Rapid lineage diversification of gray mangroves (Avicennia marina) driven by isolation in cryptic glacial refugia and extreme environmental conditions in the Arabian Peninsula

Guillermo Friis¹, Edward G. Smith², Catherine E. Lovelock³, Alejandra Ortega⁴, Alyssa Marshall⁵, Carlos M. Duarte⁶, and John A. Burt¹

¹Center for Genomics and Systems Biology, New York University - Abu Dhabi, PO Box 129188, Abu Dhabi, United Arab Emirates
²Department of Biological Sciences, University of North Carolina at Charlotte, Charlotte, NC 28262
³School of Biological Sciences, The University of Queensland, St Lucia, QLD 4072, Australia
⁴Red Sea Research Center (RSRC) and Computational Bioscience Research Center, King Abdullah
⁵Department of Marine Science and Fisheries, College of Agricultural and Marine Sciences, Sultan Qaboos University, Muscat, Oman
⁶Red Sea Research Center (RSRC) and Computational Bioscience Research Center, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900, Saudi Arabia

November 11, 2022

Running title: Evolution of Arabian mangroves

Guillermo Friis*, Edward G. Smith, Catherine E. Lovelock, Alejandra Ortega, Alyssa Marshall, Carlos M. Duarte, John A. Burt

*Corresponding author: Center for Genomics and Systems Biology, New York University — Abu Dhabi, PO Box 129188, Abu Dhabi, United Arab Emirates; Email: guillefriis@gmail.com; Tel: +97126286739.

Summary

- Plant systems occurring in ecologically heterogeneous and spatially discontinuous habitats provide an ideal opportunity to investigate the relative roles of neutral and selective factors in driving lineage diversification. Here, we analyzed fully sequenced genomes to study diversification mechanisms in the gray mangroves [Avicennia marina (Forssk.) Vierh.] of Arabia, where they occur at the edge of the species’ range and are subject to variable, often extreme, environmental conditions.

- We conducted population structure, phylogenomic and demographic analyses to reconstruct the evolutionary history of the species across Arabia. We also applied genotype-environment association methods to study adaptive mechanisms of lineage diversification.
Our analyses revealed marked genetic structure and highly supported clades among and within the seas surrounding the Arabian Peninsula. Inferred divergence times were consistent with recent periods of low marine connectivity during glacial periods, revealing the presence of (cryptic) glacial refugia in the Red Sea and the Persian/Arabian Gulf. Genetic-environment association analyses revealed high levels of adaptive differentiation, and detected signs of multi-loci local adaptation driven by temperature extremes and hypersalinity.

* These results support a process of rapid diversification resulting from the combined effects of historical factors and ecological selection, and reveal mangrove peripheral environments as relevant drivers of lineage diversity.

Introduction

Lineage diversification in plants involves both neutral and selective factors (Rieseberg & Willis, 2007), and elucidating their relative roles in the process of evolutionary divergence is essential to understand the mechanisms underlying the early stages of speciation (Coyne & Orr, 2004). Evolutionary divergence may result from the accumulation of genetic differences caused by drift in geographic isolation or isolation-by-distance (IBD, Wright, 1943; Wright, 1946), a mode of divergence driven by neutral factors (Mayr, 1954; Mayr, 1963). In turn, geographic variation in environmental conditions can result in divergent selection (Darwin, 1859; Coyne & Orr, 2004), the diversifying process that drives ecological speciation (Nosil, 2012). In ecological speciation models, reproductive barriers arise as a by-product of cumulative, ecologically adaptive changes (Mayr, 1947; Schluter, 2000; Rundle & Nosil, 2005) enabling genome-wide differentiation at both neutral and selected loci (Nosil et al., 2008; Funk et al., 2011; Shafer & Wolf, 2013; Wang & Bradburd, 2014). Ecological speciation in geographic isolation is theoretically uncontroversial, and considered key in some of the most remarkable radiations in angiosperms (Baldwin & Sanderson, 1998; Hughes & Eastwood, 2006). However, whether environment-driven processes of lineage diversification commonly occur in nature in the absence of long-term isolation and reduced gene flow remains debated in evolution research (Butlin et al., 2008; Fitzpatrick et al., 2008; Papadopulos et al., 2011; Foote, 2018). The interactions between selection and the stochastic effects derived from processes such as founder events, bottlenecks and genetic drift also remain unclear, and difficult to assess in natural systems (Barton & Charlesworth, 1984; Kliber & Eckert, 2005; Crepet & Niklas, 2009).

Plant systems occurring in the environmentally heterogeneous and often extreme habitats at species’ range edges are suitable models to investigate questions related to lineage diversification. The environment at the edges of species’ range tends to be stressful and spatially discontinuous, as well as temporally unstable (Lesica & Allendorf, 1995), frequently resulting in dynamic settings of multiple isolated populations subject to strong differential selection. The severe and stochastic character of peripheral environments is hypothesized to generate strong selective interplay between adaptation and neutral processes (Hardie & Hutchings, 2010), providing an ideal opportunity for speciation research. One such system is the gray mangrove populations of the Arabian Peninsula (Avicennia marina var. marina). The gray mangrove has the broadest distribution of any mangrove species (Spalding et al., 2010; Hogarth, 2015; Tomlinson, 2016), extending across the Indian Ocean and into the West Pacific as far as Japan and New Zealand (Fouda & Al-Muharrami, 1996; Sheppard et al., 2010; Spalding et al., 2010; Khalil, 2015). Gray mangroves present several morphological and physiological adaptations to their harsh intertidal habitat (Tomlinson, 2016), which makes them a compelling model for the study of functional genes and biological pathways involved in selection and stress tolerance in plants (Urashi et al., 2013; Xu et al., 2017). The Arabian Peninsula represents one of the northernmost edges of the species’ distribution (Duke, 1991; Spalding et al., 2010; Tomlinson, 2016), as well as a stressful habitat characterized by extreme temperatures, aridity, and often extreme salinity, factors known to be limiting for mangrove growth (Ball, 1988; Sheppard et al., 1992; Lovelock et al., 2016). Arabian marine domains are also environmentally diverse both within, and among, the main water bodies bordering the peninsula, which define three main biogeographic regions: (i) the Red Sea, where the marine system presents opposing gradients of salinity and temperature, with the highest temperature and lowest salinity
in the shallow southern basin, while the deeper northern basin has cooler temperatures but high salinity as a result of limited precipitation and high evaporation (Carvalho et al., 2019; Antón et al., 2020); (ii) the Persian/Arabian Gulf (referred to as ‘PAG’ hereafter) to the northeast of the Arabian Peninsula, where populations are subject to arid (<250 mm) to hyper-arid (<100 mm) rainfall regimes, and experience the widest range of air temperatures in the region throughout the year (Boer, 1997; Whitford & Duval, 2019); and (iii) the Arabian Sea (here including the Sea of Oman), which in contrast with former biogeographic regions, has normal oceanic salinity, and summer temperatures that are buffered by cold-water upwelling as a result of the Indian Ocean monsoon, resulting in more moderate environmental conditions (Claereboudt, 2019).

