De Novo Reference Assembly of the Upriver Orange Mangrove (*Bruguiera sexangula*) Genome

Wirulda Pootakham[1,*], Chaiwat Naktang[1,†], Chutima Sonthirod[1,†], Wasitthee Kongkachana[1], Thippawan Yoocha[1], Nukoon Jomchai[1], Chatree Maknual[2], Pranom Chumriang[2], Tamanai Pravinvongvuthi[2], and Sithichoke Tangphatsornruang[1,*]

[1]National Omics Center, National Science and Technology Development Agency (NSTDA), Pathum Thani, Thailand

[2]Department of Marine and Coastal Resources, Bangkok, Thailand

[†]These authors contributed equally to this work.

*Corresponding authors: E-mails: wirulda@alumni.stanford.edu; sithichoke.tan@nstda.or.th.

Accepted: February 1, 2022

Abstract

Upriver orange mangrove (*Bruguiera sexangula*) is a member of the most mangrove-rich taxon (Rhizophoraceae family) and is commonly distributed in the intertidal zones in tropical and subtropical latitudes. In this study, we employed the 10× Genomics linked-read technology to obtain a preliminary de novo assembly of the *B. sexangula* genome, which was further scaffolded to a pseudomolecule level using the *Bruguiera parviflora* genome as a reference. The final assembly of the *B. sexangula* genome contained 260 Mb with an N50 scaffold length of 11,020,310 bases. The assembly comprised 18 pseudomolecules (corresponding to the haploid chromosome number in *B. sexangula*), covering 204,645,832 bases or 78.6% of the 260-Mb assembly. We predicted a total of 23,978 protein-coding sequences, 17,598 of which were associated with gene ontology terms. Our gene prediction recovered 96.6% of the highly conserved orthologs based on the Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis. The chromosome-level assembly presented in this work provides a valuable genetic resource to help strengthen our understanding of mangroves' physiological and morphological adaptations to the intertidal zones.

Key words: *Bruguiera sexangula*, mangrove, 10× Genomics, genome assembly.

Significance

The upriver orange mangrove (*Bruguiera sexangula*) is adapted to hostile intertidal environments characterized by hypersalinity, hypoxia, high temperature, tidal fluctuation and strong UV light by developing special features such as aerial root systems, salt tolerance and viviparous mode of reproduction. Here, we generated a chromosome-level reference genome assembly of the upriver orange mangrove and presented evidence of recent whole genome duplication. This reference genome provides a valuable resource for future studies that will shed light on mangroves' physiological and morphological adaptations to the intertidal zones.

Introduction

Mangrove forest ecosystems are of great ecological and economic importance, supporting wetland communities of plants and animals including commercially important marine species such as fish, crab and prawn (Nagelkerken et al. 2008; Carugati et al. 2018). Mangrove species have successfully colonized intertidal habitats at the interface between terrestrial and marine ecosystems. They have adapted to hostile intertidal environments characterized by hypersalinity, hypoxia, high temperature, tidal fluctuation and strong UV light (Rothschild and Mancinelli 2001; Sandilyan 2010; Giri et al. 2011) by developing unique features such as salt tolerance,

aerial root systems and vivipary (Ball 1988; Parida and Jha 2010; Tomlinson 2016). Upriver orange mangrove (*Bruguiera sexangula*) is a member of Rhizophoraceae family and is widely distributed in the intertidal zones along sheltered coastal areas in tropical and subtropical latitudes. Mangrove ecosystems across the world are vulnerable to human exploitation. The world's mangrove forests have diminished in area by 50%, and the remaining forests have suffered partial degradation (Sandilyan and Kathiresan 2012).

