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

Running title: Genome assembly of the gray mangrove

Guillermo Friis[1*], Joel Vizueta[2], 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], 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, Spain*

[3]*Center for Desert Agriculture, Division of Biological and Environmental Sciences and Engineering, 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: Center for Genomics and Systems Biology, New York University — Abu Dhabi, PO Box 129188, Abu Dhabi, United Arab Emirates; Email: gfm3@nyu.edu; Tel: +97126286739.

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 sparse. We sequenced and assembled the genome of *A. marina* from the Arabian Gulf, which is the harshest region that the species occupies and at the northern-most limit of its distribution. We used proximity ligation libraries Chicago and Dovetail HiC, and the HiRise assembly pipeline, producing a 456,556,596 bp long genome. The largest 32 scaffolds (22.4 Mb to 10.5 Mb) accounted for 98 % of the genome assembly, with the remaining 2% distributed among much shorter 3,777 scaffolds (62.4 Kb to 1 Kb). We annotated 23,331 protein-coding genes using tissue-specific RNA-seq data, from which 13,312 were associated to GO terms. Genome assembly and annotated set of genes yield a 96.7% and 92.3% 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.

Key words: gray mangrove, *Avicennia marina*, genome assembly, HiC, Arabia

Significance statement

This study reports the first chromosome-level genome assembly for *Avicennia marina*, the most widely distributed mangrove species of the world. Our new assembly provides an excellent reference for phylogenomics and population genomics studies focused on *A. marina* and related species. An annotation based on tissue-specific mRNA data of unprecedent quality is also provided, enabling functional identification of candidate genes. Overall, the data here reported enables a range of comparative genomics, molecular ecology, and functional divergence tools for the study of an ecological and socioeconomically relevant group of organisms.

Introduction

Mangroves are a polyphyletic group of trees and shrubs that inhabit the intertidal zone of the tropic and sub-tropic coasts (Hogarth 2015; Nagelkerken, et al. 2008; Polidoro, et al. 2010; Primavera, et al. 2018). Mangroves share several morphological and physiological adaptations to their harsh intertidal 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, and shelter areas for a range of biota, as well as serving as a foraging habitat (Carugati, et al. 2018; Lee, et al. 2014; Nagelkerken, et al. 2008; 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 provide compelling systems for studying patterns of trait evolution, lineage divergence and speciation (e.g., Urashi, et al. 2013; Xu, et al. 2017b; Zhou, et al. 2007).

Of the approximately 70 mangrove species described, the gray mangrove *Avicennia marina* has the broadest latitudinal and longitudinal distribution (Hogarth 2015; Spalding, et al. 2010; Tomlinson 2016). At least three sub-specific, 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

85 presence across diverse environmental gradients (e.g. temperature, freshwater, sediment and
86 nutrient supply, salinity, tidal range) and spatial settings (e.g. open coastlines, coastal lagoons,
87 estuaries, deltas, coral fringes) (Duke 1990; Quisthoudt, et al. 2012). Along with its widespread
88 distribution, several aspects of *A. marina* biology make it a promising model organism among
89 mangroves for the study of evolution based on genomic and molecular approaches. First, the
90 broad environmental gradients present across the *A. marina* range are mirrored by remarkable
91 geographic variation in phenotypic and life-history traits (Duke 1990; Tomlinson 2016), all of
92 which makes them an ideal natural system for studying evolutionary processes related to
93 dispersal, directional selection, and neutral evolution. Previous studies have reported the
94 phylogenetic relationships for the sub-varieties of *A. marina* and congeneric species (Duke, et al.
95 1998; Li, et al. 2016; Nettel, et al. 2008), yet the specific drivers of lineage diversification remain
96 understudied. Furthermore, the extensive gray mangrove populations from the Eastern African
97 and Arabian coasts have rarely been included in reported DNA sequence-based analyses, either
98 for diversity screening or other purposes (e.g., Duke, et al. 1998; Li, et al. 2016). Second, *A.*
99 *marina* can tolerate highly variable and extreme environmental conditions, and often occupies
100 marginal, biologically limiting environments at the edges of its distribution (Morrisey, et al.
101 2010). While several biological structures and mechanisms of the gray mangrove physiology
102 have been described (Naidoo 2016; Nguyen, et al. 2015), the genetic basis and pathways
103 underlying such tolerance are still largely unknown (Jithesh, et al. 2006; Mehta, et al. 2005; Xu,
104 et al. 2017a). Understudied Arabian populations represent a particular gap due to their
105 occurrence at the extremes of air and water temperature, salinity, and aridity (Ben-Hasan and
106 Christensen 2019; Camp, et al. 2018). Third, the *A. marina* genome is moderately small and
107 structurally simple compared with other eukaryotes, and presents a limited amount of

