A high-quality genome assembly and annotation of the gray mangrove, *Avicennia marina*

Guillermo Friis [ID],[1,*] Joel Vizueta [ID],[2] Edward G. Smith [ID],[1] David R. Nelson,[1] Basel Khraiwesh,[1,3] Enas Qudeimat,[1,3] Kourosh Salehi-Ashtiani,[1] Alejandra Ortega,[4] Alyssa Marshell,[5] Carlos M. Duarte,[4] and John A. Burt[1]

[1]Center for Genomics and Systems Biology, New York University - Abu Dhabi, PO Box 129188, Abu Dhabi, United Arab Emirates
[2]Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona 08007, Spain
[3]Division of Biological and Environmental Sciences and Engineering, Center for Desert Agriculture, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia
[4]Red Sea Research Center (RSRC) and Computational Bioscience Research Center, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia
[5]Department of Marine Science and Fisheries, College of Agricultural and Marine Sciences, Sultan Qaboos University, Muscat, Oman

*Corresponding author: Guillermo Friis Center for Genomics and Systems Biology, New York University—Abu Dhabi, PO Box 129188, Abu Dhabi, United Arab Emirates. E-mail: gfm3@nyu.edu

Abstract

The gray mangrove [*Avicennia marina* (Forsk.) Vierh.] is the most widely distributed mangrove species, ranging throughout the Indo-West Pacific. It presents remarkable levels of geographic variation both in phenotypic traits and habitat, often occupying extreme environments at the edges of its distribution. However, subspecific evolutionary relationships and adaptive mechanisms remain understudied, especially across populations of the West Indian Ocean. High-quality genomic resources accounting for such variability are also sparse. Here we report the first chromosome-level assembly of the genome of *A. marina*. We used a previously release draft assembly and proximity ligation libraries Chicago and Dovetail HiC for scaffolding, producing a 456,526,188-bp long genome. The largest 32 scaffolds (22.4–10.5 Mb) accounted for 98% of the genome assembly, with the remaining 2% distributed among much shorter 3,759 scaffolds (62.4–1 kb). We annotated 45,032 protein-coding genes using tissue-specific RNA-seq data in combination with *de novo* gene prediction, from which 34,442 were associated to GO terms. Genome assembly and annotated set of genes yield a 96.7% and 95.1% completeness score, respectively, when compared with the eudicots BUSCO dataset. Furthermore, an F_{ST} survey based on resequencing data successfully identified a set of candidate genes potentially involved in local adaptation and revealed patterns of adaptive variability correlating with a temperature gradient in Arabian mangrove populations. Our *A. marina* genomic assembly provides a highly valuable resource for genome evolution analysis, as well as for identifying functional genes involved in adaptive processes and speciation.

Keywords: gray mangrove; *Avicennia marina*; genome assembly; HiC; Arabia

Introduction

Mangroves are a polyphyletic group of trees and shrubs that inhabit the intertidal zone of the tropic and subtropic coasts (Nagelkerken *et al.* 2008; Polidoro *et al.* 2010; Hogarth 2015; Primavera *et al.* 2018). Mangroves share several morphological and physiological adaptations to their habitat, including traits for tolerance of high salinity, alternating dessication and submergence of soils across tidal cycles, and exposure of roots to hypoxic, sulfide-rich soils (Tomlinson 2016). Mangroves are of great ecological and economic importance, providing key functions such as high productivity, much of which is exported to surrounding ecosystems, and acting as breeding, nesting, nursery, foraging, and shelter areas for a range of biota (Nagelkerken *et al.* 2008; Lee *et al.* 2014; Carugati *et al.* 2018; Primavera *et al.* 2018). Mangroves also serve as an important carbon sink, supporting climate change mitigation and adaptation potential (Duarte *et al.* 2013). The diversity of evolutionary origins and adaptive

mechanisms found in mangroves provides compelling systems for studying patterns of trait evolution, lineage divergence, and speciation (*e.g.*, Zhou *et al.* 2007; Urashi *et al.* 2013; Xu *et al.* 2017b).

Of the approximately 70 mangrove species described, the gray mangrove *Avicennia marina* has the broadest latitudinal and longitudinal distribution (Spalding *et al.* 2010; Hogarth 2015; Tomlinson 2016). At least three subspecific, partially allopatric taxa or "varieties" have been described: *A. marina* var. *australasica*, restricted to southeastern Australia and New Zealand; *A. marina* var. *eucalyptifolia*, in northern regions of Australasia; and *A. marina* var. *marina* that ranges from New Caledonia in the Pacific and across the entire Indian Ocean (Duke 1991; Li *et al.* 2016; Tomlinson 2016). The broad geographic distribution of *A. marina* is reflected in its presence across diverse environmental gradients (*e.g.*, temperature, freshwater, sediment and nutrient supply, salinity, tidal range) and spatial settings (*e.g.*, open coastlines, coastal lagoons, estuaries, deltas, coral fringes) (Duke 1990;