In an attempt to understand the molecular basis of mangroves' physiological adaptations in order to better conserve and restore mangrove ecosystems, multiple mangrove genomes have been sequenced including *Rhizophora apiculata* (Xu et al. 2017), *Kandelia obovata* (Hu et al. 2020), *Bruguiera parviflora* (Pootakham, Sonthirod, Naktang, Kongkachana, Sangsrakru, et al. 2022) and *Ceriops zippeliana* (Pootakham, Sonthirod, Naktang, Kongkachana, U-Thoomporn, et al. 2022) from the Rhizophoraceae family, *Avicenna marina* (Friis et al. 2021) from the Acanthaceae family, and *Aegiceras corniculatum* (Feng et al. 2021) from the Primulaceae family. Transcriptomic resources from various mangrove species have also been reported in the past few years (Yamanaka et al. 2009; Yang Y, Yang S, Li, Deng, et al. 2015, Yang Y, Yang S, Li J, Li X, et al. 2015; Krishnamurthy et al. 2017; Xu et al. 2017). Nonetheless, a high-quality, chromosome-scale assembly of the *B. sexangula* genome has not been reported. In this work, we employed the 10× Genomics linked-read technology (Paajanen et al. 2019) to obtain a draft assembly of the *B. sexangula* genome. We subsequently utilized RagTag software to scaffold this preliminary assembly using a chromosome-level genome assembly of a closely related mangrove species, *B. parviflora* (Pootakham, Sonthirod, Naktang, Kongkachana, Sangsrakru, et al. 2022), as a reference. We also obtained transcript data (RNA-seq) to assist with coding gene prediction and annotation. The final assembly was 260 Mb and comprised 20,644 scaffolds (51 scaffolds larger than 100 kb). The N50 scaffold length was 11,020,310 bases and the largest scaffold was 17,482,559 bases. We predicted a total of 26,038 gene models, of which 23,978 (92.09%) were functionally annotated and 17,598 (67.59%) had gene ontology (GO) assignments. We believe this high-quality genome assembly along with its annotation data will be of interest to researchers studying mangrove species and anyone working in the plant comparative genomics and evolution. The genomic and transcriptomic resources presented here will also be valuable for mangrove forest conservation and restoration programs.

Results and Discussion

Genome Assembly

To generate a draft genome sequence for *B. sexangula*, high-quality genomic DNA from one mature individual located in the protected mangrove forest in Ranong (supplementary fig. 1, Supplementary Material online) was used for 10× Genomics linked-read library preparation. A total of 798,923,120 paired-end reads totaling 119.83 Gb was obtained from the library. The preliminary assembly generated from the linked-read data was 263,981,354 bases with an N50 contig length of 1,414,200 bases (table 1). This assembly was further scaffolded with RagTag using a chromosome-scale assembly of a closely related mangrove species, *B. parviflora* (Pootakham, Sonthirod, Naktang, Kongkachana, Sangsrakru, et al. 2022), as a reference. The final assembly contained 260,518,658 bases in 20,644 scaffolds with an N50 length of 11,020,310 bases (fig. 1 and table 1). The 18 largest pseudomolecules, corresponding to the *B. sexangula* haploid chromosome number ($2n = 36$), covered 204,645,832 bases or 78.6% of the 260-Mb assembly. From here on, we will refer to these pseudomolecules as chromosomes, numbered according to size. The assembly size (260 Mb) corresponded extremely well with our genome size estimation using DNA flow cytometry (259 Mb) and the analysis of *k*-mer distribution of sequencing reads (263 Mb; supplementary fig. 2, Supplementary Material online).

We evaluated the quality of the final assembly by aligning short-read DNA sequences and Trinity-assembled transcript sequences to our assembly. We observed that 96.65% and 93.96% of the Illumina short reads and transcript sequences, respectively, could be aligned to the genome sequence, suggesting that our genome assembly is of high quality. We also employed Benchmarking Universal Single-Copy Orthologs (BUSCO) to evaluate the quality of the assembly using the plant-specific database of 1,614 genes from the Embryophyta OrthoDB release 10. *B. sexangula* gene predictions recovered 96.6% of the highly conserved orthologs (92.4% classified as complete and single copy, 4.2% as complete and duplicated, and 1.2% as fragmented) whereas 2.2% of the Embryophyta conserved orthologs were missing from the gene prediction.

