from seedlings in Mauritius that were collected in the Philippines by the French naturalist Pierre Sonnerat (1748–1814) (Ragone, 1997). Leaf tissue from *A. camansi* (living collection at McBryde Garden at the National Tropical Botanical Garden, Kalaheo, Hawaii: NTBG 960576.001; voucher: EG149 [CHIC]) was collected and dried on silica. Genomic DNA was extracted using the QIAGEN DNeasy Plant Mini Kit following the manufacturer’s protocol (QIAGEN, Valencia, California, USA). Two Illumina TruSeq libraries were prepared: a paired-end library with a mean insert size of 180 bp and a mate-pair library with a mean insert size of 854 bp (Illumina, San Diego, California, USA). The paired-end library was sequenced in a single lane on an Illumina HiSeq 2000 (2 × 100 bp, paired–end), and the mate-pair library was sequenced in one-half lane. Library preparation and

[1]Manuscript received 12 February 2016; revision accepted 1 June 2016.

The authors thank S. Greenwald (Argonne National Laboratory) for library preparation and Illumina sequencing and A. Devault (MYcroarray) for assistance with bait design. This research was funded by National Science Foundation grants to N.J.C.Z. (DEB-0919119) and N.J.W. (DEB-1239992) and a grant to N.J.C.Z. from the Institute for Sustainability and Energy at Northwestern University.

[5]Author for correspondence: egardner@u.northwestern.edu

doi:10.3732/apps.1600017

 © 2016 Gardner et al. Published by the Botanical Society of America. This work is licensed under a Creative Commons Attribution License (CC-BY-NC-SA).