transposable elements (Das, et al. 1995; Lyu, et al. 2018; Xu, et al. 2017a), 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 multi-species 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 (http://evolution.sysu.edu.cn/). 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 near-chromosome level assembly obtained using proximity ligation libraries as a resource for genome-based studies on *A. marina* and related mangrove species. An structural and functional annotation based on RNA-seq data from multiple tissues is also provided. We used resequencing data from multiple individuals to evaluate the performance of the assembly as a reference to study patterns of adaptive variability at the genomic level. Resequencing data is also used for assembling the mitochondrial genome.

Materials and Methods

Genome sequencing and assembly

The sequenced sample was leaf tissue obtained from an individual located at Ras Ghurab Island

in the Arabian Gulf (Fig 1; Abu Dhabi, United Arab Emirates; 24.601°N, 4.566 °E), corresponding to the *A. m. marina* variety. A high-quality genome was produced using proximity ligation libraries and the software pipeline HiRise (Putnam, et al. 2016) at Dovetail Genomics, LLC. Briefly, for Chicago and the Dovetail HiC library 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 to scaffold genome assemblies (Putnam, et al. 2016). A previously reported draft genome of *Avicennia marina* (Lyu, et al. 2018; Xu, et al. 2017a; available at evolution.sysu.edu.cn/index.html) was used in the assembly pipeline, excluding scaffolds shorter than 1Kb 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.