The Arabian Peninsula has experienced large fluctuations in spatial and environmental conditions throughout glacio-eustatic cycles that largely impacted the biodiversity of the region, in particular for the enclosed water bodies of the Red Sea and the PAG (DiBattista et al., 2016a). Throughout the last 400,000 years the Red Sea has remained connected to the Indian Ocean, yet cross-sectional area along the Strait of Bab al Mandab that connects these water bodies was, at times of glacial maxima, as low as 2% of that today, resulting in major increases in salinity and temperature within the Red Sea as well as near-complete isolation at times (Lambeck et al., 2011). For several sustained periods during the last two glacial cycles, the minimum channel width connecting the Red Sea to the Arabian Sea was less than 4 km wide and remained narrow whenever the local sea levels were 50 meters below current levels (Lambeck et al., 2011). In contrast, models show that the PAG was nearly completely drained during the peak of the last glaciation until c.a. 14,000 years ago (Lambeck, 1996). A marine incursion into the southern PAG basin started approximately 12,500 years ago, extending towards the northern basin over the following millennia, with the present day PAG shorelines forming just 6,000 years ago (Lambeck, 1996). In contrast, as an open ocean habitat, the Arabian Sea coast has only experienced vertical migration of sea levels during these glacial periods, without geographic isolation.

The combination of extreme environmental conditions, differential changes in habitat and dynamic barriers to gene flow makes the seas bordering the Arabian Peninsula one of the most variable marine environments in the world, with a high potential for speciation driven by both neutral and selective factors (DiBattista et al., 2016b). Although the phylogenetic relationships for varieties of A. marina and congeneric species have been reported for other regions (Duke et al., 1998; Nettel et al., 2008; Li et al., 2016), the extensive gray mangrove populations from the Arabian coasts have rarely been included in reported DNA sequence-based analyses (see Duke et al., 1998; Maguire et al., 2000; Li et al., 2016; Al-Qthanin & Alharbi, 2020). The specific drivers and molecular basis of local adaptation and lineage diversification in A. marina remain understudied both in Arabia and across its global distribution.

Here, we used the Arabian gray mangrove complex to examine how extreme habitat conditions and heterogeneous spatial settings have shaped genetic diversity at the highly variable edge of the species’ range using whole genome and georeferenced environmental data. First, we analyzed patterns of population structure and reconstructed the evolutionary and demographic history of the species in the Arabian Peninsula. Two general competing hypotheses about the evolutionary history of the Arabian mangroves were tested in this study: (i) mangroves from the Red Sea and PAG were extirpated during the glacial cycles of the Pleistocene, followed by a recolonization after the last glacial maximum (LGM); and (ii) mangroves remained within the enclosed seas in glacial refugia during glacial periods, and expanded once sea levels rose. Second, we studied patterns of adaptive variability applying genotype-environment association (GEA) analysis. We used redundancy analysis (RDA) combining environmental and single nucleotide polymorphisms (SNP) data to survey the genome and jointly identify environmental variables and functional genes potentially involved in local adaptation and lineage divergence.

Materials and Methods

Population sampling
We sampled a total of 200 *Avicennia marina* individuals from 19 sites across the Arabian Peninsula (var. *marina*, \(N = 190\)), and one site from Australia (var. *australasica*, \(N = 10\)) (Table 1, Fig. 1; Table S1, Supporting Information). Leaf tissue was collected from trees separated by at least 20 meters, and preserved in silica beads for up to ten days before extraction. Geographic coordinates for each tree were recorded. Genomic DNA was extracted from ground leaf tissue using the DNeasy 96 plant kit (Qiagen, Valencia, CA) according to the manufacturer’s protocol.

**Genome resequencing and variant calling** Illumina paired-end 150 bp libraries with insert size equal to 350 bp were prepared and sequenced in a Novaseq platform. Reads were mapped against a previously published reference genome for *A. marina* (Friis *et al.*, 2020), SNP calling was carried out with the GATK Genome Analysis Toolkit (GATK; McKenna *et al.*, 2010). The resulting dataset consisted of 178 individuals (Table 1; Table S1, Supporting Information) and 15,702,886 biallelic SNPs with a per-individual average coverage of 16.8 and a missing data rate of 0.11. This dataset was further filtered and customized for downstream analyses (See details in Supporting Information).

