Low-Coverage, Whole-Genome Sequencing of Artocarpus camansi (Moraceae) for Phylogenetic Marker Development and Gene Discovery

Authors: Gardner, Elliot M., Johnson, Matthew G., Ragone, Diane, Wickett, Norman J., and Zerega, Nyree J. C.

Source: Applications in Plant Sciences, 4(7)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1600017
**Genomic Resources Note**

**LOW-COVERAGE, WHOLE-GENOME SEQUENCING OF ARTOCARPUS CAMANSI (MORACEAE) FOR PHYLOGENETIC MARKER DEVELOPMENT AND GENE DISCOVERY**

**Elliott M. Gardner**, **Matthew G. Johnson**, **Diane Ragone**, **Norman J. Wickett**, **Nyree J. C. Zerega**

2Department of Plant Science, Chicago Botanic Garden, 1000 Lake Cook Road, Glencoe, Illinois 60022 USA; 3Plant Biology and Conservation, Northwestern University, 2205 Tech Drive, Hogan 2-144, Evanston, Illinois 60208 USA; and 4Breadfruit 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 Lancelin 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 (Illimina, 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 sequencing were completed at the North Central Regional Genomics Facility (University of Illinois, Urbana-Champaign, USA) using the Illumina technology platform.

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.

1Manuscript 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.

1Author for correspondence: egardner@u.northwestern.edu
doi:10.3732/apps.1600017

Applications in Plant Sciences 2016 4(7): 1600017; http://www.bioone.org/loi/apps © 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).
In addition to the markers targeted for their phylogenetic utility, target sequences for 125 additional genes were developed as follows. These sequences were selected based on predicted function only and were not screened for optimal hybridization characteristics such as long exons, high alignment specificity, or low copy number. Protein-coding regions of all scaffolds were predicted with Augustus (Stanke et al., 2008) using default parameters and allowing partial gene models. These predictions differed from the ESTScan predictions in their tendency to find even small, widely spaced, and partial exons, which is important for resequencing of functional genes, but not necessarily ideal for optimizing hybridization. Putative MADS-box genes were identified using a BLASTP search against all embryophyte genes in the NCBI nonredundant protein database, and those 98 sequences whose top BLASTP hit was a MADS-box gene were retained. To fill out the bait set, an additional 27 genes homologous to genes involved in biosynthesis of floral volatile compounds were extracted using the same search method, except that the query sequences were searched against the original contigs rather than against contigs smaller than 1000 bp. To ensure that baits would not hybridize to multiple regions of the genome, the remaining Artocarpus sequences were searched against the original A. camansi contigs using BLASTN, and any query sequence with more than one hit (>90% identity and >150 bp) was discarded. Finally, 333 predicted genes with exons between 500 and 2200 bp were selected. We retained only long exons to maximize the likelihood of hybridization success in divergent taxa, reasoning that longer exons would have a greater likelihood of containing at least a short span of conserved sequence that would hybridize across greater phylogenetic distances.

Table 1. Sequencing statistics in Artocarpus camansi.

| Sequence statistics | Paired-end library | Mate-pair library |
|---------------------|--------------------|-------------------|
| Mean insert size (bp) | 180                | 854               |
| Read pairs          | 182,485,953        | 68,892,583        |
| Read pairs surviving QC | 161,937,040 | 53,824,945        |
| GenBank SRA accession | SRR2910988    | SRR2911031        |

Table 2. Assembly statistics in Artocarpus camansi.