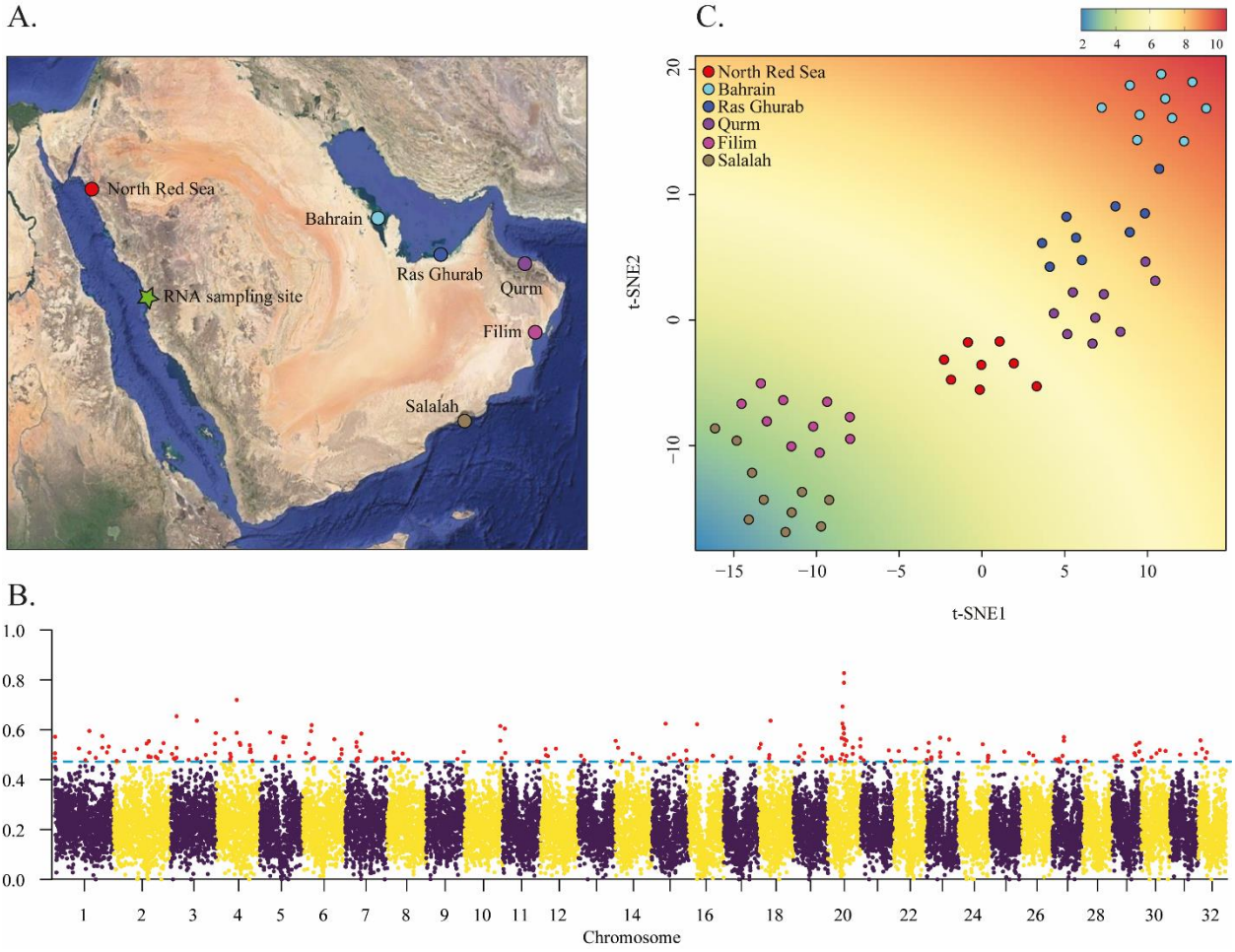
Genome annotation

We performed the annotation of the *A. marina* genome using mRNA data from a set of tissues of conspecific individuals. 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 a HiSeq2000 platform (Illumina, Inc.). Messenger RNA reads were mapped with HISAT2 (Kim, et al. 2015), and genome-referenced transcripts for each tissue were produced and merged with StringTie (Pertea, et al. 2015). Prediction of coding regions was performed with TransDecoder (Haas and Papanicolaou 2015). 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. The Trinotate (Haas 2015) pipeline was then implemented to conduct a homology-based functional annotation by using Swiss-Prot (Bairoch and Apweiler 2000) and pfam (Bateman, et al. 2002) databases, generating a final set of annotated functional genes. Further details on mRNA sequencing and annotation scripts are provided in the Supplementary Information.

Repetitive regions were first modelled *ab initio* using RepeatModeler v2.0.1 (Flynn, et al. 2019) in all scaffolds longer than 100 Kbp 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).

Fig. 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 200 SNP outliers linked to functional genes. The background shows the correlation between 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.



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 2,326 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 populations from each of the major seas bordering Arabia (Figure 1A; Table S1, Supplementary Information), including populations in the Red Sea, the Arabian Gulf and the Arabian Sea/Sea of Oman. Arabia's regional seas are characterized by extreme, but divergent, environmental conditions for mangrove growth. The northern Red Sea is characterized by cold 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) of 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 (Claereboudt 2019). Using vcftools (Danecek, et al. 2011), we conduct an F_{ST} survey across the six populations based on 20 Kb sliding windows and identifiy outlier loci associated with functional genes. We then use these loci to explore geographic patterns of adaptive variability by means of t-SNE analysis, testing for correlations between variability in sea surface temperature and t-SNE scores. Details on sequencing, variant calling, and analytical procedures are available in the Supplementary Information.