**Population structure analyses**

To explore genome-wide signs of population structure in Arabian mangroves, we first generated a dataset of high quality, independent and putatively neutral SNPs, consisting of a data matrix of 143,900 variable sites and 170 samples (Table S2, Supporting Information). We conducted a principal components analysis (PCA) as implemented the R package SNPRelate (Zheng, 2012). We also examined patterns of population divergence using a sparse non-negative matrix factorization method (SNMF; Frichot *et al.*, 2014). We ran the program five times per \(K\) value, with \(K\) ranging from 2 to 20. Similarity scores among runs and graphics were computed with CLUMPAK (Kopelman *et al.* 2015). To further explore patterns of geographic variation in Arabian mangroves, we tested for IBD using the neutral SNP dataset. We computed pairwise Nei’s genetic distance values with the hierfstat R package (Goudet *et al.*, 2015). By-sea, pairwise geographic distances were measured based geographic coordinates. A Mantel test was implemented, and significance was computed through 9,999 matrix permutations.

**Phylogenetic analysis**

A maximum likelihood phylogeny was produced using the program IQ-TREE (Nguyen *et al.*, 2015). For this analysis, we applied the same filters than for population structure tests, yet retained the Brisbane population from Australia to be used as outgroup (SNP matrix = 29,433, \(N = 178\); Table S2, Supporting Information). The ascertainment bias correction and the generalized time-reversible substitution model were implemented.

**Population and demographic history analyses**

We performed model comparisons under the likelihood framework developed in fastSIMCOAL2 v2.6 (Excoffier *et al.*, 2013) to estimate demographic parameters and date cladogenetic events among mangroves populations, and to test the competing hypotheses of colonization after the LGM versus isolation in glacial refugia in the enclosed seas around Arabia. Informed by the phylogenetic relationships (See results), three sets of models were independently analyzed for (i) the Red Sea, (ii) the PAG plus the Sea of Oman, and (iii) the entire Arabian Peninsula. Three populations were used as lineage representatives for each set of models. The Brisbane population was once again included to calibrate times of divergence, using a coalescence time with the Arabian lineages of 2.7 million years in all models (Li *et al.*, 2016; He *et al.*, 2019; He *et al.*, 2020). Topologies and times of divergence tested in each set of models were analyzed under scenarios of ‘strict isolation’ and ‘isolation with migration’. Overall, six, twelve and two models were compared for the Red Sea, the PAG and the entire Arabian Peninsula, respectively (Table S4, Fig. S1, Supporting Information). As input data we used the folded site frequency spectra (SFS) generated from resequencing data using easySFS (https://github.com/isaacovercast/easySFS). Details about fastSIMCOAL2 analyses, input data, model sketches and parameter files are provided in the Supporting Information.

TreeMix v1.13 (Pickrell & Pritchard, 2012) was used to further model historical patterns of gene flow between mangrove populations. The corresponding SNP dataset was built applying the same filters as for
the IQ-TREE phylogeny, with the exemption of linkage disequilibrium filters as it can be controlled for in the TreeMix command line (SNP matrix = 797,949; N = 178). We ran TreeMix for 0–15 migrations, grouping SNPs in blocks of 50. Migration edges were plotted until 99.8% of the variance in ancestry between populations was explained by the model (Pickrell & Pritchard, 2012). The consistency of migration edges was evaluated by running TreeMix with 50 total replicates for each added migration edge number using a different, randomly generated seed. Results from the seed that yielded the highest likelihood are reported.

**Candidate gene identification with genotype-environment association analysis**

We used genotype-environment association (GEA) analysis to identify candidate genes evolving under specific environmental pressures, and to estimate their contribution to patterns of local adaptation in mangroves from the Arabian Peninsula. We applied a redundancy analysis approach (Van Den Wollenberg, 1977; Legendre & Legendre, 1998; Borcard et al., 2011) as implemented in the vegan R package (Oksanen et al., 2016). As explanatory variables, we used georeferenced environmental data extracted for the coordinates corresponding to the sampling sites and averaged over populations. The set of explanatory variables included maximum of sea surface salinity and minimum of sea surface temperature averaged over months from the MARSPEC database (Sbrocco & Barber, 2013); as well as isothermality and maximum air temperature of the warmest month from the WorldClim database (Hijmans et al., 2005; Fick & Hijmans, 2017). As response variables, we used population allele frequencies at each variant position from a dataset of 2,488,560 SNPs (N = 170; Table S2, Supporting Information).

Two GEA analyses were implemented: a simple RDA to test for genotype-environment associations between allele frequencies and environmental predictors; and a partial redundancy analysis (pRDA), in which in addition, we controlled for population structure effects. Covariates accounting for population structure consisted of the first two PCs of a PCA based on the allele frequencies matrix after filtering out positions putatively under selection. Following the procedure described in Capblancq et al. (2018), we used the redundancy analyses to identify candidate SNP loci potentially involved in divergent selection, based on Mahalanobis distances estimated between each SNP and the center of the RDA space (Capblancq & Forester, 2021), and applying a p-value threshold < 0.01 with the Bonferroni correction for multiple testing. In addition, to explore differential association patterns of candidate loci at the individual level, we conducted a complementary RDA with the individual genotypes of the SNP outliers as response variables. The first two RDA axes were plotted for visual inspection, as well as pairwise Euclidean distances in the ordination space for all the populations included in the analysis. Further details on implemented GEA analyses and candidate gene identification are provided in the Supporting Information.

**Results**

**Population structure and genetic diversity analyses**

A principal components analysis (PCA) revealed marked levels of population structure. A plot of the first two axes recovered a pattern of clustering that matched the geographic distribution of the sampled populations. Populations from the Red Sea and the PAG clustered apart, showing high overlapping within each biogeographic region. Populations along the coasts of the Arabian Sea also showed marked genetic structure: populations at the west (Salalah and Taqah) grouped together and apart from the remaining genetic clusters, while populations from the Sea of Oman at the east (Shinas and Qurm) clustered close to the PAG. The intermediate population of Filim occupied a central position along the PC2. The third and fourth axes revealed high levels of differentiation within biogeographic regions. Northern and southern basins of the PAG clustered apart along PC3, as well as populations from the Arabian Sea, yet to a lesser extent. Populations of the Red Sea differentiated along PC4 (Fig. 2A).