| Assembly statistics | Ray assembly | Scaffolds with transcriptome |
|---------------------|--------------|-----------------------------|
| Scaffolds           | 401,889      | 388,956                     |
| Scaffold size (bp)  | 632,414,964  | 633,895,366                 |
| Scaffold N50        | 2426         | 2574                        |
| Contigs             | 415,187      | 414,958                     |
| Contig size (bp)    | 630,980,246  | 630,983,028                 |
| Contig N50          | 2315         | 2317                        |
RNA scaffolding, gene annotation, and Ks analysis—To scaffold the genome assembly further before a full genome annotation was performed, we used a transcriptome assembly of *A. camansi* (seed offspring of NTBG 000501.005, voucher: EG140 [CHIC], BioProject PRJNA311339) (Laricchia, 2014) in L_RNA_scaffolder (Xue et al., 2013). This final assembly was used for all further analyses. Gene prediction on the extended scaffolds was performed using AUGUSTUS (Stanke et al., 2008), guided by alignments from the transcriptome assembly, and limited to complete gene models only. Predicted amino acid sequences were annotated with Swiss-Prot hits and Gene Ontology (GO) terms using Trinotate (https://trinotate.github.io/). We determined orthology between the AUGUSTUS protein predictions and the published proteomes of *M. notabilis* and *C. sativa* with Proteinortho (Lechner et al., 2011), using default settings. *Artocarpus* proteins were reduced to a nonredundant set using CD-HIT-EST (Fu et al., 2012) to remove sequences that were 98% similar at the nucleotide level along 90% of their length. Chromosome counts support a hypothesis of two whole-genome duplications within Artocarpeae: the one at the root of the tribe, evidenced by a haploid chromosome number of 14 in *Clarisia biflora* Ruiz & Pav. (Oginuma and Tobe, 1995), and another in the genus *Artocarpus*, which has a haploid number of 28 (Banerji and Hakim, 1954; Oginuma and Tobe, 1995; Ragone, 2001). By contrast, *Morus* has a base haploid number of 7, although a polyploid series exists within that genus as well (He et al., 2013) (Fig. 2). To determine the age of possible whole-genome duplication (WGD) in *Artocarpus*, we selected Proteinortho orthogroups that were single copy in *Morus* and *Cannabis*, but multicopy in *Artocarpus*. We generated pairwise alignments of *Artocarpus* protein sequences (paralogs) in each of these orthogroups using MAFFT, and back-translated the sequences to nucleotides using PAL2NAL (Suyama et al., 2006). We calculated the synonymous substitution rate (Ks) for each paralog pair with KaKs_Calculator (Zhang et al., 2006) using the GY
method (also known as F3 × 4) (Goldman and Yang, 1994). We also used this procedure to calculate Ks for orthogroups that were single copy in all three proteomes to generate three ortholog Ks distributions: Morus-Artocarpus, Morus-Cannabis, and Cannabis-Artocarpus.

### Sequencing and assembly results

Sequencing and assembly statistics appear in Tables 1 and 2. Based on the k-mer counting method, the Ray assembler predicted a genome size of 669 Mb, with a peak coverage of 17×. The Ray assembly was 631 Mb (93% of the predicted size by the k-mer method implemented in Ray), and the scaffold N50 was 2426 bp. After additional scaffolding with L_RNA_scaffold, the scaffold N50 increased to 2574, and the number of scaffolds dropped from 401,899 to 388,956 (Table 2).