Results and Discussion

Sequencing and genome assembly

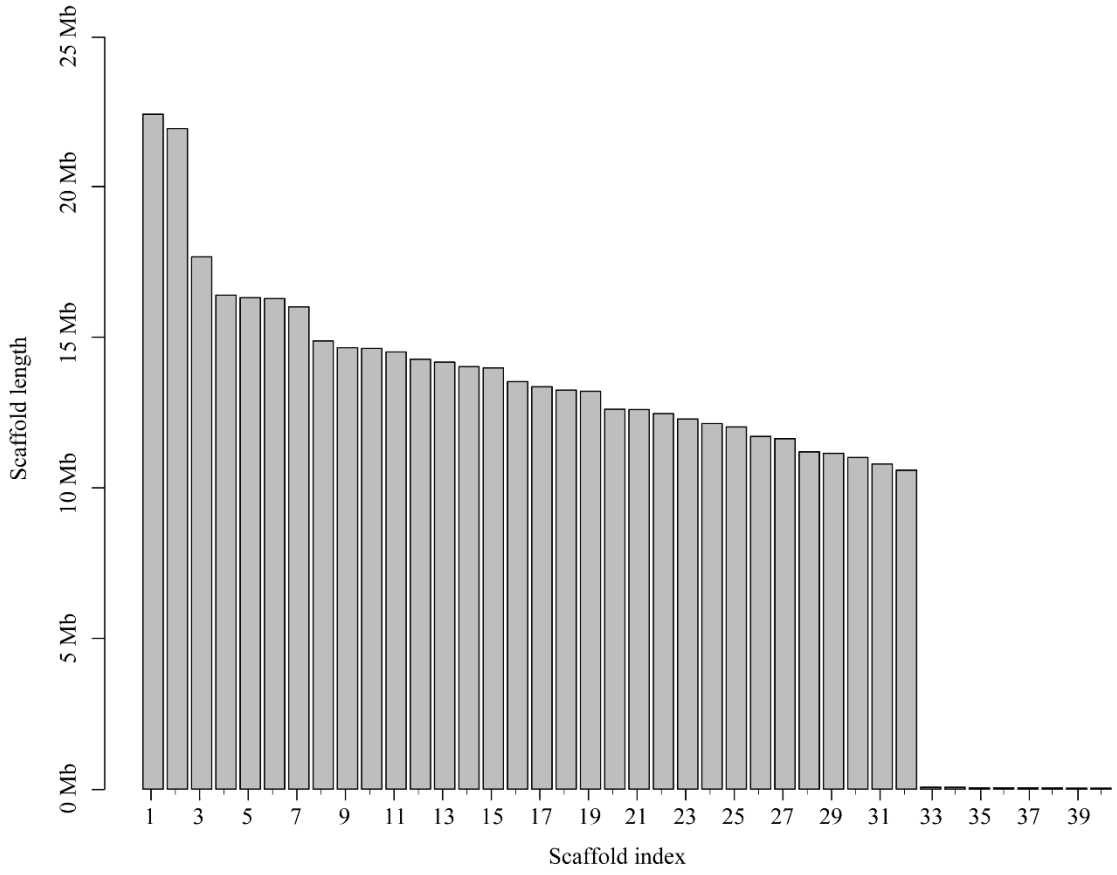
We sequenced and assembled a reference genome of a gray mangrove individual from the Arabian Gulf, an extreme environment at the northern limit of the species' distribution. Chicago and Dovetail HiC libraries produced 235 million and 212 million 2x150 bp paired-end reads, respectively, and 134.1 Gb data overall. Genome scaffolding with HiRise yielded an assembly of 456.6 Mb for a final sequence coverage of 293X; an L50/N50 equal to 15 scaffolds/13.98 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 towards extreme values (Table 1, Figure 2). The 32 longest scaffolds, ranging from 22.40 Mb 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 3,777 scaffolds ranging from 62.5 Kb to over 1 Kb (median = 1.8 Kb). The large number of small scaffolds may be due to the high fragmentation of the draft genome used in the assembly pipeline (Dovetails, 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). 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.