A SNMF analysis recovered variability patterns consistent with the PCA results. At K = 2, populations clustered into two groups separating the Red Sea from the Arabian Sea and the PAG. The plot for K = 3 also revealed the more isolated, adjacent populations from the west of the Arabian Sea (Salalah and Taqah)
as a differentiated group, with populations along the northeast Oman coast and into the PAG showing decreasing degrees of shared ancestry. At K = 4, populations from the northern basin of the PAG (Bahrain and Dammam) appeared as a differentiated cluster. At K = 5, populations from northern Red Sea (Duba, Al Wajh 1 and Al Wajh 2) also appeared as a genetically differentiated cluster, yet showing a signal of coancestry decreasing with latitude with proximal populations from the central Red Sea (Al Kharrar and King Abdullah Economic City, hereafter KAEC) and to a lesser extent, Al Lith. In the plot for K = 6, populations from the Sea of Oman, Qurum and Shinas, clustered together with Filim, yet Shinas showed intermediate levels of shared ancestry with the genetic cluster of the southern PAG basin, which included Ras Al Khaimah (hereafter RAK), Umm Al Quawain (hereafter UAQ), Ras Ghurab and Suweihat. At K = 7, populations from southern Red Sea (Al Lith, Farasan Banks 1 and Farasan Banks 2; hereafter FB1 and FB2) clustered apart from the central Red Sea populations, yet Al Lith presented intermediate levels of coancestry with the central Red Sea population KAEC (Fig. 2B).

A Mantel test based on selectively neutral, independent SNP loci revealed a significant correlation between by-sea pairwise distances and Nei’s genetic distances among sampled populations (r = 0.765, p-value < 10^{-4}; Table S3, Supporting Information).

Phylogenetic analysis

A maximum likelihood phylogenetic reconstruction showed marked differentiation between and within biogeographic regions, with almost all populations resolved in monophyletic clades with high node support. Arabian mangrove populations grouped in three major, reciprocally monophyletic clades or phylogenetic lineages: the Red Sea lineage (Major Clade I), the southern Arabian Sea lineage (Major Clade II) and the PAG plus the Sea of Oman lineage (Major Clade III). Within the Red Sea’s Major Clade I, the phylogenetic analysis revealed a sequence of cladogenetic events consistent with the geographic distribution of the sampled populations, with the first split separating northern (Duba, Al Wajh 1 and Al Wajh 2) and central populations (Al Kharrar and KAEC) from those of the south (Al Lith, FB1 and FB2); adjacent Al Kharrar and KAEC populations from central Red Sea showed little differentiation and grouped into a single clade. The Arabian Sea’s Major Clade II encompassed the populations Salalah, Taqah and Filim, all occurring in southern Oman, which appeared as the sister group of the Major Clade III, which includes all Sea of Oman and PAG populations. Major Clade III was further divided into two clades separating the populations of the northern basin of the PAG (Bahrain and Dammam) from those of the southern basin (RAK, UAQ, Ras Ghurab and Suweihat) and the Sea of Oman (Qurm and Shinas). Populations of these two latter subregions, which occur on each side of the Strait of Hormuz, appeared as reciprocally monophyletic, sister groups (Fig. 3).

Population and demographic history of the Arabian gray mangrove

Demographic models were compared under fastSIMCOAL2. In the Red Sea, the model with the greatest likelihood score (ΔAIC = 318, Fig. 4; Table S4, Supporting Information) revealed a single cladogenetic event for the three analyzed populations under an isolation with migration scenario, and a splitting time dating back to 99,200 years (95% CI = [40,062 – 295,152]), consistent with a process of rapid, simultaneous lineage differentiation. Gene exchange among lineages was particularly high from Al Kharrar in the central Red Sea towards Farasan Banks 2 towards the southern entrance (MIG Al Kharrar - FB2 = 60, 95% CI = [13 – 109] migrants per generation). For the remaining migration bands, 95% confidence intervals varied by orders of magnitude and included values near to zero, revealing little impact of these parameters in the SFS (Fig. 4; Table S5, Supporting Information). In the case of the PAG plus the Sea of Oman, the model that best fitted our data (ΔAIC = 1,615, Fig. 4; Table S4, Supporting Information) also revealed a scenario of simultaneous cladogenesis in isolation with migration dating back 37,760 years (95% CI = [18,950-161,565]), prior to the last glacial maximum. Signs of some gene flow were detected among PAG lineages. While narrower than in the Red Sea estimates, the ranges also included values near to zero in most cases, reducing the certainty in the model (Fig. 4; Table S5, Supporting Information). In the analysis with representatives of each biogeographic region of the Arabian Peninsula, the highest likelihood corresponded once again to the model including migration (ΔAIC = 1,067, Fig. 4; Table S4, Supporting Information), and revealed
splitting times of 70,180 years (95% CI = [14,074-161,976]) and 153,140 years (95% CI = [90,714-416,863]) (Fig. 4; Table S5, Supporting Information).

Phylogenetic relationships recovered with TreeMix were consistent with the maximum likelihood tree generated with IQ-TREE. A tree model with ten migration events explained up to 99.8% of the SNP variance, revealing signs of both historical and contemporary gene flow among mangrove populations. Two recovered bands of ancestral migration between the outgroup lineage of Brisbane and the northern populations of both the Red Sea and the PAG seemed highly unlikely, so that these results should be interpreted with caution (Fig. 5).