Gene Annotation

We employed a combination of ab initio prediction, homology-based search and transcript evidence to annotate the genome. We obtained a total of 26,038 predicted gene models and 23,978 protein-coding genes (92.09% of the predicted genes; supplementary tables 1 and 2, Supplementary Material online). GO was assigned to 17,598 (67.59%) protein-coding genes. Of the predicted gene models, 71.98%, 29.68%, and 17.52% could be functionally annotated using the Swissprot, EC and KEGG databases, respectively (supplementary table 2, Supplementary Material online). Additionally, Infernal 1.1 (Nawrocki and Eddy 2013) was used to perform homology search and annotate noncoding RNA sequences (supplementary table 3, Supplementary Material online). A total of 3,378 microRNAs, 246 ribosomal RNAs, 478 transfer RNAs and 11,905 small nuclear RNAs were identified in the genome.

Table 1
Bruguiera sexangula Genome Assembly Statistics

	10× Genomics	10× Genomics + RagTag Scaffolding
N50 contig/scaffold size (bases)	1,414,200	11,020,310
L50 contig/scaffold number	45	10
N75 contig/scaffold size (bases)	252,961	7,984,485
L75 contig/scaffold number	139	17
N90 contig/scaffold size (bases)	6,222	7,028
L90 contig/scaffold number	1,623	1,019
Assembly size (bases)	263,981,354	260,518,658
Number of scaffolds	21,624	20,644
Number of scaffolds ≥ 100 kb	199	51
Number of scaffolds ≥ 1 Mb	65	19
Number of scaffolds ≥ 10 Mb	0	11
Longest scaffold (bases)	9,126,008	17,482,559
% *N*	0.57	0.62
GC content (%)	34.13	34.12
BUSCO evaluation (% completeness)	96.3	96.6

The average GC content of the *B. sexangula* genome assembly was 35.6% (table 1), which was close to the average GC content in introns (34.9%), whereas the average GC content in exons was higher at 45.7% (supplementary table 1, Supplementary Material online).

Repetitive Elements in the *B. sexangula* Genome

We identified and annotated 99.51 Mb (38.2%) of the *B. sexangula* genome as repetitive sequences (supplementary table 4, Supplementary Material online). These repeat elements comprised 3.30 Mb (1.26%) of DNA transposons, 4.44 Mb (1.71%) of simple sequence repeats, 50.27 Mb (19.3%) of unclassified elements and 41.50 Mb (15.93%) of retrotransposons (supplementary table 4, Supplementary Material online). The majority of the classified repetitive sequences were retrotransposons, representing 41.7% of the total repeat elements in the genome. Most of the retrotransposons were classified as long terminal repeats (39.76 Mb), which could further be categorized into *Copia*-like (15.60 Mb) and *Gypsy*-like (19.07 Mb) elements (supplementary table 4, Supplementary Material online). The percentage of the *B. sexangula* assembly containing repetitive elements (38.2%) was higher than that observed in the *K. obovata* (24%) (Hu et al. 2020), *B. parviflora* (26%) (Pootakham, Sonthirod, Naktang, Kongkachana, Sangsrakru, et al. 2022) and *R. apiculata* (29%) (Xu et al. 2017) genomes but comparable to the figure reported for the *Populus trichocarpa* genome (40%) (Zhou and Xu 2009).

Phylogenetic Analysis of *B. sexangula* and Related Mangrove Genomes

We analyzed the gene sets from five mangrove species (*B. parviflora*, *Bruguiera gymnorhiza*, *Ceriops tagal*, *K. obovata* and *R. apiculata*), five dicots (*A. thaliana*, *C. melo*, *C. sativus*, *P. trichocarpa*, and *R. communis*), and one monocot (*Oryza sativa*) in order to explore the relationships among mangrove and other plant species. A total of 334,622 proteins (out of 351,322 input proteins from 12 species; 95.25%) were clustered into 23,601 orthologous groups. A maximum-likelihood tree was constructed based on sequence information from 3,108 single-copy orthologous genes, and the divergence time was estimated based on the topology and the branch length. *Bruguiera gymnorhiza* and *B. sexangula* diverged approximately 8.06 million years ago (Ma; supplementary fig. 3*A*, Supplementary Material online), and their last common ancestor diverged from *B. parviflora* 23.68 Ma. The last common ancestor of *Bruguiera* and the last common ancestor of three other Rhizophoraceae members (*C. tagal*, *K. obovata* and *R. apiculata*) diverged roughly 49.8 Ma (supplementary fig. 3*A*, Supplementary Material online).

