How and when should subspecies be defined? Analyses of geographical populations of the mangrove tree, Avicennia marina

Zhengzhen Wang[1], Zixiao Guo[1], Cairong Zhong[2], Haomin Lyu[1], Xinnian Li[1], Norman Duke[3], and SUHUA SHI[4]

[1]Sun Yat-Sen University
[2]Hainan Dongzhai Harbor National Nature Reserve
[3]James Cook University
[4]Sun Yat-sen University

April 28, 2020

Abstract

The designation of subspecies has long been controversial in systematics. In addition to phenotypic divergence, subspecies designation may need to incorporate population genetic analyses. In this study, we perform such a survey on three subspecies of the mangrove tree Avicennia marina, distributed along the Indo-West Pacific coasts. Samples from 16 populations (577 individuals) were collected and 94 nuclear genes were sequenced. We identify four genetic features that support the subspecies designation in this genus. First, genetic divergence that delineates the three subspecies is evident, with discordance found mainly in zones of secondary contact. Moreover, levels of genetic diversity within local populations differ among subspecies. Second, the three subspecies have separate demographic histories inferred by computational modeling. Third, gene flow is detected between subspecies indicating little or no reproductive isolation. Fourth, the delineation of the subspecies varies from locus to locus across the genome, thus hinting continual but uneven exchanges of genes. All these features indicate that the three taxa have proceeded far beyond structured populations. Since they have not satisfied the criteria for full-species designation, the subspecies designation is warranted. We believe these considerations can be generalized to other taxa.

INTRODUCTION

Taxonomic rank below species has been controversial. E. Mayr (1940, 1963) defined subspecies as "a geographically defined aggregate of local populations which differ taxonomically from other subdivisions of the species." Although critiques had challenged this subspecies classification and some taxonomists refuse to describe subspecies (Wilson & Brown, 1953), the value and utility of the subspecies rank was appreciated by others (Durrant, 1955; Mayr, 1982; Phillimore & Owens, 2006). By definition, subspecies is now used to identify populations distinct mainly in three aspects: isolated geographic range or habitat, phylogenetically concordant phenotypic characters, and separate history (O'Brien & Mayr, 1991).

Cases where morphological and genotype-based designations disagree have proven to be particularly challenging (Hawlitschek, Nagy, & Glaw, 2012; Phillimore & Owens, 2006; Torstrom, Pangle, & Swanson, 2014). Morphologically defined taxa are often paraphyletic in phylogenetic analyses (Moritz, 1994; Phillimore & Owens, 2006). Methods incorporating multiple lines of evidence, including morphological, genetic and ecological, have been proposed to resolve this impasse (Patten, 2015).

The current definition of subspecies by Mayr emphasized that speciation mostly occurs in allopatry. However, the conventional BSC view that the genome evolves as a single cohesive unit has been challenged (Wu, 2001; Wu & Ting, 2004; Feder, Egan, & Nosil, 2012; Feder, Flaxman, Egan, Comeault, & Nosil, 2013; Foote, 2018; Jiggins, 2019). An increasing number of cases indicate that speciation occurs with gene flow and without

geographical isolation (Brandvain, Kenney, Flagel, Coop, & Sweigart, 2014; Clarkson et al., 2014; Harr, 2006; Poelstra et al., 2014; Wang, He, Shi, & Wu, 2020).

In this vein, some authors proposed to modify the definition of subspecies to "heritable geographic variation in phenotype." (Patten, 2015; Patten & Remsen, 2017) This implies there is a gene or a set of genes determining phenotypic variation between subspecies. In other words, the test of monophyly on the phylogenetic tree constructed by several molecular markers and several specimens may not be fully reliable.

Here we investigated the genetic architecture of three varietal groups of the mangrove tree *Avicennia marina* to assess when and whether populations within a species have attained a sufficient level of genetic divergence to be recognized as subspecies. Notably, in previous literature, the three groups were referred to as varieties or subspecies by different authors (Duke, 2006; Duke, Benzie, Goodall, & Ballment, 1998; Maguire, Peakall, Saenger, & Maguire, 2002; Maguire, Saenger, Baverstock, & Henry, 2000). Although botanists might have used different terms such as "subspecies", "varieties" or "forms", these assignments are conceptually consistent (Mallet, 2007).

Inhabiting the intertidal zones of tropical and subtropical coasts,*Avicennia marina* is an ideal model for addressing the genetic nature of subspecies because its linear distribution makes it much easier to ascertain the range of its three varietal groups and their contact regions. *A. marina* reaches the most marginal regions of the Indo-West Pacific (IWP), due to its outstanding tolerance to salinity, drought and temperature (Tomlinson, 2016). Three putative subspecies have been identified, namely *A. m. marina, A. m. eucalyptifolia* and *A. m. australasica* . There are reports that these three groups are genetically disjunct but no fixed divergence was found among them (Duke, 1991, 2017; Duke et al., 1998). The distinction between populations and subspecies is of particular significance in the conservation efforts since mangroves are under the threat of climate change, in combination with more direct human disturbances(Gilman, Ellison, Duke, & Field, 2008; Z. Guo et al., 2018).

METHODS

Morphological characters, sampling and DNA extraction

A. m. marina is widely distributed from eastern Africa, through the Middle East, South Asia, Southeast Asia and north to South China. It is also found in western Australia. *A. m. eucalyptifolia* is mainly distributed in northern Australia and extends to southern Philippines, western Indonesia and the Southwestern Pacific islands. There is a significant range overlap of the two groups in western Australia. The third putative subspecies *A. m. australasica* is restricted to south-eastern Australia and northern New Zealand (Figure 1). *A. m. australasica* can be morphologically distinguished from the other two groups by its fully pubescent calyx lobes and bracts (Duke, 2006). These structures are more glabrous in the other groups. The bark of *A. m. australasica* is grey fissured, with short longitudinal fissures or reticulate lines, while barks of the other two subspecies are smooth green or chalky white with flaky patches. *A. m. eucalyptifolia* is mainly distinguished by its lanceolate leaves (as opposed to ovate to elliptic), as well as the style in open flowers which are positioned level with upper edges of anthers (instead of the lower edges of anthers) (Duke, 2006). *A. m. marina* may also be distinguished by its larger flowers and thicker leaves. However, these distinctions in morphological characters may be inconclusive where two putative subspecies coexist (Duke, 2006).