Candidate genes and environmental adaptation patterns in Arabian gray mangroves

Environmental predictors explained 33.2% of the total genetic variance in the simple RDA (Adjusted \( R^2 = 0.332 \), p-value < 0.01). The scores of the simple RDA revealed distinct association patterns among biogeographic regions. Red Sea mangroves showed high, negative loading values on RDA axis (RD) 1, while the remaining populations correlated positively with this axis, and heavily in the case of the Sea of Oman and PAG southern basin. In contrast, the Arabian Sea populations showed high negative association values with RD2, while Red Sea and PAG mangroves showed positive association scores along this axis, yet the former with much lower correlation values. In the pRDA, environmental variables explained 6.4% of the total genetic variance after controlling for population structure (Adjusted \( R^2 = 0.064 \), p-value < 0.01). Scores revealed a more scattered pattern of variation among and within biogeographic regions, where populations from northern PAG showed high positive correlation with pRD1, while populations of the southern PAG and Arabian Sea, as well as from Central Red Sea, showed more limited, negative loading values. In the case of the pRD2, highest loading values corresponded to Qurm, with a positive correlation, and to Ras Ghurab and Suweihat, that correlated negatively (Table S6, Supporting Information).

The simple and the partial RDA were jointly used to test for signs of selection applying the method described in Capblancq et al. (2018). The first to axes of the simple RDA accounted for an unadjusted proportion of explained variance equal to 38.8% (\( R^2 = 0.388 \)) out of 48.0% explained by the full model. In turn, the four axes of the partial RDA accounted for 32.3% (\( R^2 = 0.323 \)). Two and four RDA axes from the simple and partial RDA, respectively, were thereby used for candidate loci identification. The analyses revealed 446 genetic variants showing significant associations with one or more of the four environmental predictors in both the RDA and pRDA (\( N_{\text{RDA}} = 3,015; N_{\text{pRDA}} = 73,671; N_{\text{joint}} = 446 \); Fig. S2, Supporting Information). Of the 446 outliers, 70 were located within 31 functionally annotated genes. Reported functions of the identified genes and gene variants in flowering plants potentially relevant for mangroves in this marginal Arabian environment include the following: tolerance to chronic heat stress (MGE2; Hu et al., 2012); drought resistance by stomatal aperture and density regulation, as well as root development (ABIL2; Li et al., 2015); root hydrotropism and drought tolerance; water transport and osmotic pressure control in response to drought and salt stress (PIP; Katsuhara & Hanba, 2008; Mahdieh et al., 2008; Rodríguez-Gamir et al., 2011); cell wall biosynthesis and development (BXL2; Goujon et al., 2003; Zhao et al., 2010); organ growth adjustments in response to stress conditions (MOB1A; Pinosa et al., 2013); Nitrogen uptake and coordination against biotic stress (WRK50; Cheng et al., 2021); response to salinity stress (SOS2; Shirazi et al., 2019); drought tolerance and regulation of reactive oxygen species through sterol biosynthesis (SQE2; Posé et al., 2009); terpenoid metabolism in response to drought stress (4CLL7; Madritsch et al., 2019; Zhang et al., 2022); flower bud differentiation in response to light (AI5L5; Yi et al., 2021; Liu et al., 2022); and salt sensitivity in association with the ABA signaling pathway (PUM23; Huang et al., 2018) (Table S7, Supporting Information).

To visually explore how adaptive variability is structured at the individual level, we conducted a complementary redundancy analysis based on the individual genotypes of the 446 candidate loci alone. The plot of the first two RDA axes revealed distinctive associations patterns only partially structured by biogeographic region: Populations from the Red Sea differentiated along a gradient loading on maximum air temperature, with northern populations (Dubai, Al Wajh 1 and Al Wajh 2) showing high negative correlation values, central Red Sea (Al Lith, FB1 and FB2) clustering close to the origin of coordinates, and southern Red
Sea showing positive correlations. Populations from the southwest of the Arabian Sea (Taqah and Salalah) showed a strong, positive GEA signal with isothermality, while correlated negatively with a maximum salinity gradient. In contrast, populations from the Sea of Oman presented low association values and clustered around the origin of coordinates. Populations from the PAG showed an opposite GEA pattern to those of the Arabian Sea, also differentiating along the salinity/isothermality gradient, yet associations were weaker and more scattered in the southern basin than in the northern one (Fig. 6A). Individual Euclidean distances averaged over populations in the RDA ordination space showed general similar trends (Fig. 6B), with within-region, highest differences among association patterns occurring along the Red Sea and outrunning divergence between biogeographic regions (Fig. 6B).

Discussion

Marked population structure and phylogenetic data reveal a process of lineage diversification consistent with IBD/geographic isolation in Arabian mangroves

Population structure analyses recovered remarkably high levels of divergence across Arabia. Both a PCA and a SNMF analysis based on putatively neutral SNPs recovered signs of widespread differentiation, and identified groups of genetic clusters that mirrored the geographic distribution of mangroves across the various seas surrounding the Arabian Peninsula. The analyses also showed marked genetic differentiation between the populations from the north, the center and the south of the basin of the Red Sea; as well as between the northern and southern basins of the PAG. Considerable levels of shared ancestry between the populations of the southern PAG basin and the Sea of Oman were also detected, suggesting either incomplete lineage sorting or ongoing gene flow between them. We also found a significant correlation between neutral genetic differentiation and geographic distance among populations, suggesting that IBD may account for a large fraction of the population structure observed in the PCA and SNMF analysis.