Quisthoudt *et al.* 2012). Along with its widespread distribution, several aspects of *A. marina* biology make it a promising model organism among mangroves for the study of evolution based on genomic and molecular approaches. First, the broad environmental gradients present across the *A. marina* range are mirrored by remarkable geographic variation in phenotypic and life-history traits (Duke 1990; Tomlinson 2016), which makes them an suitable system for studying evolutionary processes related to dispersal, directional selection, and neutral evolution. Previous studies have reported the phylogenetic relationships for the varieties of *A. marina* and congeneric species (Duke *et al.* 1998; Nettel *et al.* 2008; Li *et al.* 2016), yet the specific drivers of lineage diversification remain understudied. In particular, the extensive gray mangrove populations from the eastern African and Arabian coasts have rarely been included in reported DNA sequence-based analyses, either for diversity screening or other purposes (*e.g.*, Duke *et al.* 1998; Li *et al.* 2016). Second, *A. marina* can tolerate highly variable and extreme environmental conditions, and often occupies marginal, biologically limiting environments at the edges of its distribution (Morrisey *et al.* 2010). While several biological structures and mechanisms of the gray mangrove physiology have been described (Nguyen *et al.* 2015; Naidoo 2016), the genetic basis and pathways underlying such tolerance are still largely unknown (Mehta *et al.* 2005; Jithesh *et al.* 2006; Xu *et al.* 2017a). Understudied Arabian populations represent a particular gap due to their occurrence at the extremes of air and water temperature, salinity, and aridity (Camp *et al.* 2018; Ben-Hasan and Christensen 2019). Third, the *A. marina* genome is moderately small and structurally simple compared with other eukaryotes and presents a limited amount of transposable elements (Das *et al.* 1995; Xu *et al.* 2017a; Lyu *et al.* 2018), which facilitates the identification of short polymorphisms and structural variants.

Previous studies on *A. marina* using high-throughput sequencing techniques and genomic approaches have been released. Two multispecies studies, including *A. marina*, explored patterns of convergent evolution at functional genes (Xu *et al.* 2017a) and transposable elements loads (Lyu *et al.* 2018), based on draft genomes that were recently made available online. Whole-genome assemblies of *A. marina* and several other mangrove taxa have recently been used for demographic inference (Guo *et al.* 2018) and convergent evolution analysis (He *et al.* 2020), but the underlying genomic data are not publicly accessible.

Here, we report a high-quality, chromosome-level assembly obtained using proximity ligation libraries as a resource for genome-based studies on *A. marina* and related mangrove species. A structural and functional annotation based on RNA-seq data from multiple tissues is also provided. In addition, we generated whole-genome shotgun data for a set of resequenced individuals from six different populations along the coasts of the Arabian Peninsula. We used these data to evaluate the performance of the assembly as a reference to study patterns of adaptive variability at the genomic level. Resequencing data are also used for organelle assembling.

Materials and methods
Genome sequencing and assembly
A high-quality genome was produced combining a previously released draft assembly of *A. marina* var. *marina* (GenBank assembly accession: GCA_900003535.1, GCA_900174615.1; Xu *et al.* 2017a; Lyu *et al.* 2018) and newly generated sequence data from proximity ligation libraries. Preparation of proximity ligation libraries Chicago and HiC as well as scaffolding with the software pipeline

HiRise (Putnam *et al.* 2016; https://dovetailgenomics.com) was conducted at Dovetail Genomics, LLC. The sequenced sample was leaf tissue obtained from an individual of *A. marina* var. *marina*, located at Ras Ghurab Island in the Arabian Gulf (Figure 1; Abu Dhabi, United Arab Emirates; 24.601°N, 4.566°E). Briefly, for Chicago and Dovetail HiC libraries preparation, chromatin was fixed with formaldehyde. Fixed chromatin was then digested with DpnII and free blunt ends were ligated. Crosslinks were reversed, and the DNA purified from protein, which was then sheared to ~350 bp mean fragment size. Libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters, and sequencing was carried out on an Illumina HiSeq X platform. Chicago and Dovetail HiC library reads were then used as input data for genome assembly for HiRise, a software pipeline designed specifically for using proximity ligation data for high-level scaffolding of draft genomes (Supplementary Figures S1 and S2; Putnam *et al.* 2016). The previously reported draft genome of *A. marina* was used in the assembly pipeline, excluding scaffolds shorter than 1 kb since HiRise does not assemble them. Further details are provided in the Supplementary information.

The mitochondrial genome was assembled using NOVOplasty2.7.2 (Dierckxsens *et al.* 2017) and resequencing data based on Illumina paired-end 150-bp libraries from a conspecific individual (see below; Supplementary information). The maturase (matR) mitochondrial gene available in NCBI (GenBank accession no. AY289666.1) was used for the input seed sequence. The assembly of the chloroplast yielded a high number of alternative, unsolved assemblies and is thereby no reported.