Table 1. Summary statistics for the genome assembly and annotation of *A. marina*. CDS indicates protein-coding sequences.

Genome assembly	
Total length	456,556,596 bp
Number of scaffolds	3,809
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 scaffolds	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	23,331
Number of annotated genes	21,147
Number of genes with GOs	13,312
Average gene length	5,383.7 bp
Number of CDS	59,445
Average CDS length (bp)	11,46.33 bp
Number of exons	493,357
Average exon length (bp)	307.06 bp
Number of introns	433,912
Average intron length (bp)	625.19 bp
BUSCO eukaryota database	C:97.2% [S:78.8%, D:18.4%], F:1.6%, M:1.2%, N:255
BUSCO eudicots database	C:92.3% [S:85.2%, D:7.1%], F:1.6%, M:6.1%, 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; and N: Total BUSCO groups searched. CDS indicates protein-coding sequences.

Fig. 2. Length bar-plot of the longest 40 scaffolds arranged by decreasing size. The genome was sequenced using proximity ligation libraries Chicago and Dovetails HiC, and the assembly was carried out with HiRise pipeline.



Genome annotation

We identified 23,331 protein-coding genes for which 21,147 orthologs from other species were identified, and 13,312 were associated to GO terms. The average gene length was 5.38 Kb, with a mean of 8.3 exons and 18.6 introns per gene. BUSCO integrity analysis reported a 97.2% of recovered BUSCOs for the eukaryota database, and a 92.3% in the case of the eudicots (Table 1). We also found that a total of 40.20% (188.5 Mb) of the *A. marina* assembly consisted of

255 repetitive elements, a value moderately larger than the 30.4% previously reported for the species
256 (Xu, et al. 2017a). The greatest proportions corresponded to long terminal repeats and
257 unclassified elements (20% and 16.7%, respectively; Table S2, Supplementary Information).
258
259 *Adaptive variability analysis and functional assessment of* A. marina *genome*
260 We resequenced 56 individuals of *A. marina* from six different populations across the
261 environmentally diverse coasts of the Arabian Peninsula (Figure 1A) with a coverage of 85X.
262 After SNP calling and a strict quality filtering, we obtained a dataset of 538,185 SNPs. An F_{ST}
263 scan based on sliding 20 Kb windows revealed a heterogeneous landscape of differentiation and
264 detected a peak of high divergence at the Scaffold 20 (Figure 1B). A total of 200 highly
265 divergent loci were identified, from which 43 overlapped with annotated genes associated to GO
266 terms (Table S2, Figure S1). Several of these genes are involved in the development of shoots,
267 leaves and flowers (BAM2 and BAM7; DeYoung, et al. 2006; Reinhold, et al. 2011), root and
268 seeds (FEI1; Basu, et al. 2016), as well as in protein storage (VTI12; Sanmartín, et al. 2007).
269 Importantly, we also found signals of differentiation in genes involved in plant sensitivity to salt
270 and osmotic stress (WRKY40; Chen, et al. 2010), drought and palatability to detritivorous
271 crustaceans (LOX6; Grebner, et al. 2013), supporting the role of abiotic and biotic factors in the
272 differentiation of Arabian mangroves (Table S3, Supplementary Information). A t-SNE based
273 324 SNPs extracted from the functionally annotated, highly divergent loci showed clear
274 clustering patterns among sampled populations (Figure 1C). Loading scores of retained t-SNE
275 axes revealed a high correlation with the gradient of sea surface temperature (SST; p-values
276 below 8.3×10^{-14} and 2.0×10^{-16} for t-SNE1 and t-SNE2, respectively), also congruent with a