We sampled 16 populations, 577 individuals (16 to 40 individuals per population) from East Africa, South China, Southeast Asia, Australia to New Zealand, covering *A. marina* 's range (Table 1, Figure 1). Leaves of each individual were dried, labeled, and stored for DNA extraction. DNA content of each extraction was measured by NanoDrop 2000. We pooled equal amounts of DNA of individuals from the same population to make one DNA mixture. In total, 16 DNA mixtures were used in the following experiments.

PCR and Illumina high-throughput sequencing

Based on about 200 DNA sequences from a library of expressed sequence tags of *A. marina* (J. Huang et al., 2014), we developed a new set of 94 pairs of primers anchored at exons but spanning at least one intron. All primers produced amplicons with length 500 to 1500 bps. All 94 amplicons from one *A. marina*

individual were sequenced using the Sanger method to obtain reference sequences. We also did this for one *A. alba* individual for use as an outgroup. For the DNA mixture of each population, polymerase chain reaction (PCR) amplification was performed using each of the 94 primer pairs. To reduce amplification errors, TaKaRa high-fidelity PrimerStar HS DNA polymerase was used. The 30 μL PCR mixture consists of 3 μL 10x TaqBuffer (Mg2+), 3 μL dNTPs (2mM/μL), 1.5 μL of each primer (10μM/μL), 0.5 μL HS DNA Polymerase, 3 μL DNA template (˜10ng/μL) and 19 μL deionized water. The PCR program was: 4 min at 94°C; 30 cycles of 10 s at 94°C, 30 s of annealing at the corresponding temperature (Supplementary Table 1), extension at 72°C for 2 min; followed by 8 min final extension at 72°C. Reactions were held at 16°C before PCR products were subjected to electrophoresis on 1.2% agarose gels. Target bands were excised under ultraviolet light and extracted using the Pearl DNA Gel Extraction Kit (Pearl, Guangzhou, China). Extracted DNA was examined by NanoDrop 2000 to ensure that the amount of each gene product was no less than 100ng. PCR products of the 94 loci from the same population were again pooled, using 100 ng of DNA per locus. We thus obtained 16 PCR product mixtures, each including amplicons from 94 loci.

PCR product mixtures from each population were delivered for sequencing on the Illumina HiSeq 2000 platform at BGI (Shenzhen) following the manufacturer's instructions. DNA libraries of 200 bp were constructed for these mixtures and an 8 bp index in the adapter was used to distinguish the populations. Method details used for library construction were the same as those detailed in the Supplementary materials of our previous publication (Zixiao Guo et al., 2016). Raw reads produced were 90 or 130 bps in length.

Read mapping and variant calling

The quality of short reads produced by the HiSeq2000 platform was first examined by FastQC and then mapped to the previously obtained reference sequences using MAQ 0.7.1(H. Li, Ruan, & Durbin, 2008). In mapping and pileup, the mutation rate between reference and read was set to 0.002, the threshold of mismatch base quality sum was 200, and the minimum mapping quality of reads was 30. To exclude false-positive mismatches, we counted the mismatch rate for each site across the read and mismatch rate for each base quality. We trimmed the first and last 10 bases of each read and filtered bases with quality score less than 30, using in-house Perl scripts (available on GitHub: https://github.com/GgamerL/AvicenniaSolexa/tree/SolexaAvicennia).

Variant sites were also identified using MAQ 0.7.1. To avoid bias introduced by sequencing errors, we discarded sites with insufficient site coverage (<100 reads) and those with minor allele frequency less than 0.01 (He et al., 2013). Single nucleotide polymorphisms (SNP) were used in the subsequent analyses. To reduce false SNPs introduced by homopolymers or insertions/deletions, putative variants in those regions were masked.

Genetic divergence and diversity estimation

To estimate absolute genetic divergences between populations of the three subspecies, we computed pairwise D_{XY}following the formula derived by Nei (Nei & Li, 1979). When calculatingD_{XY} , two alleles at each SNP were interpreted as two haplotypes, and corresponding allele frequencies as haplotype frequencies. Each pairwise D_{XY} values were summed over all SNPs and the sum was normalized by sequence length. To test subspecies distinctness, the obtained D_{XY} matrix was used in multidimensional scaling using the 'cmdscale' package implemented in R (Figure 2), as well as neighbor-joining tree constructing using MEGA7 (Kumar, Stecher, & Tamura, 2016). We also performed Principal Component Analysis (PCA) on the SNP frequency matrix (summarizing the frequency of each SNP in each population) using the "prcomp" function in R (Venables & Ripley, 2002) to test whether the populations of three subspecies contain distinct polymorphisms. Finally, to assess the extent to which genetic polymorphisms were fixed,F_{ST} statistics between subspecies and between populations were computed following a method for a large number of SNPs (Nei & Miller, 1990; Willing, Dreyer, & van Oosterhout, 2012).

To test whether the three subspecies maintain different levels of genetic diversity, we computed statistics of π and Watterson's ϑ. The π summarizes the average number of nucleotide differences between two sequences randomly sampled from a population (Nei, 1987), while Watterson's ϑ estimates nucleotide polymorphism based on the number of observed segregating sites (Watterson, 1977). To correct the systematic errors of high-throughput sequencing, we computed the ϑ values following a published algorithm (He et al., 2013).

These parameters, i.e.D_{XY} , F_{ST} , ϑ and π, were calculated using in-house Perl scripts (available from the aforementioned GitHub repository).