### Identification of target sequences

Proteinortho found 2041 putative single-copy orthogroups with exactly one copy in each of the three species included that were nonorganellar and were present on contigs longer than 1000 bp. Of the 617 orthogroups with only a single BLASTN hit in the original A. camansi assembly hit (>90% identity and >150 bp), the 333 selected for bait design were between 504 and 2166 bp, and these totaled 354,171 bp. Comparison of these exons with full gene predictions by AUGUSTUS in IGV revealed that ESTs can tend to find single-exon genes or the longest exon in multixen genes, which was expected to increase the efficiency of hybridization. GenBank accession numbers, Artocarpus scaffold coordinates, and orthology assignments to M. notabilis and C. sativa are presented in Appendix S1. The maximum-likelihood tree (Fig. 3) indicates that of the 98 putative MADS-box sequences, 51 are tentatively assignable to domain MIKC (Type 2) and 47 are tentatively assignable to domain M (Type 1). However, these assignments should be treated with caution due to the fragmentary nature of the assembly. Exons for the putative MADS-box genes and the additional 27 terpenoid synthesis genes totaled 101,871 bp. The median ortholog hit ratio, calculated against the top BLASTP hit, was 0.60, indicating that most sequences likely contained the majority of a gene; 16 had an ortholog hit ratio of 1, indicating complete coding sequences. GenBank accession numbers, Artocarpus scaffold coordinates, and the best BLASTP hit are presented in Appendix S2. Fine-scale name assignments based on BLASTP hits should be treated with caution, particularly for the MADS-box genes, as they do not always agree with the phylogenetic placement in the maximum-likelihood tree. The 458 Artocarpus sequences identified here, combined with Morus orthologs for the 333 phylogenetic markers, maximized the available target space for a bait set made up of 20,000 120-mer baits with 3× tiling (MYcroarray, Ann Arbor, Michigan, USA). The use of these bait sequences to capture sequences from taxa representing all Moraceae tribes, including differences in hybridization efficiency at various phylogenetic distances, and between the 333 phylogenetic markers and the other genes, is reported in Johnson et al. (2016).

### Gene annotation and Ks analysis

AUGUSTUS predicted 83,061 gene sequences on the extended scaffolds; of these, Trinotate reported annotations for 49,089 sequences, including 22,865 sequences annotated with 1226 unique GO terms. A search for reciprocal best BLASTP hits between the predicted proteins for A. camansi and M. notabilis found 14,089 orthogroups collectively containing 15,634 Artocarpus genes and 14,506 Morus genes, which is 54% of the Morus total. Among the AUGUSTUS protein predictions, the M. notabilis proteome, and the C. sativa proteome, we generated 18,657 orthogroups using Proteinortho, 10,925 of which contained proteins from all three proteomes, and 8,539 were single copy in all three genomes. There were 1391 orthogroups containing multiple sequences from Artocarpus but only one sequence from each of the published genomes (many-to-one orthogroups). The distribution of synonymous substitutions (Ks) between pairs of Artocarpus paralogs in the many-to-one orthogroups reflects evidence of a possible lineage-specific WGD in A. camansi (Fig. 4). We then compared the Ks distribution of Artocarpus paralogs in many-to-one orthogroups to the distribution of Ks between pairs of orthologs in the single-copy orthogroups generated by Proteinortho. The Artocarpus paralog distribution indicates WGD in Artocarpus is more recent than the divergence of Artocarpus from M. notabilis, a member of the sister tribe Moreae (Fig. 4). This is consistent with a hypothesis of WGD events in the tribe Artoceae based on published chromosome counts (Fig. 2, discussed above). However, as confirmed chromosome counts exist for only a few of the ca. 70 Artocarpus species, more work is required to arrive at a firm conclusion regarding the precise phylogenetic position of the duplication within the genus.

### CONCLUSIONS

This study constitutes a substantial increase in the genomic resources available for Artocarpus in particular and Moraceae in general. The 333 phylogenetic markers will enable large-scale phylogenomic studies within the family using a Hyb-Seq (target enrichment) approach. Because sequence capture is anchored by conserved exons but can extend into flanking noncoding regions if read lengths are sufficient (Johnson et al., 2016), these markers can be employed at both deep and shallow phylogenetic scales. The gene predictions provide a resource that can be mined to explore gene families; for example, the MADS-box sequences identified provide a starting point for characterizing this important gene family in Artocarpus. Finally, the shotgun assembly presented here, while fragmentary, covers an estimated 93% of the genome, and represents a first step toward a reference genome assembly for Artocarpus. Thus the study demonstrates how existing reference genomes (Morus, Cannabis, and Arabidopsis).
can be leveraged to increase the value of de novo shotgun assemblies from shallow genome sequencing, enabling the inexpensive development of genome-level resources for nonmodel organisms with moderately large genomes.