277 pattern of adaptive divergence driven by environmental factors. Further details on analytical
278 procedures are reported in the Supplementary Information.
279
280 In conclusion, we report the first chromosome-scale assembly for the *Avicennia marina* genome
281 along with a comprehensive annotation based on tissue-specific RNA-seq data. The genome is
282 highly contiguous and complete, and we demonstrated that it is valuable resource for variant
283 calling and the identification of functional, candidate genes underlying phenotypic and
284 environmental divergence among mangrove taxa. Improved scaffolding also enables the
285 identification of regions putatively under selection, including structural variants such as
286 chromosome rearrangements or copy number variations, all relevant for investigating questions
287 related to evolutionary biology and molecular ecology in this ecological and socioeconomically
288 important species.
289
290
291 **Data deposition**
292 Genome assembly and annotation, and multi-sample sequence alignment from resequenced
293 individuals have been deposited at DRYAD and can be accesed at DRYAD
294 (doi:10.5061/dryad.3j9kd51f5). The genome assembly has also been deposited at
295 DDBJ/ENA/GenBank under the accession JABGBM000000000. The version described in this
296 paper is version JABGBM010000000. Bioproject (SRA) accession: PRJNA629068; Biosample
297 accession: SAMN14766548.
298

299 Datasets relating to the RNA-seq analysis have been deposited in Mendeley
300 (doi:10.17632/9tsp7fr28r). The RNA-seq reads have been deposited at the National Center for
301 NCBI under the Bioproject (SRA) accession: PRJNA591919, Biosample accession:
302 SAMN13391520.
303
304
305 **Acknowledgements**
306 The authors are grateful to NYUAD's Core Bioinformatics for analysis and computational
307 assistance. We thank Mark Priest, Dain McParland, Noura Al-Mansoori, Anique Ahmad, and
308 Ada Kovaliukaite for their participation in the fieldwork. We gratefully acknowledge the
309 Environment Agency Abu Dhabi (permit reference number 20181823a), and the Oman Ministry
310 of Environment and Climate Affairs, Director General of Nature Conservation (permit number:
311 6210/10/75), for providing permits for research related to this manuscript. This work was
312 supported by New York University Abu Dhabi Institute Grant (73 71210 CGSB9), NYUAD
313 Faculty Research Funds (AD060) and King Abdullah University of Science and Technology
314 baseline funding to CMD. We dedicate this manuscript to the memory of Basel Khraiwesh, who
315 had produced the RNA-seq data used for annotation in this study.
316

References

Bairoch A, Apweiler R 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Research 28: 45-48.

Basu D, et al. 2016. Glycosylation of a fasciclin-like arabinogalactan-protein (SOS5) mediates root growth and seed mucilage adherence via a cell wall receptor-like kinase (FEI1/FEI2) pathway in Arabidopsis. PLOS One 11.

Bateman A, et al. 2002. The Pfam protein families database. Nucleic Acids Research 30: 276-280.

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. Global ecology and conservation 17: e00556.

Camp EF, et al. 2018. The future of coral reefs subject to rapid climate change: lessons from natural extreme environments. Frontiers in Marine Science 5: 4.

Carugati L, et al. 2018. Impact of mangrove forests degradation on biodiversity and ecosystem functioning. Scientific Reports 8: 13298.

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

Chen H, et al. 2010. Roles of Arabidopsis WRKY18, WRKY40 and WRKY60 transcription factors in plant responses to abscisic acid and abiotic stress. BMC plant biology 10: 281.

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

Danecek P, et al. 2011. The variant call format and VCFtools. Bioinformatics 27: 2156-2158.