Demographic history simulation

To test whether the three putative subspecies represent separately evolving lineages, we built models and simulated sequences under these models using the ms software (Hudson, 2002). Ten models were constructed to reflect demographic history of the three subspecies (Simulation 1): (1) The three subspecies are not separate; (2) *A. m. australasica*and *A. m. eucalyptifolia* are not separate; (3) *A. m. marina* and *A. m. eucalyptifolia* are not separate; (4) *A. m. marina* and *A. m. australasica* are not separate; (5) All three subspecies are separate and *A. m. marina* diverged first; (6) All three subspecies are separate and *A. m. eucalyptifolia* diverged first; (7) All three subspecies are separate and *A. m. australasica* diverged first. (8-10) have the same divergence topology as models 5, 6 and 7 but with gene flow allowed between *A. m. marina* and *A. m. eucalyptifolia* (Figure 3b). Long-term effective population sizes of the populations (N) and coalescent times (T) were common among all models. Notably, to reduce the complexity of parameter setting and speed up computation, all population size parameters were derived from a single parameter N_0 randomly picked from the prior distribution. In models with more than one lineage, N_0 was assigned to one lineage and the N for other specific lineage was produced by multiplying N_0 by ϑ_x/ϑ_0, where ϑ_x and ϑ_0 are the observed ϑ of the specific and the assigned lineage respectively.

For each model, we performed 100,000 coalescent simulations using the ms program (Hudson, 2002). Each simulation contained 80 loci of 1000 base pairs. Mutation rate was set at $3.26e^{-8}$/generation/bp, which was estimated by phylogenomic comparisons with closely related species on whole genomes (He et al., 2020). The sample size of each population was consistent with our field sampling described previously (Table 1). Demographic parameters were drawn randomly from a uniform prior distribution. Identical prior distributions of corresponding parameters were set for models within each set (Supplementary Table 2).

Ten summary statistics were calculated for each simulated data set, including segregating site number (S), Watterson's estimator (ϑ), nucleotide polymorphism (π) and Tajima's D of each population, as well as D_{XY} and F_{ST} . Summary statistics were calculated for each simulation independently. Euclidean distances were calculated by comparing simulated statistics with corresponding observed summary statistics. The tolerance of retaining simulated data was set to 0.05. Bayesian posterior probabilities of each model were then estimated following the Approximate Bayesian Computation (ABC) schema (Beaumont, Zhang, & Balding, 2002) using the "abc" package in R (Csilléry, François, & Blum, 2012). The "postpr" function together with "neuralnet" option in the "abc" R package was used to perform model selection.

We also built four models to test whether maBB genetically belongs to*A. m. marina* or *A. m. eucalyptifolia*(Simulation 2). In model v1 and v2, maBB (constant effective population size of N_{bb}) and *A. m. marina*(N_{ma}) coalesced at vT_1 generations ago, and then the common ancestor further coalesced with *A. m. eucalyptifolia* (effective population size N_{eu}) at vT_0 generations ago (vT_0>vT_1). Model v1 differed from v2 by presence or absence of gene flow ($m_1$and m_2) between maBB and *A. m.eucalyptifolia* . Similarly, in models v3 and v4, maBB (N_{bb}) coalesced with *A. m. eucalyptifolia*(N_{eu}) at vT_1 generations ago. The common ancestor then coalesced with *A. m. marina* (effective population size N_{ma}) at vT_0 generations ago (vT_0>vT_1). Nine summary statistics, namely Watterson's estimator (ϑ) for each population and pairwise F_{ST} and $D_{XY,}$ were used in the model selection procedure similar to the one previously described.

Detection of gene flow between subspecies

We used the statistical model implemented in TreeMix to infer patterns of population splits and mixtures among subspecies (Pickrell & Pritchard, 2012). As revealed from the F_{ST}statistic above, some populations are genetically similar, indicating panmixia within a relative wide geographic range, e.g. Andaman Sea on the west of Malay Peninsula, South China Sea (Gulf of Thailand and Hainan Island). Hence, one representing population from each region was used in this analysis. The eleven populations were related to the common ancestor through a graph of ancestral populations, which was inferred by allele frequency and a Gaussian approximation to genetic drift (Pickrell & Pritchard, 2012). Gene flow events were inferred by adding

admixtures onto the Maximum Likelihood population splitting topology.

Haplotype inference and population structure mapping

Haplotypes of genes were inferred following an expectation-maximization algorithm (Bilmes, 1998; Dempster, Laird, & Rubin, 1977). We used an in-house Perl script to perform this haplotype inference, employing short reads to extract SNP linkage information (available from the above GitHub repository). If two adjacent SNPs were not covered by any read pair, we broke the gene into segments. In this case, the midpoint of the two adjacent SNPs would be defined as the breakpoint of two consecutive segments. Because the inference process uses a maximum likelihood method to compare haplotype alternatives, it is prone to yield short segments when a large number of populations is considered. Therefore, we selected eight populations representing different varieties and different regions for inferring haplotypes: two *A. m. eucalyptifolia* (euCA and euDW), two *A. m. australasica* (auAK and auBS), and four *A. m. marina* (maBB, maLS, maTN, and maSY). Finally, genes were split into 454 linked segments and haplotypes were inferred for each segment (Supplementary Table 3). Before constructing haplotype networks, we filtered segments with length less than 100bp or with missing data. For each of the 231 retained segments, we computed a haplotype network using the NETWORK software (Polzin & Daneshmand, 2003).

RESULTS

Among-group genetic divergence

We obtained 76 to 87 kb of DNA sequence covering 88 to 94 genes from each population (Table 1). By mapping short reads to reference sequences, we identified 74 to 1657 segregating sites (Table 1). We calculated among-population pairwise D_{XY} values to assess genetic divergence and used the resulting distance matrix to construct a neighbor-joining tree (Supplementary Figure 1). TheD_{XY} matrix showed clear divergence between the three varietal groups previously designated as subspecies (Figure 2b), the maBB is an outlier discussed below. The largestD_{XY} values were observed between the *A. m. australasica* populations and the other two subspecies, ranging from 7.7 to 9.9/kb (Supplementary Table 4). Relatively lower divergence was observed between *A. m. eucalyptifolia* and *A. m.marina* populations, with D_{XY} values between 6.5 and 7.4/kb. By pooling populations within putative subspecies, we estimated the D_{XY} to be 8.2/kb between *A. m. eucalyptifolia* and *A. m. australasica* , 6.7/kb between *A. m. marina* and *A. m. eucalyptifolia* and 9.1/kb between *A. m. marina* and *A. m. australasica* .