Phylogenetic inference revealed a generalized pattern of reciprocal monophyly among Arabian mangrove populations, and identified major, clearly differentiated phylogenetic lineages occurring in the Red Sea, the Arabian Sea and the PAG, with the latter including the populations of the Sea of Oman. Patterns of phylogenetic divergence among Arabian biogeographic regions have been previously reported for other marine taxa (Saenz-Agudelo et al., 2015; Smith et al., 2017; DiBattista et al., 2020; Ketchum et al., 2020; Smith et al., 2022), congruent with the geographic distribution of the species and with physical barriers to gene flow (DiBattista et al., 2016b; DiBattista et al., 2020). However, few marine organisms in Arabia exhibit comparable levels of differentiation at such fine geographic scales as the gray mangrove, suggesting a process of population differentiation occurring within the historically enclosed water bodies of the Red Sea and the PAG. Neutral population structure and phylogenetic relationships here reported support the hypothesis of mangroves finding refuge within the enclosed water bodies of the Arabian Peninsula during Pleistocene glacial cycles, where lineages could diverge in long-term geographic isolation. Geological and paleoclimatic data suggest that areas of the Red Sea basin could indeed have acted as glacial refugia for mangroves (Lambeck et al., 2011), as it has been proposed for other endemic marine lineages of the region (DiBattista et al., 2016a; DiBattista et al., 2016b). This scenario is particularly striking in the case of the PAG, where the presence of potentially suitable habitats was arguably very limited during glacial periods, where only the estuary of the historic Shatt-Ur river outlet to the Sea of Oman near the Strait of Hormuz had been thought to occur (Lambeck, 1996). Interestingly, populations from the southern basin of the PAG appeared as more closely related to the populations of the Sea of Oman, on the outside of the Strait of Hormuz, than with those of the northern PAG basin. This sequence of cladogenetic events also better fits a scenario of differentiation in glacial refugia where populations in the north would have become isolated earlier as the shore lines retreated, as opposed to a scenario of recent colonization throughout the PAG after the LGM, where northern populations would represent the most recent phylogenetic split (e.g. Smith et al., 2022).

Demographic inference supports lineage divergence in (cryptic) glacial refugia followed by gene flow upon
A total of 20 demographic models were compared with fastSIMCOAL2 to identify the most likely scenario under which mangroves from the Red Sea, the PAG and the Arabian Peninsula overall, diversified. For the Red Sea, AIC revealed that the most likely scenario was consistent with a simultaneous cladogenesis occurring during a period spanning the three last glacial cycles, suggesting that mangroves may have colonized the Red Sea basin shortly after the Strait of Bab al-Mandab opened to the Indian Ocean ca. 400,000 years ago (Lambeck et al., 2011) and persisted thereafter. Glacial cycles in the Red Sea are characterized by changes in the eustatic sea level and connectivity, but also by periods of desiccation and sharp changes in salinity and temperature. Therefore, the results reported here suggest that the levels of population and phylogenetic diversity found in the Red Sea are the product of historical events involving changes in aspects of population history such as effective size and geographic isolation, but also in environmental conditions (DiBattista et al., 2016a; DiBattista et al., 2020).

In the case of the PAG, the model that better fit our data revealed a scenario of simultaneous cladogenesis dating to nearly 40K years ago, prior to the LGM and the later infilling of the enclosed sea. Inferred splitting time suggests the presence of cryptic glacial refugia within the PAG. Drastic changes in environmental conditions during glacial maxima, particularly in terms of salinity and temperature may have presented exceptional physiological challenges to resident marine life. Until now, survival through glacial periods has not been documented for any marine taxa in the PAG. However, conditions of intertidal zones in marine swamps, while extreme, may have remained locally suitable for gray mangroves, which occur in a wide range of salinities globally. Lambeck (1996) reconstructions of historic shorelines at times of the LGM show the potential development of lagoons and lakes in several localities, with the main ones corresponding to what are now northern and southern basins, as well as in the Strait of Hormuz. Indeed, today *A. marina* occurs in a number of inland, non-tidal saline (e.g. Stoddart et al., 1973; Ellison & Simmonds, 2003) or even freshwater lakes (Beard, 1967; Taylor, 1986) under conditions that could resemble those of the PAG during glacial periods. Also, previous examples of marine cryptic glacial refugia have been documented for taxa such as seaweeds, rays and snails in places like southwest Ireland, southwest Greenland, the northern Brittany-Hurd Deep area of the English Channel, the Azores islands or the northwest Iberian Peninsula (Bringloe et al.; Provan et al., 2005; Chevolot et al., 2006; Hoarau et al., 2007; DiBattista et al., 2016a), suggesting that the persistence of benign local conditions in marine environments during glacial periods may be relatively common. In addition to a split occurring prior to the LGM, the best fitting demographic model revealed potential gene flow between PAG basins and between the southern basin and the Sea of Oman. The TreeMix analysis also detected significant levels of gene flow between both terminal and ancestral branches, suggesting that some migration could occur as mangrove populations became isolated and also after the LGM, once the infilling of the PAG started. Potential barriers such as the narrow Strait of Hormuz or the marine eddies between Qatar and Iran separating the northern and southern basin may, therefore, not be as limiting for mangrove propagule dispersal as has been proposed for other marine-associated, non-plant species (Hoolihan et al., 2004; Smith et al., 2022).

The highest likelihood model for the entire Arabian Peninsula revealed times of divergence among major lineages congruent with a process of diversification occurring over the different biogeographic regions of the peninsula, arguably driven by geographic isolation and large-scale changes in environmental conditions and connectivity resulting from the glacio-eustatic cycles of the Pleistocene. Coalescence time between the representative populations of the PAG (Ras Ghurab) and the Arabian Sea (Taqah) dated back 70K, lending support for the hypothesis of a lineage split prior to the LGM. The splitting time between the Red Sea (represented by Farasan Banks 2) and the PAG/Arabian Sea ancestral lineage dates back to, at least, the onset of the last glacial period and potentially and at most, to times close the opening of the Red Sea to the Indian Ocean. Both fastSIMCOAL2 and TreeMix analyses detected a certain degree of gene exchange among ancestral and terminal branches, suggesting a lack of reproductive barriers among lineages upon secondary contact.