Genome annotation
We performed the annotation of the *A. marina* genome using mRNA data from a set of tissues of conspecific individuals in combination with *de novo* gene prediction using BRAKER2 v2.1.5 (Hoff *et al.* 2016). Samples were collected on the coast of the Eastern Central Red Sea north of Jeddah in the Kingdom of Saudi Arabia (22.324°N, 39.100°E; Figure 1A). Total RNA was isolated from root, stem, leaf, flower, and seed using TRIzol reagent (Invitrogen, USA). RNA-seq libraries were prepared using TruSeq RNA sample prep kit (Illumina, Inc.), with inserts that range in size from approximately 100–400 bp. Library quality control and quantification were performed with a Bioanalyzer Chip DNA 1000 series II (Agilent) and sequenced in an HiSeq2000 platform (Illumina, Inc.). First, repetitive regions were modeled *ab initio* using RepeatModeler v2.0.1 (Flynn *et al.* 2019) in all scaffolds longer than 100 kb with default options. The resulting repeat library was used to annotate and soft-mask repeats in the genome assembly with RepeatMasker 4.0.9 (Smit *et al.* 2015). Next, messenger RNA reads were mapped against the soft-masked genome assembly with HISAT2 (Kim *et al.* 2015). Gene prediction was conducted with BRAKER2 using both the RNA-seq data and the conserved orthologous genes from BUSCO Eudicots_odb10 as proteins from short evolutionary distance to provide hints and train GeneMark-ETP and Augustus (–etpmode; Lomsadze *et al.* 2005; Stanke *et al.* 2006; Gotoh 2008; Stanke *et al.* 2008; Li *et al.* 2009; Barnett *et al.* 2011; Iwata and Gotoh 2012; Lomsadze *et al.* 2014; Buchfink *et al.* 2015; Hoff *et al.* 2019; Bruna *et al.* 2020). The obtained gene annotation gff3 file was validated and used to generate the reported gene annotation statistics with GenomeTools (Gremme *et al.* 2013) and in-house Perl scripts. Finally, we conducted a similarity-based approach to assist the functional annotation of the predicted proteins. We integrated InterProScan v5.31 (Jones *et al.* 2014) and BLAST (Tatusova and Madden 1999) searches using the UniProt Swiss-Prot database and the annotated proteins
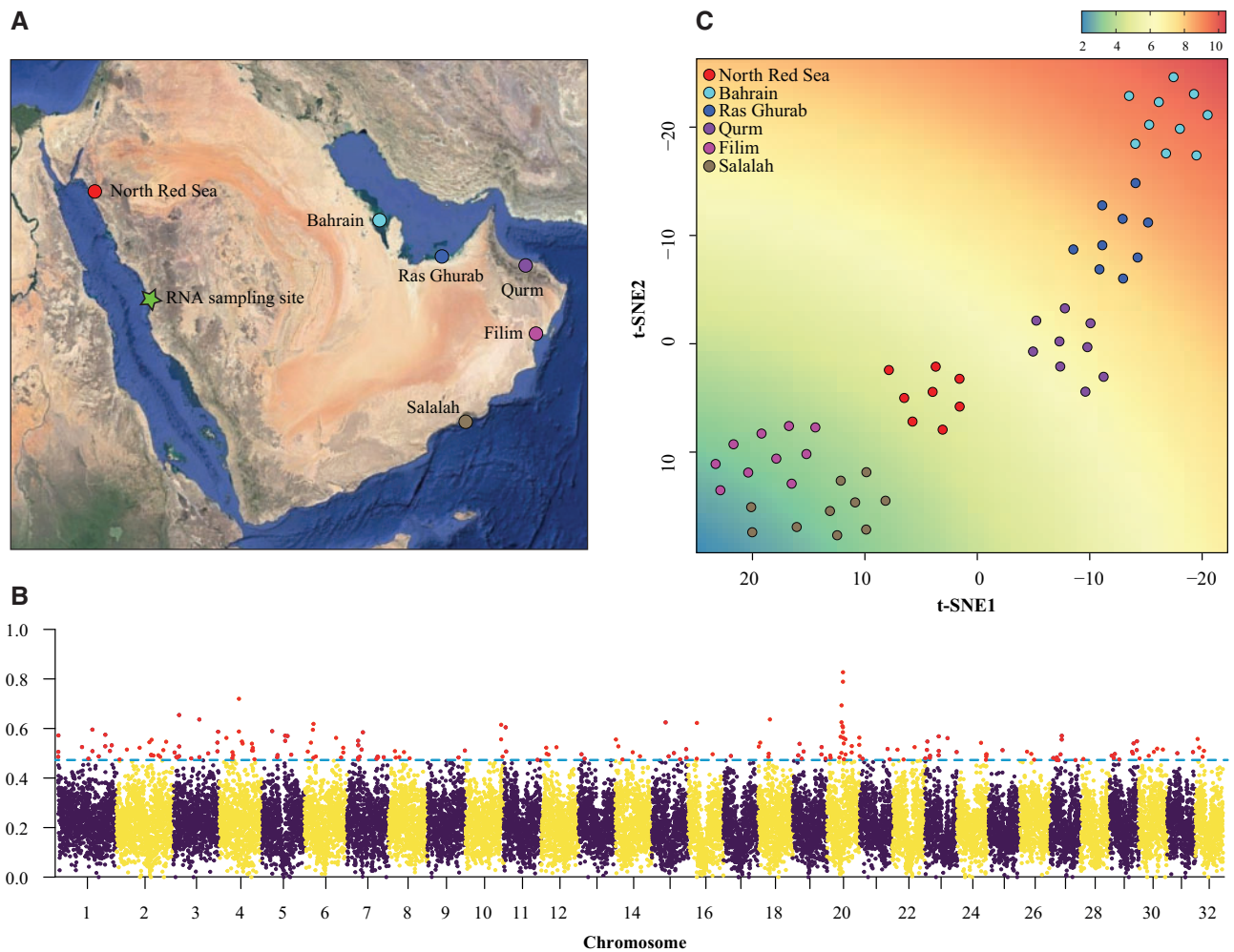
Figure 1 Geography and adaptive variability in Arabian gray mangroves. (A) Locations of the six stands sampled for whole-genome resequencing (colored circles) and for RNA-seq (green star). (B) F_{ST} genome scan based on 22,181 windows of 20 kb. Boxplot outliers (coefficient = 1.5) are marked in red (C) t-SNE based on 613 SNP outliers linked to functional genes. The background shows the correlation of t-SNE1 and t-SNE2 with the annual temperature range registered in each one of the sampling locations. Temperature depicted in the legend is in °C.

from the *Arabidopsis thaliana* genome (UniProt Consortium 2019) to generate a final set of annotated functional genes (Supplementary Figure S4). Further details on mRNA sequencing and annotation scripts are provided in the Supplementary information.