We used the 4DTv approach to estimate the timing of evolutionary divergence between *B. sexangula* and related mangrove species (supplementary fig. 3*B*, Supplementary Material online). Comparison of 7,319 pairs of paralogous genes residing within 291 duplicated colinear blocks (defined as regions of genome that harbor at least ten colinear genes with fewer than six intervening genes) within *B. sexangula* genome revealed a prominent peak at 0.115 4DTv, suggesting that *B. sexangula* has experienced a recent genome-wide duplication event (supplementary fig. 3*B*, Supplementary Material online). This was also supported by the extensive presence of intragenomic synteny blocks throughout the genome (fig. 1). Additionally, we observed distance peaks at approximately 0.12 4DTv in *K. obovata* and *R. apiculata*, suggesting that these two closely related mangrove species had undergone a genome-wide duplication during the same period as *B. sexangula*. Interestingly, the *B. sexangula*–*K. obovata* and *B. sexangula*–*R. apiculata* peaks displayed a lower 4DTv distance (0.042) than the 4DTv distances between paralogous genes within *B. sexangula*, *K. obovata* and *R. apiculata* (∼0.12), raising the possibility that the observed whole-genome duplication event indeed occurred

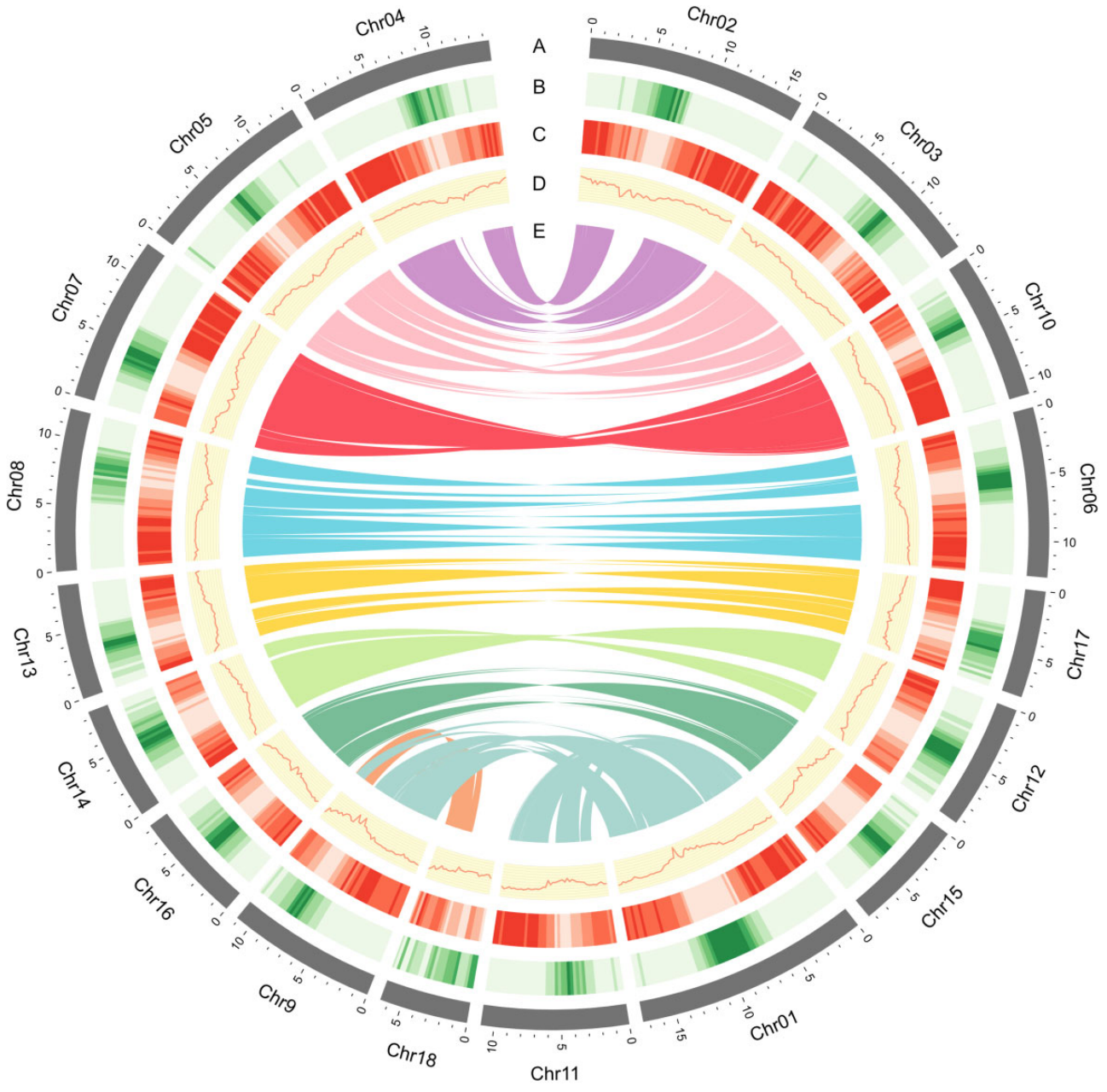

Fig. 1.—Genomic landscape of *B sexangula*. (*A*) A physical map of 18 chromosomes numbered according to size (Mb scale). (*B*) Repeat density represented by the fraction of genomic regions covered by repetitive sequences in 250-kb windows. (*C*) Gene density represented by the number of genes in 250-kb windows. (*D*) GC content represented by the percentage of G + C bases in 250 kb windows. (*E*) Syntenic regions in the genome are illustrated by connected lines.

in the common ancestor of the Rhizophoreae before the speciation of *B. sexangula*, *K. obovata* and *R. apiculata*.

Materials and Methods

Plant Material Collection and Extraction of Nucleic Acid