339 Das AB, Basak UC, Das P 1995. Variation in nuclear DNA content and karyotype analysis in
340 three species of Avicennia, a tree mangrove of coastal Orissa. Cytobios 84: 93-102.
341 DeYoung BJ, et al. 2006. The CLAVATA1-related BAM1, BAM2 and BAM3 receptor kinase-
342 like proteins are required for meristem function in Arabidopsis. The Plant Journal 45: 1-16.
343 Dierckxsens N, Mardulyn P, Smits G 2017. NOVOPlasty: de novo assembly of organelle
344 genomes from whole genome data. Nucleic Acids Research 45: e18-e18.
345 Duarte CM, Losada IJ, Hendriks IE, Mazarrasa I, Marbà N 2013. The role of coastal plant
346 communities for climate change mitigation and adaptation. Nature Climate Change 3: 961-968.
347 Duke N 1990. Phenological trends with latitude in the mangrove tree Avicennia marina. The
348 Journal of Ecology: 113-133.
349 Duke N 1991. A systematic revision of the mangrove genus Avicennia (Avicenniaceae) in
350 Australasia. Australian Systematic Botany 4: 299-324.
351 Duke NC, Benzie JA, Goodall JA, Ballment ER 1998. Genetic structure and evolution of species
352 in the mangrove genus Avicennia (Avicenniaceae) in the Indo-West Pacific. Evolution 52: 1612-
353 1626.
354 Flynn JM, et al. 2019. RepeatModeler2: automated genomic discovery of transposable element
355 families. bioRxiv: 856591.
356 Grebner W, Stingl NE, Oenel A, Mueller MJ, Berger S 2013. Lipoxygenase6-dependent oxylipin
357 synthesis in roots is required for abiotic and biotic stress resistance of Arabidopsis. Plant
358 Physiology 161: 2159-2170.
359 Gremme G, Steinbiss S, Kurtz S 2013. GenomeTools: a comprehensive software library for
360 efficient processing of structured genome annotations. IEEE/ACM Transactions on
361 Computational Biology and Bioinformatics 10: 645-656.

362 Guo Z, et al. 2018. Extremely low genetic diversity across mangrove taxa reflects past sea level
363 changes and hints at poor future responses. Global Change Biology 24: 1741-1748.
364 Haas B. 2015. Trinotate: Transcriptome functional annotation and analysis. In.
365 Haas B, Papanicolaou A 2015. TransDecoder (find coding regions within transcripts). Github, nd
366 https://github. com/TransDecoder/TransDecoder (accessed May 17, 2018).
367 He Z, et al. 2020. Convergent adaptation of the genomes of woody plants at the land-sea
368 interface. National Science Review.
369 Hogarth PJ. 2015. The biology of mangroves and seagrasses: Oxford University Press.
370 Jithesh M, Prashanth S, Sivaprakash K, Parida A 2006. Monitoring expression profiles of
371 antioxidant genes to salinity, iron, oxidative, light and hyperosmotic stresses in the highly salt
372 tolerant grey mangrove, Avicennia marina (Forsk.) Vierh. by mRNA analysis. Plant cell reports
373 25: 865-876.
374 Kim D, Langmead B, Salzberg S 2015. hisat2. Nature methods.
375 Lee SY, et al. 2014. Ecological role and services of tropical mangrove ecosystems: a
376 reassessment. Global Ecology and Biogeography 23: 726-743.
377 Li X, et al. 2016. Re-evaluation of phylogenetic relationships among species of the mangrove
378 genus Avicennia from Indo-West Pacific based on multilocus analyses. PLOS One 11:
379 e0164453.
380 Lyu H, He Z, Wu CI, Shi S 2018. Convergent adaptive evolution in marginal environments:
381 unloading transposable elements as a common strategy among mangrove genomes. New
382 phytologist 217: 428-438.