Gene completeness assessment

We assessed gene completeness in the genome assembly, and gene annotation, using BUSCO (Benchmarking Universal Single-Copy Orthologs) v4.0.5 (–auto-lineage-euk option; Waterhouse *et al.* 2018). BUSCO evaluations were conducted using the 255 and 2326 single-copy orthologous genes in Eukaryota_odb10 and Eudicots_odb10 datasets, respectively.

Adaptive variability analysis and functional assessment of A. marina *genome*

To test the potential of the assembly and annotation reported here as a resource for genomic-based studies, we checked for regions of high divergence across the genome of *A. marina* using newly generated whole-genome data. We resequenced 60 individuals from six different sampling sites from each of the major seas bordering Arabia (Figure 1A; Supplementary Table S1),

including populations in the Red Sea, the Arabian Gulf, and the Arabian Sea/Sea of Oman. Arabia's regional seas present extreme, but divergent, environmental conditions for mangrove growth. The northern Red Sea is characterized by colder winter temperatures and high salinity (Carvalho *et al.* 2019), while the southern Persian/Arabian Gulf is the world's hottest sea each summer and is also hypersaline, with both areas considered arid to hyperarid with limited (<200 mm) rainfall (Vaughan *et al.* 2019). In contrast, the Arabian Sea and Sea of Oman have normal oceanic salinity, and summer temperatures that are buffered by cold-water upwelling as a result of the Indian Ocean monsoon, resulting in more benign environmental conditions (Burt *et al.* 2016; Claereboudt 2019). Using our new *A. marina* genome assembly as a reference, we mapped the resequenced data with the mem algorithm implemented in the Burrows-Wheeler Aligner (BWA; Li and Durbin 2009) and conducted the variant calling with the Genome Analysis Toolkit (GATK; McKenna *et al.* 2010). Applying heavy quality filters, a matrix of genome-wide SNPs was generated. Using vcftools (Danecek *et al.* 2011), we computed Weir and Cockerham (W&C; Weir and Cockerham 1984) Weighted F_{ST} values based on 20-kb sliding windows, and defining each sampling site as a population to be compared. Based on

the distribution of F_{ST} values, boxplot outlier loci (coefficient = 1.5) associated with functional genes were identified. We then used these loci to explore geographic patterns of adaptive variability by means of t-SNE analysis, testing for correlations between variability in sea surface temperature (SST) and t-SNE scores. Details on sequencing, variant calling, and analytical procedures for this section are available in the Supplementary information.

Data availability

Genome assembly, genome annotation, and related supporting resources for facilitating the use of these data have been deposited at DRYAD (doi:10.5061/dryad.3j9kd51f5). SNP matrix from resequenced individuals used in adaptive variation analyses is also provided in this dataset in vcf format. Raw resequencing data have been deposited in GenBank under the accessions SRA: SRP265707; BioProject: PRJNA629068.

The genome assembly has also been deposited at DDBJ/ENA/GenBank along with raw sequence data from Chicago and HiC libraries under the accession JABGBM000000000, SRA: SRP265707; BioProject: PRJNA629068.

Datasets relating to the RNA-seq analysis have been deposited in Mendeley (doi:10.17632/9tsp7fr28r). The RNA-seq reads have been deposited at GenBank under the Bioproject (SRA) accession: PRJNA591919; Biosample accession: SAMN13391520.

Supplementary material is available at figshare DOI: https://doi.org/10.25387/g3.13270154.

Results and discussion

Sequencing and genome assembly

We sequenced and assembled a reference genome of the gray mangrove. Chicago and Dovetail HiC libraries produced 235 and 212 million 2 × 150 bp paired-end reads, respectively, and 134.1 Gb data overall. Genome scaffolding with HiRise yielded an assembly of 456.5 Mb; an L50/N50 equal to 15 scaffolds/13.98 Mb and an L90/N90 of 29 scaffolds/11.14 Mb; and a relatively large number of ambiguous bases (*i.e.,* N) inserted in the genome (10.6%; Table 1). The scaffold length distribution was heavily skewed toward extreme values (Table 1, Figure 2A). The 32 longest scaffolds, ranging from 22.40 to 10.58 Mb (median = 13.44 Mb) accounted for 98.03% of the whole assembled genome, congruent with a chromosome number 2N = 64 reported earlier (He *et al.* 2020). The remaining 1.97% of the genome was distributed among another 3,759 scaffolds ranging from 62.5 kb to over 1 kb (median = 1.8 kb; Figure 2A). The large number of small scaffolds may be due to the high fragmentation of the draft genome used in the assembly pipeline (Dovetail, personal communication). The integrity assessment of the *A. marina* genome retrieved a 98.8% and a 96.7% of the tested BUSCO groups for the eukaryote and the eudicots databases, respectively (Table 1, Figure 2B). The remarkable discontinuity in length sizes, as well as the integrity and high quality of the scaffolding lends considerable support to the hypothesis of 32 chromosomes; further sequencing efforts involving long-read sequencing are warranted for confirmation. The mitochondrial genome assembly was 22,019 pb long with a 46.4% GC content.