Leaf tissues from one mature *B. sexangula* individual located in the natural mangrove forest under the protection of the Department of Marine and Coastal Resources (Ranong Province; 10°10′15.45″N 98°42′26.89″E) were collected, immediately frozen, and stored in liquid nitrogen until use. High-molecular weight genomic DNA was extracted and evaluated following the protocol in Pootakham, Sonthirod, et al. (2021). For transcriptome sequencing, we isolated total RNA from leaf tissues collected from the same individual used for genome sequencing following the protocol in Pootakham, Nawae, et al. (2021). Poly(A) mRNA was enriched using the Dynabeads mRNA Purification Kit (ThermoFisher Scientific,

Waltham, USA). Prior to the library construction, the integrity of the RNA samples was evaluated on the Fragment Analyzer System (Agilent, Santa Clara, USA).

Library Preparation and Sequencing

For whole-genome sequencing, 1 ng of DNA was used for the 10× Genomics linked-read library preparation using the Chromium Genome Library Kit & Gel Bead Kit v2 according to the manufacturer's instructions (10× Genomics, Pleasanton, USA). The resulting 10× library was subsequently sequenced on the Illumina HiSeq X Ten (150-bp paired-end reads). To obtain short-read RNA sequences, a sequencing library was prepared according to the methods reported in Pootakham, Naktang, et al. (2021). The library was sequenced on the MGISEQ-2000RS using the MGISEQ-2000RS Sequencing Flow Cell V3.0 (MGI Tech, Shenzhen, China).

Preliminary Genome Assembly and Scaffolding

Linked-read data were assembled using the Supernova assembler v2.1.1 using the default settings (https://support.10xgenomics.com/de-novo-assembly/software/pipelines/latest/using/running; last accessed: February 14, 2022; 10× Genomics, Pleasanton, USA). The preliminary assembly was further scaffolded with RagTag v1.1.0 (https://github.com/malonge/RagTag; last accessed: February 14, 2022) (Alonge et al. 2019) using the *B. parviflora* genome assembly (Pootakham, Sonthirod, Naktang, Kongkachana, Sangsrakru, et al. 2022), as a reference.