Genetic divergences were generally lower among populations than among subspecies. The two *A. m. australasica* populations diverged little from each other (D_{XY} =2.2/kb), while the*A. m. eucalyptifolia* pair diverged more but still less than differentiation among subspecies (D_{XY} = 5.48/kb). Within *A. m. marina* , we see two major groups, one containing maMC, maLS, and maPN (west of the Malay Peninsula), the other maTN, maBK, maSS, maSY, maWC, maSB, maCB, and maBL (east of the Malay Peninsula, Supplementary Figure 2). D_{XY} per kb ranges from 1.27 to 3.75 within the first and from 0.94 to 4.69 within the second group. Between the two groups, D_{XY}ranges from 4.32 to 5.69, still lower than between subspecies. The maBB population is an outlier and has diverged far from other *A. m. marina* populations (D_{XY} = 7.76-8.43/kb), to a level comparable with among subspecies differentiation.

Genetic divergences can also be observed by comparing the polymorphism frequencies within populations. By plotting principal components of the allele frequency matrix, we again see clear differences among the three subspecies. The two *A. m. eucalyptifolia* populations also show obviously different polymorphism frequency patterns (Figure 2a).D_{XY} quantifies the absolute divergence between groups, while F_{ST} reflects how much genetic divergence is fixed in each subspecies (Cruickshank & Hahn, 2014). The 120 values of pairwise F_{ST} calculated for the 16 populations are generally high, with the average value of 0.61 (first and third quartiles are 0.50 and 0.76 respectively).F_{ST} values among five *A. m. marina*populations from the South China Sea, i.e. maTN, maBK, maSS, maSY and maWC cluster at the bottom of the distribution. If we exclude the ten values or pairs of these five populations, the average of remaining 110 pairs is elevated to 0.66 (first and third quartiles 0.54 and 0.77). Notably, the F_{ST} values among populations from*A. m eucalyptifolia* and *A. m. australasica* are 0.19 and 0.42 respectively. The prevalence of such highF_{ST} value

indicates that genetic polymorphisms are fixed to a large extent between subspecies as well as between some populations within *A. m. marina* .

Both the nucleotide diversity (π) and Watterson's estimator of nucleotide polymorphism (ϑ) identified different levels of within-population genetic variation. The two *A. m.eucalyptifolia* populations have the highest genetic diversity, with average ϑ (across segments) = 2.82 and 3.94/kb and average π = 3.41 and 4.06/kb (Figure 3a). In contrast, *A. m. marina* populations are low in genetic diversity, with average ϑ ranging from 0.21 to 0.91/kb and average π ranging from 0.15 to 1.39/kb (Table1, Figure 3a). The two *A. m. australasica* populations have very different levels of genetic diversity (Table1, Figure 3a). The very low level of diversity in the auAK population is likely due to its marginal location. Marginal populations of *A. m. marina* such as maWC and maSY are also very monomorphic. The auBS population of *A. m. australasica*has an intermediate level of genetic diversity, higher than *A. m. marina* but lower than *A. m. eucalyptifolia* .

The three subspecies have separate demographic history

The highly variable levels of genetic diversity in populations from different subspecies hint that the three putative subspecies have separate demographic histories. To test whether they indeed have evolved separately, we built ten demographic models and used the ABC method to select the models that best fit our observed data (Figure 3b). These simulations indicate that model 10 provides the best fit. This model treats the three subspecies as separate lineages, with gene flow between*A. m. marina* and *A. m. eucalyptifolia* . The posterior probability of this model is 0.3774, much higher than the second-best model 2 (Table 2). Notably, the gene flow between *A. m. marina*and *A. m. eucalyptifolia* allowed in model 10 indicates that these subspecies are not completely isolated, though separate.

Gene flow between subspecies

We wanted to examine the magnitude of among-subspecies gene flow in more detail. The samples collected on the west coast of Australia (population maBB) were morphologically diagnosed as *A. m. marina.*However, our genetic data show relatively lowF_{ST} values between maBB and *A. m. eucalyptifolia* . F_{ST} values between maBB and other *A. m. marina* populations were comparable with those between maBB and *A. m. australasica* populations (Supplementary Table 4). The NJ tree also clusters maBB with the two *A. m. eucalyptifolia*populations (Supplementary Figure 1). This contrast between morphological and genetic lines of evidences hints strongly at gene flow between *A. m. marina* and *A. m. eucalyptifolia* in Western Australia where the two subspecies grow together. The demographic model assuming that maBB has descended from *A. m. marina* but exchanges genes with *A. m. eucalyptifolia* is the most likely in our ABC analyses (model v2, posterior probability 0.933, Table 2 and Figure 4a).

We also used TreeMix to capture more potential gene flow events among populations (Figure 4b). We identified six such events on the population splitting graph. Among the six events, three were between subspecies, with one gene flow event between each pair of the three subspecies. Two gene flow events occurred between pairs of *A. m. marina*populations. The last gene flow event occurred between maBB and the outgroup species *A. alba* .

Haplotype network variation across the genome

Despite clear genomic and morphological differences among *A. marina* subspecies, we found evidence of gene flow. Thus, the genomes of these groups are expected to be mosaic, with regions of high divergence interspersed among undifferentiated loci subject to exchange of genetic material among subspecies. To verify this, we inferred haplotype networks across the close to 100 loci we sequenced on the genome. Using an expectation-maximization method to infer the linkage of the SNPs detected from these loci, we split these regions into 454 linked segments (Supplementary Table 3). Segments with missing data and with length less than 100bp were discarded and 231 segments were retained for haplotype network reconstruction, with *A. alba* as the outgroup (Figure 5).