Genome annotation

We identified 45,032 protein-coding sequenced for which 35,604 showed homology with proteins from other species, and 34,442 were associated to GO (gene ontology) terms. The average gene length was 3.15 kb, with a mean of 5.2 exons and 4.2 introns per gene. BUSCO integrity analysis reported a 98.9% of recovered

Table 1 Summary statistics for the genome assembly and annotation of *A. marina*

Genome assembly	
Total length	456,526,188 bp
Number of scaffolds	3,791
N50/L50	13,979,447 bp/15 scaffolds
N90/L90	11,144,373 bp/29 scaffolds
Chromosome scale	10,583,658 bp/32 scaffolds
Longest scaffold	22,400,447 bp
Missingness	10.6%
GC content	35.2 %
BUSCO eukaryota database	C: 98.8% [S: 81.2%, D: 17.6%], F: 0.8%, M: 0.4%, N = 255
BUSCO eudicots database	C: 96.7% [S: 89.2%, D: 7.5%], F: 0.8%, M: 2.5%, N = 2,326

Genome annotation	
Number of genes	41,206
Number of annotated genes	35,604
Number of genes with GOs	34,442
Average gene length	3,152.28
Number of CDS	45,032
Average CDS length (bp)	1,097.74
Number of exons	233,312
Average exon length (bp)	211.87
Number of introns	188,280
Average intron length (bp)	536.98
BUSCO eukaryota database	C: 98.9% [S: 82.4%, D: 16.5%], F: 0.8%, M: 0.3%, N = 255
BUSCO eudicots database	C: 95.1% [S: 87.3%, D: 7.8%], F: 1.4%, M: 3.5%, N = 2,326

BUSCO parameters are C, complete BUSCO; S, complete and single-copy BUSCOs; D, complete and duplicated BUSCOs; F, fragmented BUSCOs; M, missing BUSCOs; N, total BUSCO groups searched. CDS indicates protein-coding sequences.

complete BUSCOs for the eukaryota database, and a 95.1% in the case of the eudicots (Table 1, Figure 2B). We also found that a total of 40.2% (188.5 Mb) of the *A. marina* assembly consisted of repetitive elements, a value moderately larger than the 30.4% previously reported for the species (Xu *et al.* 2017a). The greatest proportions corresponded to long terminal repeats and unclassified elements (20% and 16.7%, respectively; Figure 2A; Supplementary Table S2).

Adaptive variability analysis and functional assessment of A. marina genome

We resequenced 60 individuals of *A. marina* from six different populations across the environmentally diverse coasts of the Arabian Peninsula (Figure 1A, Supplementary Table S1) with a coverage of 85X. After SNP calling and a strict filtering for quality and missing data, we obtained a dataset of 538,185 SNPs for 56 individuals. A W&C Weighted F_{ST} scan based on sliding 20-kb windows revealed a heterogeneous landscape of differentiation and detected a peak of high divergence at the Scaffold 20 (Figure 1B). A total of 200 highly divergent loci were identified, from which 123 (61.5%) overlapped with 109 annotated genes associated to GO terms (Supplementary Table S3 and Figure S4). Importantly, we found signals of differentiation in genes involved in regulatory networks and biological pathways related to salinity stress response (*HDA19, NHL6,* and *ASG2*; Chen and Wu 2010; Bao *et al.* 2016; Dutilleul *et al.* 2016; Ueda *et al.* 2017); drought resistance and stomatal conductance (*NPK1* and *NCED*; Ning *et al.* 2008; Speirs *et al.* 2013); leaf cuticle development (*GPAT4*; Fabre *et al.* 2016); response to heat stress (*IDM1*; Zhao *et al.* 2014); sensitiveness to UV-B radiation (*SAD2*; Chen *et al.* 2013) and red light (*GIL1*; Allen *et al.* 2006); as well as other responses to abiotic stressors including root development and homeostasis maintenance
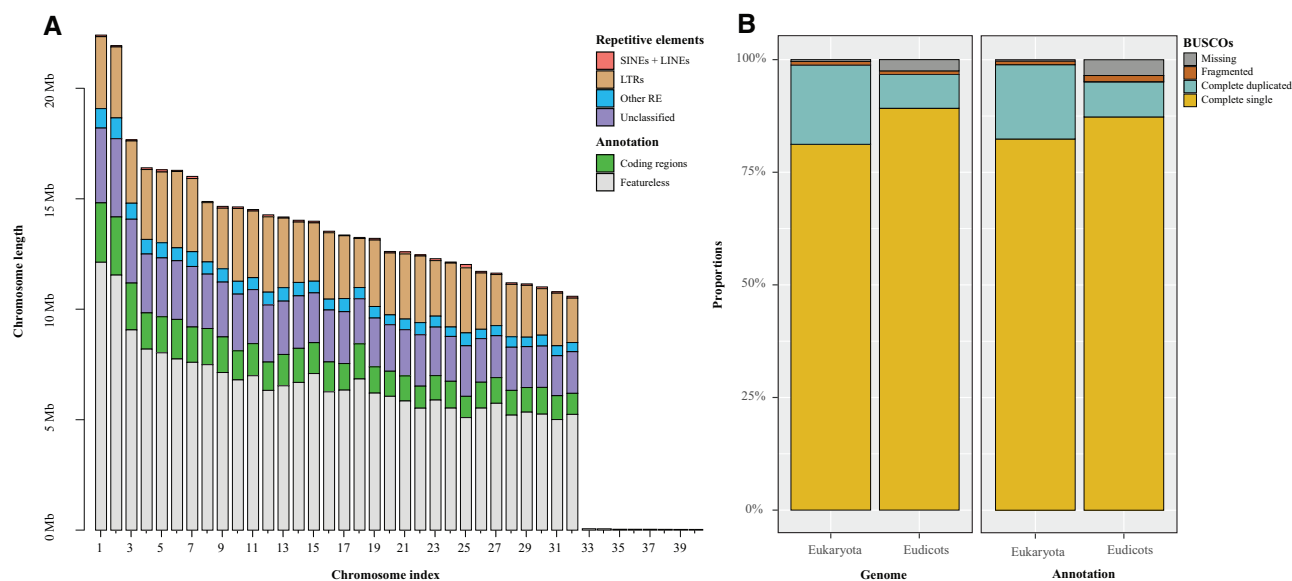
Figure 2 High-quality assembly and annotation for *A. marina*. (A) Length bar-plot of the longest 40 scaffolds arranged by decreasing size. Bars show per-chromosome genomic proportions corresponding to identified gene regions; and repetitive elements (SINEs + LINEs: short and long interspersed nuclear elements; LTRs, long terminal repeats). (B) BUSCO completeness percentages using the eukaryota ($N = 255$) and eudicots ($N = 2,326$) databases for the genome assembly (left) and for the annotation (right).