LITERATURE CITED

ALTICHL, S. F., W. GISH, W. MILLER, E. W. MYERS, and D. J. LIPMAN. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

BARNEKI, L., and A. HAKIM. 1954. A contribution to the life-history of *Artocarpus lakoocha* Roxb. *Proceedings of the Indian Academy of Sciences, Section B* 92: 128–132.

BLEISCH, P. D., A. J. WENZEL, and A. D. WOLFER. 2014. Gene prediction and annotation in *Penstemon* (Plantaginaceae): A workflow for marker development from extremely low-coverage genome sequencing. *Applications in Plant Sciences* 2: 1400044.

BOISVERT, S., F. LAVIOLETTE, and J. CORBEL. 2010. Ray: Simultaneous assemblies of reads from a mix of high-throughput sequencing technologies. *Journal of Computational Biology* 17: 1519–1533.

BOGERS, A. M. M. LOUBE, and B. USSANZEL. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* 30: 2114–2120.

FU, L., B. NIU, Z. ZHU, S. WU, and W. LI. 2012. CD-HT: Accelerated for clustering the next generation sequencing data. *Bioinformatics (Oxford, England)* 28: 3150–3152.

GARDNER, E. M., K. M. LARICHEA, M. MURPHY, D. RAGONE, B. E. SCHEFFLER, S. SIMPSON, E. W. WILLIAMS, and N. J. C. ZERIEGA. 2015. Chloroplast microsatellite markers for *Artocarpus* (Moraceae) developed from transcriptome sequences. *Applications in Plant Sciences* 3(9): 1500049.

GARDNER, E. M., M. G. JOHNSON, D. RAGONE, N. J. WICKETT, and N. J. C. ZERIEGA. 2016. Data from: Low-coverage, whole-genome sequencing of *Artocarpus camansi* (Moraceae) for phylogenetic marker development and gene discovery. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.3293r

GOLDMAN, N., and Z. YANG. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution* 11: 725–736.

HE, N., C. ZHANG, X. QU, S. ZHAO, and Y. TAO. 2013. Draft genome sequence of the mulberry tree *Morus notabilis*. *Nature Communications* 4: 2445.

JONES, A. M. P., D. RAGONE, N. G. TAVANA, D. W. BERNDTAS, and S. J. MURCH. 2011. Beyond the bounty: Breadfruit (*Artocarpus alitiss*) for food security and novel foods in the 21st century. *Ethisbotany Research and Applications* 9: 129–149.

KANEHIRA, M., and S. GOTO. 2000. KEGG: Kyoto Encyclopaedia of Genes and Genomes. *Nucleic Acids Research* 28: 27–30.

KATOH, K., and D. M. STANLEY. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.

LAMIESCH, P., T. Z. BERARDINI, D. LI, D. SWARBRECK, C. WILKS, R. SASIDHARAN, R. MULLER, ET AL. 2012. The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Research* 40: D1202–D1210.

LARICHEA, K. 2014. Transcriptome analysis of domesticated breadfruit and its wild relatives. Master’s thesis, Northwestern University, Evanston, Illinois, USA.

LECHNER, M., S. FINDER, L. STEINER, M. MARZ, P. F. STADLER, and S. J. PROHASKA. 2011. Proteinortho: Detection of (co-)orthologs in large-scale analysis. *BMC Genomics* 12: 124.

O’NEIL, S. T., and S. J. EMRICH. 2013. Assessing de novo transcriptome assembly metrics for consistency and utility. *BMC Genomics* 14: 465.

OGINUMA, K., and H. TOBE. 1995. Karyomorphology of some Moraceae and Cecropiaceae (Urticales). *Journal of Plant Research* 108: 313–326.

RAGONE, D. 1997. Breadfruit, *Artocarpus alitlis* (Parkinson) Fosberg. Promoting the conservation and use of underutilized and neglected crops. IPGRI, Rome, Italy.