Among these segments, 134 (58.0%) were not genetically distinguishable among groups with only one or a few haplotypes identified and all haplotypes closely related to each other and shared among the three

subspecies. The other 66 segments (28.6%) reliably distinguished *A. m. australasica* from the other two subspecies. Among these 66 segments, the maBB population shared haplotypes with *A. m. australasica* rather than *A. m. marina* in seven loci. The third type of segments, which was 14 in total (6.1%), delimited *A. m. marina* from the other two subspecies. Five segments (2.2%) distinguish *A. m. eucalyptifolia* , but maBB shared haplotypes with *A. m. eucalyptifolia* in all cases. Most importantly, 11 segments (4.8%) allow clear delimitation of all three subspecies, with haplotypes split into three groups and each subspecies containing haplotypes from a single group. However, in eight of the 11 segments, maBB shared haplotypes with *A. m. eucalyptifolia* . Finally, one segment (0.4%) separated *A. m. marina* and *A. m. australasica,* but *A. m. eucalyptifolia* contained haplotypes from both subspecies.

DISCUSSION

The three subspecies are distinct in morphology, genetic divergence and demography

In this study we comprehensively sampled *A. marina* populations across their geographical range, assembled an extensive SNP data set and used it to test the genetic basis for distinguishing three morphologically recognized subspecies. Our study indeed confirms a robust genetic split of *A. marina* into three varietal groups, noting that this divergence was observed both in the genetic distance D_{XY} matrix and in PCA clustering based on a SNP frequency matrix. Given the almost mutually exclusive distribution of these three varieties (Figure 1), the genetic divergence between subspecies is expected to be shaped to a large extent by geographical barriers.

The boundary of current distributions of *A. m.eucalyptifolia* and *A. m. marina* in Asia roughly aligns with Weber's Line. Similar to the cases identified in other mangrove trees, such as *Rhizophora apiculata* (Zixiao Guo et al., 2016),*Sonneratia alba* (Yang et al., 2017), *Ceriops tagal* (Y. Huang et al., 2012) and *Xylocarpus granatum* (Zixiao Guo et al., 2018), we presumes that the Indonesia-through flow in the Wallacea Zone was the geographic barrier isolating *A. m. marina* in Asia and *A. m. eucalyptifolia* in Australia and New Guinea (Gordon, 2005; Hall, 2009). On the east coast of Australia, the occurrence of *A. m. eucalyptifolia* grades into *A. m. australasica* between Rockhampton and Brisbane, where the North Caledonian Jet bifurcates into the North Queensland Current and the East Australian Current (Ganachaud et al., 2007; Schiller et al., 2008). Hence, we also infer that this bifurcating ocean current may underlie the divergence of *A. m. eucalyptifolia* and *A. m. australasica* . Although geographical isolation is no longer considered as the only factor driving speciation, ocean currents seem to have played an important role in the driving divergence of these three subspecies.

Sister subspecies are postulated to be separate evolutionary lineages, although some gene flow can be accommodated if they come into contact. Genetic structure between populations can result from reduction in gene flow, but evidence from multiple aspects is necessary to establish that such events occur. The very different levels of genetic diversity among subspecies may provide such evidence. We find that the model that treats the three subspecies as separate lineages, with gene flow between *A. m. marina* and *A. m. eucalyptifolia,* fits our data best. A previous study estimated that *A. m. australasica* diverged from the other two subspecies at about 2.7 MYA, while *A. m.eucalyptifolia* diverged from *A. m. marina* at about 1.8 MYA (X. Li et al., 2016). The split between *A. m.marina* and *A. m. eucalyptifolia* coincides with the beginning of Pleistocene, when sea level significantly dropped. The sea level drop emerged shallow seas into lands in the Indo-Australian Archipelago region, possibly leading to separation of nascent subspecies ranges.

Gene flow and uneven divergence across the genome

The establishment of reproductive isolation is an important landmark of speciation completion (Feder, Egan, & Nosil, 2012; Wu, 2001; Wu & Ting, 2004). Interbreeding between complete species is impeded by various forms of behavioral, ecological or genetical incompatibilities (Seehausen et al., 2014). Subspecies, on the other hand, are in the middle to late stages of speciation, before the establishment of strong reproductive isolation. This permits gene flow between subspecies wherever hard geographic isolation is absent. In our case, *A. m. marina* and *A. m. eucalyptifolia* co-occur on Western Australian coasts. A recent study identified seven genetic clusters of populations along the 2,400 kilometers of these coasts. Although the

authors claimed equal possibility to sample individuals from both subspecies, their clustering analyses didn't reveal subspecies differentiation (Binks et al., 2019). Our data provide direct evidence for the gene flow between subspecies in that location, consistent with observations from the previous study. Moreover, gene flow is not limited to Western Australia, but is common to all three subspecies.

We find that *A. marina* subspecies have achieved stable morphological and genetic distinctness but not strong reproductive isolation. At this stage, the genome has not yet diverged as a collective unit. Thus, using a few markers randomly selected from the genome is not likely to be sufficient to stably distinguish these varietal groups. Instead, some regions, usually accounting for a small portion of the genome (for example, about 5% of the genome in this study), diverge significantly but the rest of the genome is mixed by gene flow. The highly divergent part of the genome is not necessarily a single block on one chromosome, but is usually sparsely dispersed across the genome. Even though specific genes underlying the divergent phenotypes are not mapped, this genomic pattern of uneven divergence is typical for the subspecies stage.