(*CTL1*; Hermans *et al.* 2010; Gao *et al.* 2017), and inhibition of responses to osmotic stress by means of chromatin condesantion (*SLK2*; Shrestha *et al.* 2014), supporting the role of environmental, divergent selective pressures in the differentiation of Arabian mangroves (Supplementary Table S3). A t-SNE based on 613 SNPs extracted from the functionally annotated, highly divergent loci identified in the F_{ST} scan showed clear clustering patterns among sampled populations (Figure 1C). Loading scores of retained t-SNE axes showed high correlation with the gradient of SST (*P*-values below 2.0×10^{-16} for both t-SNE1 and t-SNE2), also congruent with a differentiation process driven by environmental factors. In light of these results, further research into gene evolution and biological pathways involved in local adaptation to the extreme environment of the Arabian mangroves is in process. These questions are, however, beyond the scope of this report and thus will be presented elsewhere.

In conclusion, we report the first chromosome-scale assembly for the *A. marina* genome along with a comprehensive annotation based on tissue-specific RNA-seq data. The genome is highly contiguous and complete, and we demonstrated that it is a valuable resource for variant calling and the identification of functional, candidate genes underlying phenotypic and environmental divergence among mangrove taxa. Moreover, the successful identification of candidate genes potentially involved in local adaptation reveals the genome and annotation here reported as a relevant tool for the identification of biological pathways underlying molecular responses to extreme environmental conditions, which can in turn help to characterize adaptive mechanisms at the transcriptional and metabolomic levels. Improved scaffolding also enables the identification of regions putatively under selection, including structural variants such as chromosome rearrangements or copy number variations, all relevant for investigating questions related to evolutionary biology and molecular ecology in this ecological and socioeconomically important species.

Literature cited

Allen T, Ingles PJ, Praekelt U, Smith H, Whitelam GC. 2006. Phytochrome-mediated agravitropism in Arabidopsis hypocotyls requires GIL1 and confers a fitness advantage. Plant J. 46: 641–648.

Bao Y, Song W-M, Pan J, Jiang C-M, Srivastava R, *et al.* 2016. Overexpression of the NDR1/HIN1-like gene NHL6 modifies seed

germination in response to abscisic acid and abiotic stresses in Arabidopsis. PLos One. 11:e0148572.

Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT. 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics. 27:1691–1692.

Ben-Hasan A, Christensen V. 2019. Vulnerability of the marine ecosystem to climate change impacts in the Arabian Gulf—an urgent need for more research. Glob Ecol Conserv. 17:e00556.

Bruna T, Lomsadze A, Borodovsky M. 2020. GeneMark-EP and-EP+: automatic eukaryotic gene prediction supported by spliced aligned proteins. bioRxiv. 2019.2012. 2031.891218.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 12:59–60.

Burt JA, Coles S, Van Lavieren H, Taylor O, Looker E, et al. 2016. Oman's coral reefs: A unique ecosystem challenged by natural and man-related stresses and in need of conservation. Marine Pollution Bulletin. **105**:498–506.

Camp EF, Schoepf V, Mumby PJ, Hardtke LA, Rodolfo-Metalpa R, et al. 2018. The future of coral reefs subject to rapid climate change: lessons from natural extreme environments. Front Mar Sci. 5:4.

Carugati L, Gatto B, Rastelli E, Martire ML, Coral C, et al. 2018. Impact of mangrove forests degradation on biodiversity and ecosystem functioning. Sci Rep. 8:13298.

Carvalho S, Kürten B, Krokos G, Hoteit I, Ellis J. 2019. The Red Sea. In: C Sheppard, editor. World Seas: An Environmental Evaluation. Amsterdam, NL: Elsevier. p. 49–74.