RAGONE, D. 2001. Chromosome numbers and pollen stainability of three species of Pacific Island breadfruit (*Artocarpus* Moraceae). *American Journal of Botany* 88: 693–696.

STAMATAKIS, A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics (Oxford, England)* 22: 2688–2690.

STANKE, M., M. DICKHANS, R. BAERTHISI, and D. HAUSSLER. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics (Oxford, England)* 24: 637–644.

STRAUB, S. C. K., M. FISHBEN, T. LIVSHULTZ, Z. FOSTER, M. PARKS, K. WEITEMIER, R. C. CRONN, and A. LISTON. 2011. Building a model: Developing genomic resources for common milkweed (*Asclepias syriaca*) with low coverage genome sequencing. *BMC Genomics* 12: 211.

SUYAMA, M., D. TORRENTS, and P. BORK. 2006. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* 34: W609–W612.

TAN, Y., Q. DONG, Z. B. CHI, P. CONG, and Z. ZHOU. 2015. Genomewide identification and analysis of the MADS-box gene family in apple. *Genome* 555: 277–290.

VAN BAKEL, H., J. M. STOUT, A. G. COTE, C. M. TALLON, A. G. SHARPE, T. R. HUGHES, and J. E. PAGE. 2011. The draft genome and transcriptome of *Cannabis sativa*. *Genome Biology* 12: R102.

WEITEMIER, K., S. C. K. STRAUB, R. C. CRONN, M. FISHBEN, R. SCHMUCKL, A. MCDONNELL, and A. LISTON. 2014. Hyb-Seq: Combining target enrichment and genome scaffinging for plant phylogenetics. *Applications in Plant Sciences* 2(9): 1400042.

WITHERUP, C., D. RAGONE, T. WIESNER-HANKS, B. IRISI, B. SCHEFFLER, S. SIMPSON, F. ZEE, ET AL. 2013. Development of microsatellite loci in *Artocarpus alitlis* (Moraceae) and cross-amplification in congeneric species. *Applications in Plant Sciences* 1(7): 1200423.

XUE, W., J.-T. LI, Y.-P. ZHU, G.-Y. HOU, X.-F. KONG, Y.-Y. KUANG, and X.-W. SUN. 2013. L_RNA scaffold: Scaffolding genomes with transcripts. *BMC Genomics* 14: 604.

YU, H., J. D. NASON, L. ZHANG, L. ZHENG, W. WU, and X. GE. 2015. De novo transcriptionome sequencing in *Ficus birti* Vahl. (Moraceae) to investigate gene regulation involved in the biosynthesis of pollinator attracting volatiles. *Tree Genetics & Genomes* 11: 91.

ZERIEGA, N. J. C., D. RAGONE, T. MOTLEY, and W. ZOMLEFER. 2005. Systematics and species limits of breadfruit (*Artocarpus* Moraceae). *Systematic Botany* 30: 603–615.

ZERIEGA, N. J. C., M. N. SUPARDI, and T. J. MOTLEY. 2010. Phylogeny and recircumscription of *Artocarpae* (Moraceae) with a focus on *Artocarpus*. *Systematic Botany* 35: 766–782.

ZERIEGA, N., T. WIESNER-HANKS, D. RAGONE, B. IRISI, B. SCHEFFLER, S. SIMPSON, and F. ZEE. 2015. Diversity in the breadfruit complex (*Artocarpus*, Moraceae): Genetic characterization of critical germplasm. *Tree Genetics & Genomes* 11: 4.

ZHANG, Z., J. LI, X. Q. ZHAO, J. WANG, G. K. S. WONG, and J. YU. 2006. KaKs_Calculator: Calculating Ka and Ks through model selection and model averaging. *Genomics, Proteomics & Bioinformatics* 4: 259–263.

http://www.bioone.org/loi/apps