In conclusion, our deep survey of genetic variation among populations of*A. marina* from a wide geographic range and large genomic scope, confirms the designation of three subspecies within *A. marina* . Subspecies should show features including: (1) distinct morphological traits and higher level of genetic divergence between pairs of populations from different than from the same subspecies; (2) separate demographic history; (3) complete geographical isolation is not mandatory for subspecies and gene flow between them is possible where ranges overlap; (4) Delineation of subspecies varies from locus to locus, suggesting inconsistent genetic divergence across the genome, and hinting continual (and uneven) exchange of genes from locus to locus. These features from a population genetic prospective could be applied to assessment of subspecies in other domain of the tree of life.

The utility of subspecies classification in evolutionary studies and conservation

As a rank between population and species, the subspecies is useful for predicting the evolutionary divergence levels among geographical populations (Barrowclough, 1982). Populations defined as subspecies are expected to be more highly differentiated than within groups and should have separate demographic history. The classification of subspecies should not be the end but a byproduct of investigations of genetic variation within a species if patterns that warrant designation of subspecies are found (Barrowclough, 1982).

The other important utility of subspecies is to inform conservation decisions. Mayr proposed that subspecies are of conservation importance for their potential to evolve into full species and their acquisition of unique characteristics (O'Brien & Mayr, 1991). The emphasis on species diversity in conservation policy had driven taxonomists to revise subspecies upward to species (Mallet, 2007). More recently, managers have become increasingly aware of the necessity to protect biodiversity at all levels of life. Hence the recognition of subspecies based on well-founded evidences will be important for conserving genetic diversity below the species rank. As a mostly widespread mangrove tree,*A. marina* is important for the ecological health of coastal ecosystems, especially under the prospect of global climate change. Management efforts to protect each of the three subspecies should refer to their distinct genetic backgrounds.

ACKNOWLEDGEMENTS

We thank Chung-I Wu for insightful comments that improved the manuscript. This study was supported by the National Natural Science Foundation of China (grant numbers 31830005 and 31900200); the National Key Research and Development Plan (2017FY100705); the China Postdoctoral Science Foundation (grant number 2019M663212); and the Chang Hungta Science Foundation of Sun Yat-Sen University.

REFERENCES

Barrowclough, G. F. (1982). Geographic Variation , Predictiveness , and Subspecies. *The Auk* , *99* (3), 601–603.

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics* ,*162* (4), 2025–2035.

Bilmes, J. A. (1998). A gentle tutorial of the EM algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. *ReCALL* , *4* (510), 126. doi: 10.1080/0042098032000136147

Binks, R. M., Byrne, M., Mcmahon, K., Pitt, G., Murray, K., & Evans, R. D. (2019). Habitat discontinuities form strong barriers to gene flow among mangrove populations, despite the capacity for long-distance dispersal. *Diversity and Distributions* , *25* (2), 298–309. doi: 10.1111/ddi.12851

Brandvain, Y., Kenney, A. M., Flagel, L., Coop, G., & Sweigart, A. L. (2014). Speciation and Introgression between *Mimulus nasutus* and*Mimulus guttatus* . *Plos Genetics* , *10* (6), e1004410. doi: 10.1371/journal.pgen.1004410

Clarkson, C. S., Weetman, D., Essandoh, J., Yawson, A. E., Maslen, G., Manske, M., ... Donnelly, M. J. (2014). Adaptive introgression between Anopheles sibling species eliminates a major genomic island but not reproductive isolation. *Nature Communications* , *5* (May). doi: 10.1038/ncomms5248

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* , *23* (13), 3133–3157. doi: 10.1111/mec.12796

Csilléry, K., François, O., & Blum, M. G. B. (2012). Abc: An R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution* , *3* (3), 475–479. doi: 10.1111/j.2041-210X.2011.00179.x

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via the *EM* Algorithm.*Journal of the Royal Statistical Society: Series B (Methodological)* , *39* (1), 1–22. doi: 10.1111/j.2517-6161.1977.tb01600.x

Duke, N. C. (1991). A systematic revision of the mangrove genus*Avicennia* (Avicenniaceae) in Australasia. *Australian Systematic Botany* , *4* (2), 299. doi: 10.1071/SB9910299

Duke, N. C. (2006). *Australia's mangroves: the authoritative guide to Australia's mangrove plants* . MER.

Duke, N. C. (2017). Mangrove Floristics and Biogeography Revisited : Further Deductions from Biodiversity Hot Spots , Ancestral Discontinuities , and Common Evolutionary Processes. In *Mangrove Ecosystems: A Global Biogeographic Perspective* (pp. 17–53).

Duke, N. C., Benzie, J. A. H., Goodall, J. A., & Ballment, E. R. (1998). Genetic Structure and Evolution of Species in the Mangrove Genus*Avicennia* (Avicenniaceae) in the Indo-West Pacific.*Evolution* , *52* (6), 1612–1626.

Durrant, S. D. (1955). In Defense of the Subspecies. *Systematic Zoology* , *4* (4), 186–190.

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics* , *28* (7), 342–350. doi: 10.1016/J.TIG.2012.03.009

Ganachaud, A., Kessler, W., Wijffels, S., Ridgway, K., Cai, W., Holbrook, N., ... Aung, T. (2007). *Southwest Pacific Ocean Circulation and Climate Experiment (SPICE)* . Seattle, WA.

Gilman, E. L., Ellison, J., Duke, N. C., & Field, C. (2008). Threats to mangroves from climate change and adaptation options: A review.*Aquatic Botany* , *89* (2), 237–250. doi: 10.1016/j.aquabot.2007.12.009

Gordon, A. L. (2005). Oceanography of the Indonesian seas and their throughflow. *Oceanography* , *18* (4), 14–27.