Estimation of Genome Size

We employed both DNA flow cytometry and the *k*-mer analysis of the sequencing read distribution to estimate the nuclear genome size. The *k*-mer analysis was performed using the Jellyfish software v2.2.10, and the *k*-mer distribution was plotted with GenomeScope v1.0 ($k = 31$; http://qb.cshl.edu/genomescope/; last accessed: February 14, 2022) (Vurture et al. 2017). For DNA flow cytometry analysis, fresh leaf tissues were chopped using a sharp razor blade to release nuclei, and the procedure in Dolezel and Bartos (2005) was followed. For *B. sexangula*, the woody plant buffer reported in Loureiro et al. (2007) was used as a nuclear isolation buffer, and nuclei were stained with 50 μg/ml propidium iodide (Thermo Fisher Scientific, Waltham, USA). Arabidopsis was used as the DNA reference standard, and LB01 was used as its nuclear isolation buffer (Dolezel and Bartos 2005).

Assessment of the Genome Assembly Quality

We evaluated the quality of the final assembly by aligning short-read DNA sequences (Illumina data) and Trinity-assembled transcript sequences using BWA v0.7.17-r1188 and GMAP v2020-09-12 (Wu and Watanabe 2005), respectively, to our assembly. We also employed the BUSCO v4.0.5 (Simão et al. 2015) to evaluate the assembly by testing for the presence and completeness of the orthologs using the plant-specific database of 1,614 genes from the Embryophyta OrthoDB release 10 (Kriventseva et al. 2015).

Repetitive Sequence Identification

We employed RepeatModeler v2.0.2 (http://www.repeatmasker.org/RepeatModeler/; last accessed: February 14, 2022) to identify transposable element (TE) families in the unannotated genome assembly. RECON v1.08 and RepeatScout v1.0.5 were used to identify the boundaries of repetitive elements and to build consensus models of interspersed repeats. To confirm that repeat sequences in the library did not contain large families of protein-coding sequences that were not TEs, we aligned them to GenBank's nr protein database using BLASTX with an e-value cutoff of 10^{-6}.

Gene Prediction and Annotation

We used evidences from ab initio prediction, RNA-based prediction and homology-based prediction to identify protein-coding sequences in the unmasked genome assembly using EvidenceModeler (EVM) v1.1.1 (Haas et al. 2008). RNA-based prediction approaches utilized RNA-seq evidences from *B. sexangula*. Short-read RNA sequences were mapped to the assembly during the initial step of annotation using the PASA2 pipeline v2.4.1 (Haas et al. 2008). Protein sequences from *O. sativa*, *Mimulus guttatus*, *Sesamum indicum*, *P. trichocarpa*, and *Eucalyptus grandis* obtained from public databases were aligned to the unmasked genome using AAT (Huang et al. 1997). Two ab initio gene predictors were run on the unmasked assembly. Protein-coding gene predictions were obtained with Augustus v3.3.3 (Stanke et al. 2004) trained with *O. sativa*, *M. guttatus*, *S. indicum*, *P. trichocarpa*, and *E. grandis* and PASA (v2.4.1) transcriptome alignment assembly (Hoff et al. 2016, 2019) using RNA-seq alignment files as inputs. All gene predictions were combined by EVM to generate consensus gene models using the following weights for each evidence type: PASA2–1, AAT–0.3, and Augustus–0.3. We cross-checked the position of annotated genes with those of known repeats and excluded any gene that had more than 50% overlapping sequences with repetitive elements from the list. The remaining predicted genes were functionally annotated using OmicsBox v2.0.10 (https://www.biobam.com/download-omicsbox/; last accessed: February 14, 2022). Protein sequences were aligned with UniProtKB/Swiss-Prot and GenBank nonredundant database using local BLASTP with an *e*-value cutoff of 10^{-5}. GO terms were retrieved and assigned to *B. sexangula* query sequences, and enzyme codes (EC) corresponding to GO were retrieved and mapped to KEGG pathway annotations.

Phylogenetic Analysis

OrthoFinder was used to identify orthologous groups in *B. sexangula*, five mangrove species (*B. gymnorhiza*, *B. parviflora*, *C. tagal*, *K. obovata* and *R. apiculata*) five dicots (*Arabidopsis thaliana*, *Cucumis melo*, *Cucumis sativus*, *Ricinus communis* and *P. trichocarpa*) and one monocot (*O. sativa*). A phylogenetic tree was constructed based on protein sequences from single-copy orthologous groups using the RAxML-NG program (Stamatakis 2006) following the protocol reported in Yang Y, Yang S, Li J, Li X, et al. (2015). Divergence times were estimated using the MCMCtree software v4.0 (PAML 4 package; Yang 2007) using the Bayesian Relaxed Molecular Clock approach. The "correlated molecular clock" and "JC69" models were used with the published divergence time between the common ancestor of Rhizophoraceae, Euphorbiaceae (*R. communis*) and Salicaceae (*P. trichocarpa*) estimated at 105–120 Ma (Davis et al. 2005; Xi et al. 2012). The most recent fossil recognized as ancestral Rhizophora has been dated to the late Eocene (33.9–38 Ma) (Muller 1981; Graham 2006).