Chen L-T, Wu K. 2010. Role of histone deacetylases HDA6 and HDA19 in ABA and abiotic stress response. Plant Signal Behav. 5: 1318–1320.

Chen M, Zhang W, Lv Z, Zhang S, Hidema J, et al. 2013. Abscisic acid is involved in the response of Arabidopsis mutant sad2-1 to ultraviolet-B radiation by enhancing antioxidant enzymes. South Afr J Bot. 85:79–86.

Claereboudt MR. 2019. Oman. In: C, editor. Sheppard World Seas: An Environmental Evaluation. Amsterdam, NL: Elsevier. p. 25–48.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, 1000 Genomes Project Analysis Group, et al. 2011. The variant call format and VCFtools. Bioinformatics. 27:2156–2158.

Das AB, Basak UC, Das P. 1995. Variation in nuclear DNA content and karyotype analysis in three species of Avicennia, a tree mangrove of coastal Orissa. Cytobios. 84:93–102.

Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45:e18.

Duarte CM, Losada IJ, Hendriks IE, Mazarrasa I, Marbà N. 2013. The role of coastal plant communities for climate change mitigation and adaptation. Nature Clim Change. 3:961–968.

Duke N. 1990. Phenological trends with latitude in the mangrove tree Avicennia marina. J Ecol. 78:113–133.

Duke N. 1991. A systematic revision of the mangrove genus Avicennia (Avicenniaceae) in Australasia. Aust Syst Bot. 4: 299–324.

Duke NC, Benzie JA, Goodall JA, Ballment ER. 1998. Genetic structure and evolution of species in the mangrove genus Avicennia (Avicenniaceae) in the Indo-West Pacific. Evolution. 52: 1612–1626.

Dutilleul C, Ribeiro I, Blanc N, Nezames CD, Deng XW, et al. 2016. ASG2 is a farnesylated DWD protein that acts as ABA negative regulator in Arabidopsis. Plant Cell Environ. 39:185–198.

Fabre G, Garroum I, Mazurek S, Daraspe J, Mucciolo A, et al. 2016. The ABCG transporter PEC1/ABCG32 is required for the formation of the developing leaf cuticle in Arabidopsis. New Phytol. 209: 192–201.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, et al. 2019., RepeatModeler2: automated genomic discovery of transposable element families. bioRxiv :856591.

Gao Y-Q, Chen J-G, Chen Z-R, An D, Lv Q-Y, et al. 2017. A new vesicle trafficking regulator CTL1 plays a crucial role in ion homeostasis. PLoS Biol. 15:e2002978.

Gotoh O. 2008. A space-efficient and accurate method for mapping and aligning cDNA sequences onto genomic sequence. Nucleic Acids Res. 36:2630–2638.

Gremme G, Steinbiss S, Kurtz S. 2013. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. IEEE/ACM Trans Comput Biol and Bioinf. 10: 645–656.

Guo Z, Li X, He Z, Yang Y, Wang W, et al. 2018. Extremely low genetic diversity across mangrove taxa reflects past sea level changes and hints at poor future responses. Glob Change Biol. 24: 1741–1748.

He Z, Xu S, Zhang Z, Guo W, Lyu H, The International Mangrove Consortium, et al. 2020. Convergent adaptation of the genomes of woody plants at the land-sea interface. Nat Sci Rev. 7:978–993.

Hermans C, Porco S, Verbruggen N, Bush DR. 2010. Chitinase-like protein CTL1 plays a role in altering root system architecture in response to multiple environmental conditions. Plant Physiol. 152:904–917.

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics. 32:767–769.

Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-genome annotation with BRAKER. In: Kollmar M, editor. Gene Prediction: Methods and Protocols. New York: Springer. p. 65–95.

Hogarth PJ. 2015. The Biology of Mangroves and Seagrasses. University of York, United Kingdom: Oxford University Press.

Iwata H, Gotoh O. 2012. Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. Nucleic Acids Res. 40:e161.

Jithesh M, Prashanth S, Sivaprakash K, Parida A. 2006. Monitoring expression profiles of antioxidant genes to salinity, iron, oxidative, light and hyperosmotic stresses in the highly salt tolerant grey mangrove, Avicennia marina (Forsk.) Vierh. by mRNA analysis. Plant Cell Rep. 25:865–876.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics. 30:1236–1240.

Kim D, Langmead B, Salzberg S. 2015. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods **12**, 357–360.

Lee SY, Primavera JH, Dahdouh-Guebas F, McKee K, Bosire JO, et al. 2014. Ecological role and services of tropical mangrove ecosystems: a reassessment. Glob Ecol Biogeogr. 23:726–743.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 25:1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, 1000 Genome Project Data Processing Subgroup, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics. 25:2078–2079.

Li X, Duke NC, Yang Y, Huang L, Zhu Y, et al. 2016. Re-evaluation of phylogenetic relationships among species of the mangrove genus Avicennia from Indo-West Pacific based on multilocus analyses. PLoS One. 11:e0164453.

Lomsadze A, Burns PD, Borodovsky M. 2014. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. Nucleic Acids Res. 42:e119.

Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. 2005. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 33:6494–6506.

Lyu H, He Z, Wu CI, Shi S. 2018. Convergent adaptive evolution in marginal environments: unloading transposable elements as a common strategy among mangrove genomes. New Phytol. 217: 428–438.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, *et al.* 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20: 1297–1303.