Guo, Z., Li, X., He, Z., Yang, Y., Wang, W., Zhong, C., ... Shi, S. (2018). Extremely low genetic diversity across mangrove taxa reflects past sea level changes and hints at poor future responses. *Global Change Biology* , *24* (4). doi: 10.1111/gcb.13968

Guo, Zixiao, Guo, W., Wu, H., Fang, X., Ng, W. L., Shi, X., ... Huang, Y. (2018). Differing phylogeographic patterns within the Indo-West Pacific mangrove genus *Xylocarpus* (Meliaceae).*Journal of Biogeography* , *45* (3), 676–689. doi: 10.1111/jbi.13151

Guo, Zixiao, Huang, Y., Chen, Y., Duke, N. C., Zhong, C., & Shi, S. (2016). Genetic discontinuities in a dominant mangrove *Rhizophora apiculata* (Rhizophoraceae) in the Indo-Malesian region. *Journal of Biogeography* , *43* , 1856–1868. doi: 10.1111/jbi.12770

Hall, R. (2009). Southeast Asia's changing palaeogeography. *Blumea - Biodiversity, Evolution and Biogeography of Plants* , *54* (1), 148–161. doi: 10.3767/000651909X475941

Harr, B. (2006). Genomic islands of differentiation between house mouse subspecies. *Genome Research* , *16* (6), 730–737. doi: 10.1101/gr.5045006

Hawlitschek, O., Nagy, Z. T., & Glaw, F. (2012). Island evolution and systematic revision of comoran snakes: Why and when subspecies still make sense. *PLoS ONE* , *7* (8). doi: 10.1371/journal.pone.0042970

He, Z., Li, X., Ling, S., Fu, Y.-X., Hungate, E., Shi, S., & Wu, C.-I. (2013). Estimating DNA polymorphism from next generation sequencing data with high error rate by dual sequencing applications. *BMC Genomics* , *14* (1), 535. doi: 10.1186/1471-2164-14-535

He, Z., Xu, S., Zhang, Z., Guo, W., Lyu, H., Zhong, C., ... Shi, S. (2020). Convergent adaptation of the genomes of woody plants at the land-sea interface. *National Science Review* .

Huang, J., Lu, X., Zhang, W., Huang, R., Chen, S., & Zheng, Y. (2014). Transcriptome Sequencing and Analysis of Leaf Tissue of *Avicennia marina* Using the Illumina Platform. *PLoS ONE* , *9* (9), e108785. doi: 10.1371/journal.pone.0108785

Huang, Y., Zhu, C., Li, X., Li, X., Hu, L., Tan, F., ... Shi, S. (2012). Differentiated population structure of a genetically depauperate mangrove species Ceriops tagal revealed by both Sanger and deep sequencing. *Aquatic Botany* , *101* , 46–54. doi: 10.1016/j.aquabot.2012.04.001

Hudson, R. R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* , *18* (2), 337–338. doi: 10.1093/bioinformatics/18.2.337

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets.*Molecular Biology and Evolution* , *33* (7), msw054. doi: 10.1093/molbev/msw054

Li, H., Ruan, J., & Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research* , 1851–1858. doi: 10.1101/gr.078212.108.

Li, X., Duke, N. C., Yang, Y., Huang, L., Zhu, Y., Zhang, Z., ... Shi, S. (2016). Re-evaluation of phylogenetic relationships among species of the mangrove genus *Avicennia* from Indo-West Pacific based on multilocus analyses. *PLoS ONE* , *11* (10), 1–14. doi: 10.1371/journal.pone.0164453

Maguire, T., Peakall, R., Saenger, P., & Maguire, L. (2002). Comparative analysis of genetic diversity in the mangrove species*Avicennia marina* (Forsk.) Vierh.(Avicenniaceae) detected by AFLPs and SSRs. *TAG Theoretical and Applied Genetics* , *104* (2), 388–398.

Maguire, T., Saenger, P., Baverstock, P., & Henry, R. (2000). Microsatellite analysis of genetic structure in the mangrove species*Avicennia marina* (Forsk.) Vierh.(Avicenniaceae). *Molecular Ecology* , *9* (11), 1853–1862.

Mallet, J. (2007). Subspecies, semispecies, superspecies.*Encyclopedia of Biodiversity* .

Mayr, E. (1940). Speciation phenomena in birds. *The American Naturalist* , *74* , 249–278.

Mayr, E. (1963). *Animal Species and Evolution* . Cambridge, MA: Harvard University Press.

Mayr, E. (1982). Commentary Forum : Avian Subspecies in the 1980 ' S of What Use Are Subspecies ? *The Auk* , *99* (3), 593–595.

Moritz, C. (1994). Defining 'Evolutionarily Significant Units' for conservation. *Tree* , *9* (10), 373–375. doi: 10.1016/0169-5347(94)90057-4

Nei, M. (1987). *Molecular evolutionary genetics* . New York: Columbia University Press.

Nei, M., & Li, W.-H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America* ,*76* (10), 5269–5273. doi: 10.1073/pnas.76.10.5269

Nei, M., & Miller, J. C. (1990). A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. *Genetics* , *125* (4), 873–879.

O'Brien, S. J., & Mayr, E. (1991). Bureaucratic mischief: Recognizing endangered species and subspecies. *Science* , *251* (4998), 1187–1188. doi: 10.1126/science.251.4998.1187

Patten, M. A. (2015). Subspecies and the philosophy of science. *The Auk* , *132* (2), 481–485. doi: 10.1642/auk-15-1.1

Patten, M. A., & Remsen, J. V. (2017). Complementary roles of phenotype and genotype in subspecies delimitation. *Journal of Heredity* ,*108* (4), 462–464. doi: 10.1093/jhered/esx013

Phillimore, A. B., & Owens, I. P. F. (2006). Are subspecies useful in evolutionary and conservation biology? *Proceedings of the Royal Society B: Biological Sciences* , *273* (1590), 1049–1053. doi: 10.1098/rspb.2005.3425

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics* , *8* (11). doi: 10.1371/journal.pgen.1002967

Poelstra, J. W., Vijay, N., Bossu, C. M., Lantz, H., Ryll, B., Baglione, V., ... Wolf, J. B. W. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* ,*344* (6190), 1410–1414.