The Analysis of Genome Synteny

We employed MCscanX (Wang et al. 2012) to analyze colinearity within the *B. sexangula* genome. To identify putative paralogs, *B. sexangula* amino acid sequences were aligned against themselves using BLASTP with an *e*-value cutoff of 10^{-10}. We defined intragenic homeologous segments as regions of at least ten genes with colinear or nearly colinear runs of paralogs elsewhere in the genome with fewer than six intervening genes. The intragenic homeologous segments were plotted using CIRCOS v0.69.8 (Krzywinski et al. 2009).

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgment

This study was supported by the National Science and Technology Development Agency (NSTDA), Thailand (Grant No. P-19-52261).

Author Contributions

W.P. and S.T. designed research study and obtained the funding. W.P., T.Y., N.J., C.M., P.C., and T.P. performed laboratory work (sample collection, DNA and RNA extraction, library construction, and sequencing). C.N., C.S., and W.K. performed bioinformatics analyses. W.P. wrote and revised the manuscript, and all authors reviewed it.

Data Availability

Illumina sequencing data from the 10× Genomics library and RNA-seq data (MGISEQ) were submitted to the NCBI Sequence Read Archive (SRA) database under BioProject accession number PRJNA734123 (DNA short-read data: SRX12279148; RNA-seq data: SRX12119193). The *B. sexangula* genome assembly was deposited in the DDBJ/ENA/GenBank database under the accession number JAHLGP000000000. The annotation, coding sequences, protein sequences, and functional annotation (based on the NCBI nr database) have been deposited in figshare database under the DOI: 10.6084/m9.figshare.16625905.v1.

Literature Cited

Alonge M, et al. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 20(1):224.
Ball MC. 1988. Ecophysiology of mangroves. Trees 2(3):129–142.
Carugati L, et al. 2018. Impact of mangrove forests degradation on biodiversity and ecosystem functioning. Sci Rep. 8(1):13298.
Davis CC, Webb CO, Wurdack KJ, Jaramillo CA, Donoghue MJ. 2005. Explosive radiation of Malpighiales supports a mid-Cretaceous origin of modern tropical rain forests. Am Nat. 165(3):E36–E65.
Dolezel J, Bartos J. 2005. Plant DNA flow cytometry and estimation of nuclear genome size. Ann Bot. 95(1):99–110.
Feng X, et al. 2021. Genomic insights into molecular adaptation to intertidal environments in the mangrove *Aegiceras corniculatum*. New Phytol. 231(6):2346–2358.
Friis G, et al. 2021. A high-quality genome assembly and annotation of the gray mangrove, *Avicennia marina*. Genes Genom Genet. 11:jkaa025.
Giri C, et al. 2011. Status and distribution of mangrove forests of the world using earth observation satellite data. Global Ecol Biogeogr. 20(1):154–159.
Graham A. 2006. Paleobotanical evidence and molecular data in reconstructing the historical phytogeography of Rhizophoraceae. Ann Mo Bot Gard. 93(2):325–334.
Haas BJ, et al. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. 9(1):R7.
Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32(5):767–769.
Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-genome annotation with BRAKER. Methods Mol Biol. 1962:65–95.
Hu M-J, et al. 2020. Chromosome-scale assembly of the *Kandelia obovata* genome. Hortic Res. 7(1):75.
Huang X, Adams MD, Zhou H, Kerlavage AR. 1997. A tool for analyzing and annotating genomic sequences. Genomics 46(1):37–45.
Krishnamurthy P, et al. 2017. Transcriptomics analysis of salt stress tolerance in the roots of the mangrove *Avicennia officinalis*. Sci Rep. 7(1):10031.
Kriventseva EV, et al. 2015. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. Nucleic Acids Res. 43(Database issue):D250–D256.
Krzywinski MI, et al. 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19(9):1639–1645.
Loureiro J, Rodriguez E, Dolezel J, Santos C. 2007. Two new nuclear isolation buffers for plant DNA flow cytometry: a test with 37 species. Ann Bot. 100(4):875–888.
Muller J. 1981. Fossil pollen records of extant angiosperms. Bot Rev. 47(1):1.