Mehta PA, Sivaprakash K, Parani M, Venkataraman G, Parida AK. 2005. Generation and analysis of expressed sequence tags from the salt-tolerant mangrove species *Avicennia marina* (Forsk) Vierh. Theor Appl Genet. 110:416–424.

Morrisey DJ, Swales A, Dittmann S, Morrison MA, Lovelock CE, *et al.* 2010. The Ecology and Management of Temperate Mangroves. Boca Raton, USA: CRC Press. p. 43–160.

Nagelkerken I, Blaber S, Bouillon S, Green P, Haywood M, *et al.* 2008. The habitat function of mangroves for terrestrial and marine fauna: a review. Aquat Bot. 89:155–185.

Naidoo G. 2016. The mangroves of South Africa: an ecophysiological review. South Afr J Bot. 107:101–113.

Nettel A, Dodd RS, Afzal-Rafii Z, Tovilla-Hernández C. 2008. Genetic diversity enhanced by ancient introgression and secondary contact in East Pacific black mangroves. Mol Ecol. 17:2680–2690.

Nguyen HT, Stanton DE, Schmitz N, Farquhar GD, Ball MC. 2015. Growth responses of the mangrove *Avicennia marina* to salinity: development and function of shoot hydraulic systems require saline conditions. Ann Bot. 115:397–407.

Ning J, Liu S, Hu H, Xiong L. 2008. Systematic analysis of NPK1-like genes in rice reveals a stress-inducible gene cluster co-localized with a quantitative trait locus of drought resistance. Mol Genet Genomics. 280:535–546.

Polidoro BA, Carpenter KE, Collins L, Duke NC, Ellison AM, *et al.* 2010. The loss of species: mangrove extinction risk and geographic areas of global concern. PLoS One. 5:e10095.

Primavera JH, Friess DA, Van Lavieren H, Lee SY. 2018. The Mangrove ecosystem. In: C, editor. Sheppard World Seas: An Environmental Evaluation: Volume II: The Indian Ocean to the Pacific. Academic Press. p. 18–51.

Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, *et al.* 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. 26:342–350.

Quisthoudt K, Schmitz N, Randin CF, Dahdouh-Guebas F, Robert EM, *et al.* 2012. Temperature variation among mangrove latitudinal range limits worldwide. Trees. 26:1919–1931.

Shrestha B, Guragain B, Sridhar VV. 2014. Involvement of co-repressor LUH and the adapter proteins SLK1 and SLK2 in the regulation of abiotic stress response genes in Arabidopsis. BMC Plant Biol. 14:54.

Smit A, Hubley R, Green P. 2015. RepeatMasker Open-4.0.2013-2015. Available at www.repeatmasker.org. (Accessed: 2020 April 2).

Spalding M, Kainuma M, Collins L. 2010. World Atlas of Mangroves. London, UK: Earthscan.

Speirs J, Binney A, Collins M, Edwards E, Loveys B. 2013. Expression of ABA synthesis and metabolism genes under different irrigation strategies and atmospheric VPDs is associated with stomatal conductance in grapevine (*Vitis vinifera* L. cv Cabernet Sauvignon). J Exp Bot. 64:1907–1916.

Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 24:637–644.

Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics. 7:62.

Tatusova TA, Madden TL. 1999. BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol Lett. 174:247–250.

Tomlinson PB. 2016. The Botany of Mangroves. Cambridge University, United Kingdom: Cambridge University Press.

Ueda M, Matsui A, Tanaka M, Nakamura T, Abe T, *et al.* 2017. The distinct roles of class I and II RPD3-like histone deacetylases in salinity stress response. Plant Physiol. 175:1760–1773.

UniProt Consortium 2019. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 47:D506–D515.

Urashi C, Teshima KM, Minobe S, Koizumi O, Inomata N. 2013. Inferences of evolutionary history of a widely distributed mangrove species, *Bruguiera gymnorrhiza*, in the Indo-West Pacific region. Ecol Evol. 3:2251–2261.

Vaughan GO, Al-Mansoori N, Burt JA. 2019. The Arabian Gulf. In: C Sheppard, editor. World Seas: An Environmental Evaluation. Amsterdam, NL: Elsevier. p. 1–23.

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, *et al.* 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol. 35:543–548.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution. 38:1358–1370.

Xu S, He Z, Guo Z, Zhang Z, Wyckoff GJ, *et al.* 2017a. Genome-wide convergence during evolution of mangroves from woody plants. Mol Biol Evol. 34:1008–1015.

Xu S, He Z, Zhang Z, Guo Z, Guo W, The International Mangrove Consortium, *et al.* 2017b. The origin, diversification and adaptation of a major mangrove clade (Rhizophoreae) revealed by whole-genome sequencing. Nat Sci Rev. 4:721–734.

Zhao Y, Xie S, Li X, Wang C, Chen Z, *et al.* 2014. REPRESSOR OF SILENCING5 encodes a member of the small heat shock protein family and is required for DNA demethylation in Arabidopsis. Plant Cell. 26:2660–2675.

Zhou RC, Zeng K, Wu W, Chen XS, Yang ZH, *et al.* 2007. Population genetics of speciation in nonmodel organisms: I. Ancestral polymorphism in mangroves. Mol Biol Evol. 24:2746–2754.

Communicating editor: P. Ingvarsson