383 Mehta PA, Sivaprakash K, Parani M, Venkataraman G, Parida AK 2005. Generation and
384 analysis of expressed sequence tags from the salt-tolerant mangrove species Avicennia marina
385 (Forsk) Vierh. Theoretical and Applied Genetics 110: 416-424. doi: 10.1007/s00122-004-1801-y
386 Morrisey DJ, et al. 2010. The ecology and management of temperate mangroves. In: CRC Press,
387 Boca Raton(USA). p. 43-160.
388 Nagelkerken I, et al. 2008. The habitat function of mangroves for terrestrial and marine fauna: a
389 review. Aquatic Botany 89: 155-185.
390 Naidoo G 2016. The mangroves of South Africa: An ecophysiological review. South African
391 Journal of Botany 107: 101-113.
392 Nettel A, Dodd RS, Afzal-Rafii Z, Tovilla-Hernández C 2008. Genetic diversity enhanced by
393 ancient introgression and secondary contact in East Pacific black mangroves. Molecular Ecology
394 17: 2680-2690.
395 Nguyen HT, Stanton DE, Schmitz N, Farquhar GD, Ball MC 2015. Growth responses of the
396 mangrove Avicennia marina to salinity: development and function of shoot hydraulic systems
397 require saline conditions. Annals of Botany 115: 397-407.
398 Pertea M, et al. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-
399 seq reads. Nature Biotechnology 33: 290.
400 Polidoro BA, et al. 2010. The loss of species: mangrove extinction risk and geographic areas of
401 global concern. PLOS One 5: e10095.
402 Primavera JH, Friess DA, Van Lavieren H, Lee SY. 2018. The Mangrove Ecosystem. In:
403 Sheppard C, editor. World Seas: An Environmental Evaluation: Volume II: The Indian Ocean to
404 the Pacific: Academic Press. p. 18-51.

405 Putnam NH, et al. 2016. Chromosome-scale shotgun assembly using an in vitro method for long-
406 range linkage. Genome Research 26: 342-350.
407 Quisthoudt K, et al. 2012. Temperature variation among mangrove latitudinal range limits
408 worldwide. Trees 26: 1919-1931.
409 Reinhold H, et al. 2011. β-Amylase–Like proteins function as transcription factors in
410 Arabidopsis, controlling shoot growth and development. The Plant Cell 23: 1391-1403.
411 Sanmartín M, et al. 2007. Divergent functions of VTI12 and VTI11 in trafficking to storage and
412 lytic vacuoles in Arabidopsis. Proceedings of the National Academy of Sciences 104: 3645-
413 3650.
414 Smit A, Hubley R, Green P. 2015. RepeatMasker Open-4.0. 2013–2015. In.
415 Spalding M, Kainuma M, Collins L. 2010. World atlas of mangroves. London, UK: Earthscan.
416 Tomlinson PB. 2016. The botany of mangroves: Cambridge University Press.
417 Urashi C, Teshima KM, Minobe S, Koizumi O, Inomata N 2013. Inferences of evolutionary
418 history of a widely distributed mangrove species, Bruguiera gymnorrhiza, in the Indo-West
419 Pacific region. Ecology and Evolution 3: 2251-2261. doi: 10.1002/ece3.624
420 Vaughan GO, Al-Mansoori N, Burt JA. 2019. The Arabian Gulf. In: Sheppard C, editor. World
421 Seas: An Environmental Evaluation: Elsevier. p. 1-23.
422 Waterhouse RM, et al. 2018. BUSCO applications from quality assessments to gene prediction
423 and phylogenomics. Molecular Biology and Evolution 35: 543-548.
424 Xu S, et al. 2017a. Genome-wide convergence during evolution of mangroves from woody
425 plants. Molecular Biology and Evolution 34: 1008-1015.

426 Xu S, et al. 2017b. The origin, diversification and adaptation of a major mangrove clade
427 (Rhizophoreae) revealed by whole-genome sequencing. National Science Review 4: 721-734.
428 doi: 10.1093/nsr/nwx065
429 Zhou RC, et al. 2007. Population genetics of speciation in nonmodel organisms: I. Ancestral
430 polymorphism in mangroves. Molecular Biology and Evolution 24: 2746-2754. doi:
431 10.1093/molbev/msm209

432