Altogether, our analyses reveal a process of lineage diversification in the mangroves of the Arabian Peninsula...
spanning the last two to three glacial periods. Although our results suggest that it took place before the LGM, the Arabian mangroves system represents a strikingly fast process of diversification, occurring at temporal scales of tens to a few hundred thousand years. Generation time for the gray mangrove has been estimated at 20 years in tropical areas (He et al., 2019; Triest et al., 2021). Calibrating of our demographic models was based on a splitting time estimate using this value (Li et al., 2016), and thereby we retained it in our analyses. However, generation time in high-latitude populations of the PAG have been reported to range from 25.6 to 53.8 years (Ali et al., 2008), which implies that the diversification process may date back to as few as 2,000 to 6,000 generations or less. High levels of population divergence are expected at the edge of species’ ranges, where reduced population sizes and pronounced habitat fragmentation can lead to rapid differentiation due to drift in isolation and divergent selection (Lesica & Allendorf, 1995; Schaal & Leverich, 1996). Indeed, higher genetic structure in the margins of the species distribution compared to core populations of the gray mangrove of the West Pacific has been previously reported (Arnaud-Haond et al., 2006). Moreover, increased rates of speciation have been documented in other mangrove systems undergoing periodic cycles of isolation and gene flow due to fluctuations in sea level during glacial periods (He et al., 2019), similar to what may have occurred in the enclosed water bodies of the PAG and especially the Red Sea.

Redundancy analyses reveal multi-loci adaptive divergence driven by environmental extremes

In a simple RDA, and a partial RDA controlling for population structure, four constraining variables approximating key environmental parameters for mangrove ecology explained 33.2% and 6.4% of the total variance in population SNP frequencies, respectively. The amount of variance explained by environmental variables was comparable to the values reported in studies applying RDA to genotype-environment analysis in plants, typically ranging from 15% to 40% (e.g. Leamy et al., 2016; Garot et al., 2019; Capblancq et al., 2020; Faske et al., 2021; Chang et al., 2022). We also identified 31 functionally annotated genes associated with one or more RDA outliers. Among the functions associated to detected candidate genes, at least eleven seemed particularly relevant for adaptation to the extreme and variable habitats of the seas around the Arabian Peninsula. These functions included different mechanisms to cope with abiotic stresses such as drought, temperature extremes, salinity or light radiation; hydrotropism and development; or Nitrogen uptake.

A complementary RDA based solely on the identified SNP outliers and computed over individual genotypes revealed considerable adaptive divergence among Red Sea populations, which differentiated along a gradient of maximum temperatures from south to north. The Red Sea does present major latitudinal, climatic differences along its basin. Towards the south, strong solar insolation results in air and sea surface temperatures exceeding the 40o and 30o C respectively, yet minimum temperatures rarely drop under 20o C. In contrast, salinity increases northwards due to limited precipitation and high evaporation, and minimum annual air temperatures regularly drop below 10o C, together with a reduction in nutrient supply and productivity (Anton et al., 2020; Lopez-Sandoval et al., 2021). As a result, several authors consider the northern, central and southern portions of the Red Sea to be distinct biotopes (Sheppard et al., 1992; Carvalho et al., 2019; Friis & Burt, 2020). Differences in gene-environment correlation signals for populations of the PAG were less clear, yet the analysis recovered highly divergent association scores with respect to the populations of the west of the Arabian Sea, differentiating along salinity maxima and isothermality gradients. This is congruent with the environmental conditions found in the PAG, with air temperatures varying through the from 2 degC to 48 degC through the year in specific coastal sites (Boer, 1997; Whitford & Duval, 2019), resulting in remarkably low isothermality values.

Population history and environmental selection drive rapid lineage diversification in the gray mangroves of Arabia

Rapid diversification processes in plants are typically associated with historic large fluctuations in habitat availability and connectivity (Rieseberg & Willis, 2007). In this study, we used the Arabian mangroves complex to analyze the role of spatial environmental variation, geography and neutral evolution in driving population and lineage divergence. Phylogenetic analysis and demographic modeling revealed evolutionary relationships and times of population divergence consistent with periods of geographic isolation in glacial
refugia both in the Red Sea and the PAG, ruling out the hypothesis of a postglacial colonization. A significant signal of IBD was also detected, suggesting that geographic distance may have also played an import role in the differentiation of the phylogenetic lineages. Lineage split times were strikingly recent, providing evidence of a rapid process of diversification taking place within the enclosed Red Sea and PAG water bodies during the Pleistocene. Signs of differential local adaption were conspicuous among phylogenetic lineages identified for each of the biogeographic regions, and also among the populations within the Red Sea, revealing strong, divergent ecological pressures within this biogeographic region. Differential selection among PAG and Sea of Oman populations may have been weaker than in the Red Sea as a result of more limited environmental differences, and have acted during shorter periods, as populations seem to have differentiated more recently. Populations from the less extreme environment of the Arabian Sea also showed distinctive patterns of gene-environment correlation, further supporting a process of adaptive differentiation among the main mangrove lineages of the Arabian Peninsula. Overall, the results reported in this study support a process of rapid diversification driven by the combined factors of environmental selection and historical events, and reveal mangrove peripheral populations of Arabia as potentially important sites for lineage diversification and sources of evolutionary innovation in response to environmental extremes.

Acknowledgements
The authors are grateful to NYUAD’s Core Bioinformatics for analysis and computational assistance. We thank Mark Priest, Dain McParland, Noura Al-Mansoori, Anique Ahmad, and Ada Kovaliukaite for their participation in the fieldwork. We gratefully acknowledge the Environment Agency Abu Dhabi (permit reference number 20181823a), and the Oman Ministry of Environment and Climate Affairs, Director General of Nature Conservation (permit number: 6210/10/75), for providing permits for research related to this manuscript. This work was supported by New York University Abu Dhabi Institute Grant (73 71210 CGSB9) provided by Tamkeen, NYUAD Faculty Research Funds (AD060) and King Abdullah University of Science and Technology baseline funding to CMD.