Nagelkerken I, et al. 2008. The habitat function of mangroves for terrestrial and marine fauna: a review. Aquat Bot. 89(2):155–185.
Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29(22):2933–2935.
Paajanen P, et al. 2019. A critical comparison of technologies for a plant genome sequencing project. Gigascience 8(3):giy163.
Parida AK, Jha B. 2010. Salt tolerance mechanisms in mangroves: a review. Trees 24(2):199–217.
Pootakham W, Naktang C, et al. 2021. De novo chromosome-level assembly of the *Centella asiatica* genome. Genomics 113(4):2221–2228.
Pootakham W, Nawae W, et al. 2021. A chromosome-scale assembly of the black gram (*Vigna mungo*) genome. Mol Ecol Resour. 21(1):238–250.
Pootakham W, Sonthirod C, et al. 2021. De novo assemblies of *Luffa acutangula* and *Luffa cylindrica* genomes reveal an expansion associated with substantial accumulation of transposable elements. Mol Ecol Resour. 21(1):212–225.
Pootakham W, Sonthirod C, Naktang W, Kongkachana W, Sangsrakru D, et al. 2022. A chromosome-scale reference genome assembly of yellow mangrove (*Bruguiera parviflora*) reveals a whole genome duplication event associated with the Rhizophoraceae lineage. Mol Ecol Resour. 1–15. doi: 10.1111/1755-0998.13587.
Pootakham W, Sonthirod C, Naktang W, Kongkachana W, U-Thoomporn S, et al. 2022. A de novo reference assembly of the yellow mangrove *Ceriops zippeliana* genome. Genes Genom Genet. doi: 10.1093/g3journal/jkac025.
Rothschild LJ, Mancinelli RL. 2001. Life in extreme environments. Nature. 409(6823):1092–1101.
Sandilyan S. 2010. Climate change and mangrove wetlands. Emerg Sci. 2:18–19.
Sandilyan S, Kathiresan K. 2012. Mangrove conservation: a global perspective. Biodivers Conserv. 21(14):3523–3542.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210–3212.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22(21):2688–2690.
Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. 32(Web Server issue):W309–W312.
Tomlinson PB. 2016. The botany of mangroves. Cambridge: Cambridge University Press.
Vurture GW, et al. 2017. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33(14): 2202–2204.
Wang Y, et al. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 40(7):e49.
Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21(9):1859–1875.
Xi Z, et al. 2012. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. Proc Natl Acad Sci USA. 109(43):17519–17524.
Xu S, et al. 2017. The origin, diversification and adaptation of a major mangrove clade (Rhizophoreae) revealed by whole-genome sequencing. Natl Sci Rev. 4(5):721–734.
Yamanaka T, Miyama M, Tada Y. 2009. Transcriptome profiling of the mangrove plant *Bruguiera gymnorhiza* and identification of salt tolerance genes by Agrobacterium functional screening. Biosci Biotechnol Biochem. 73(2):304–310.
Yang Y, Yang S, Li J, Deng Y, et al. 2015. Transcriptome analysis of the Holly mangrove *Acanthus ilicifolius* and its terrestrial relative, *Acanthus leucostachyus*, provides insights into adaptation to intertidal zones. BMC Genomics. 16:605–605.
Yang Y, Yang S, Li J, Li X, et al. 2015. De novo assembly of the transcriptomes of two yellow mangroves, *Ceriops tagal* and *C. zippeliana*, and one of their terrestrial relatives, *Pellacalyx yunnanensis*. Mar Genomics. 23:33–36.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591.
Zhou F, Xu Y. 2009. RepPop: a database for repetitive elements in *Populus trichocarpa*. BMC Genomics. 10:14.

Associate editor: Maud Tenaillon