Author Contributions

Table 1. Arabian mangrove populations included in this study and genotyped individuals per site. The number of analyzed samples after filtering for missing data is shown in brackets.

<table>
<thead>
<tr>
<th>Population</th>
<th>Region</th>
<th>N (analyzed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duba</td>
<td>Red Sea</td>
<td>10 (8)</td>
</tr>
<tr>
<td>Al Wajh 1</td>
<td>10 (9)</td>
<td></td>
</tr>
<tr>
<td>Al Wajh 2</td>
<td>10 (10)</td>
<td></td>
</tr>
<tr>
<td>Al Kharrar</td>
<td>10 (10)</td>
<td></td>
</tr>
<tr>
<td>King Abdullah Economic City (KAEC)</td>
<td>10 (9)</td>
<td></td>
</tr>
<tr>
<td>Al Lith</td>
<td>10 (8)</td>
<td></td>
</tr>
<tr>
<td>Farasan Banks 1 (FB1)</td>
<td>10 (10)</td>
<td></td>
</tr>
<tr>
<td>Farasan Banks 2 (FB2)</td>
<td>10 (9)</td>
<td></td>
</tr>
<tr>
<td>Salalah</td>
<td>Arabian Sea &amp; Sea of Oman</td>
<td>10 (10)</td>
</tr>
<tr>
<td>Taqah</td>
<td>10 (10)</td>
<td></td>
</tr>
<tr>
<td>Filim</td>
<td>10 (10)</td>
<td></td>
</tr>
<tr>
<td>Qurm</td>
<td>10 (9)</td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>Site</td>
<td>Area</td>
</tr>
<tr>
<td>-------------------------</td>
<td>------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Shinas</td>
<td>10</td>
<td>(10)</td>
</tr>
<tr>
<td>Ras Al Khaimah (RAK)</td>
<td>10</td>
<td>Persian/Arabian Gulf</td>
</tr>
<tr>
<td>Umm Al Quawain (UAQ)</td>
<td>10</td>
<td>(5)</td>
</tr>
<tr>
<td>Ras Ghurab</td>
<td>10</td>
<td>(9)</td>
</tr>
<tr>
<td>Suweihat</td>
<td>10</td>
<td>(7)</td>
</tr>
<tr>
<td>Bahrain</td>
<td>10</td>
<td>(10)</td>
</tr>
<tr>
<td>Dammam</td>
<td>10</td>
<td>(10)</td>
</tr>
<tr>
<td>Brisbane</td>
<td>10</td>
<td>(8)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>200</td>
<td>(178)</td>
</tr>
</tbody>
</table>

Figure 1: **Geographic sampling.** Locations of the 19 Arabian Peninsula sampling sites, including eight from the Red Sea, five from the Arabian Sea and the Sea of Oman, and six from the Persian/Arabian Gulf. Dark green shading represents the distribution of mangrove populations along the Arabian coasts, obtained from Spalding *et al.* (2010). Leaf tissue was collected from ten trees per sampling site for DNA extraction and whole genome sequencing. An eastern Australian population was also included as an outgroup (not shown).
Figure 2: Neutral genetic structure of the Arabian mangroves. (A) Plot of axes PC1 against PC2 (left panel) and PC3 against PC4 (right panel) based on selectively neutral genome-wide SNPs. Colors match those in Fig. 1. (B) SNMF analysis based on the same SNP set for K values 2 to 7. On each plot, individuals are represented by vertical bars, and colors correspond to the degree of shared ancestry for each genetic cluster.
Figure 3: **Phylogenetic relationships among Arabian mangrove populations.** Maximum likelihood phylogeny based on selectively neutral genome-wide SNP loci. Nodes corresponding to monophyletic population groups have been collapsed, as well as nodes with bootstrap supports below 85. Branch colors correspond to those in Fig. 1.
Figure 4: **Population and demographic history of the Arabian mangroves.** Graphic summary of the demographic models that better fitted the SFS and corresponding times of divergence computed in fastSIMCOAL2 for the Red Sea, the PAG and the entire Arabian Peninsula. Historical sea level fluctuations through the glacial cycles of the Quaternary are provided in a consistent temporal scale (left panel) based on calibration points reported in Miller *et al.* (2005), represented by gray circles. Glacial periods of low sea connectivity are represented in blue and interglacial periods in gold. Full parameter estimates and CI % are provided in Table S5, Supporting Information.
Figure 5: **Historical gene flow among Arabian mangrove populations.** Phylogenetic relationships and patterns of historic gene flow among Arabian mangrove populations as computed in TreeMix. Branch lengths represent the amount of genetic drift. Arrows represent inferred migration events, colored according to the proportion of the recipient population’s ancestry that derives from the donor. SE represents the standard error of migration rates.
Figure 6: Patterns of genotype-environment association in Arabian mangroves. (A) Redundancy analysis (RDA) based on 446 SNP candidate loci, and four explanatory variables corresponding to environmental factors. Points represent the projection of individual genotype frequencies on the first two RDA axes. The explanatory variables are shown within the space defined by RDA1 and RDA2 by labeled vectors. Vector arrows indicate the direction of the gradient of variation for the corresponding environmental parameter. The loading score of each sample point on each explanatory variable can be obtained by an orthogonal projection on the corresponding plotted vector. For facilitating visual interpretation, marker colors correspond to different subregions within the Red Sea, the Arabian Sea and the Persian/Arabian Gulf. (B) Heatmap of the pairwise Euclidean distances averaged over populations within the RDA1 and RDA2 ordinal space shown in the (A) plot.

References


---

23