Polzin, T., & Daneshmand, S. V. (2003). On Steiner trees and minimum spanning trees in hypergraphs. *Operations Research Letters* ,*31* (1), 12–20. doi: 10.1016/S0167-6377(02)00185-2

Schiller, A., Oke, P. R., Brassington, G., Entel, M., Fiedler, R., Griffin, D. A., & Mansbridge, J. V. (2008). Eddy-resolving ocean circulation in the Asian–Australian region inferred from an ocean reanalysis effort. *Progress in Oceanography* , *76* (3), 334–365. doi: 10.1016/j.pocean.2008.01.003

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., ... Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics* , *15* (3), 176–192. doi: 10.1038/nrg3644

Tomlinson, P. B. (2016). *The Botany of Mangrovess* (Second Edi). Cambridge, UK: Cambridge University Press.

Torstrom, S. M., Pangle, K. L., & Swanson, B. J. (2014). Shedding subspecies: The influence of genetics on reptile subspecies taxonomy. *Molecular Phylogenetics and Evolution* , *76* (1), 134–143. doi: 10.1016/j.ympev.2014.03.011

Venables, W. N., & Ripley, B. D. (2002). Modern applied statistics with S Springer-Verlag. *New York* .

Wang, X., He, Z., Shi, S., & Wu, C.-I. (2020). Genes and speciation – Is it time to abandon the Biological Species Concept? *National Science Review* . doi: 10.1093/nsr/nwz220

Watterson, G. A. (1977). Heterosis or Neutrality? *Genetics* ,*85* (4), 789–814.

Willing, E. M., Dreyer, C., & van Oosterhout, C. (2012). Estimates of genetic differentiation measured by fst do not necessarily require large sample sizes when using many snp markers. *PLoS ONE* , *7* (8), 1–7. doi: 10.1371/journal.pone.0042649

Wilson, E. O., & Brown, W. L. (1953). The subspecies concept and its taxonomic application. *Systematic Zoology* , *2* (3), 97–111. doi: 10.2307/2411818

Wu, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology* , *14* (September), 851–865.

Wu, C.-I., & Ting, C. T. (2004). Genes and speciation. *Nature Reviews Genetics* , *5* (2), 114–122. doi: 10.1038/nrg1269

Yang, Y., Li, J., Yang, S., Li, X., Fang, L., Zhong, C., ... Shi, S. (2017). Effects of Pleistocene sea-level fluctuations on mangrove population dynamics: a lesson from Sonneratia alba. *BMC Evolutionary Biology* , *17* (1), 1–14. doi: 10.1186/s12862-016-0849-z

DATA ACCESSIBILITY

GenBank accession numbers of reference sequences for the genes we sequenced are KC928137-KC928228, KC954697 and KF918414-KF918415 (the detailed information could be found in the supplementary Table S2 of He et al (2019), doi: 10.1093/nsr/nwy078).

AUTHOR CONTRIBUTIONS

S. Shi and Z. Guo designed and supervised the project. S. Shi, C. Zhong, X. Li, H. Lyu and N. C. Duke collected the samples. Z. Wang and H. Lyu produced the data. Z. Wang and Z. Guo analyzed the data. Z. Guo and Z. Wang wrote the manuscript. S. Shi and N. C. Duke helped in improving the manuscript. All the authors read and approved the final manuscript.

Table 1 Sample information and population genetic statistics.

	Location	Longitude & Latitude	Site ID	N[1]	G[2]	Total reads	Depth	Total length	S[3]
1	Meed Creek, Kenya	39°58'6"E, 3deg20'33"S	maMC	16	92	6870508	4670	83438	97
2	Laemson, Thailand	98°27'57"E, 9deg36'14"N	maLS	35	91	10373578	5966	85999	322
3	Penang, Malaysia	100°22'5"E, 5deg31'34"N	maPN	26	93	11894482	6979	88648	287
4	Thongnian, Thailand	99°48'10"E, 9deg18'6"N	maTN	35	93	10605220	6100	87742	275
5	Samut Sakon, Thailand	100° 2'6"E, 13deg22'28"N	maSS	19	93	12150330	6998	87532	384
6	Ban Kunsha, Thailand	100°26'33"E, 13deg30'1"N	maBK	35	93	12291212	6990	87583	382
7	Sanya, China	109°41'16"E, 18deg15'33"N	maSY	100	91	15241634	8087	85329	136
8	Wenchang, China	110°50'0"E, 19deg33'35"N	maWC	100	93	15431782	7512	86924	118
9	Cebu, Philippines	124° 0'25"E, 10deg21'57"N	maCB	26	94	11863938	6938	89399	366
10	Sabah, Malaysia	117°59'27"E, 5deg48'44"N	maSB	35	93	11763230	6567	86849	89
11	Bali, Indonesia	115°14'8"E, 8deg42'59"S	maBL	35	93	10450180	5837	87181	268
12	Bunbury, Australia	115°39'0"E, 33deg19'33"S	maBB	40	93	6834914	3789	82804	358
13	Darwin, Australia	130°54'14"E, 12deg27'44"S	euDW	40	92	6746212	4084	84700	165
14	Cairns, Australia	145°47'37"E, 16deg57'22"S	euCA	35	88	11609894	6518	77737	104
15	Brisbane, Australia	153° 6'42"E, 27deg21'3"S	auBS	40	93	11274220	6062	87426	759
16	Auckland, New Zealand	174°40'44"E, 36deg52'28"S	auAK	22	88	11468068	5929	76119	74

Note: [1] N is the sample size, [2] G is the number of genes sequenced, [3] S is the number of segregating sites. "ma" stands for var. *marina,* "eu" stands for var. *eucalyptifolia* and 6062"au" stands for var.*australasica* .

Table 2 Posterior probabilities of models using Approximate Bayesian Computation

Posterior probability of simulation 1: Ten models	Model 1	Model 2	Model 3	Model 4	Model5
	0.1005	0.2008	0.0997	0.0000	0.0000
	Model 6	Model 7	Model 8	Model 9	Model 10
	0.0000	0.1218	0.0001	0.0998	0.3774
Posterior probability of simulation 2: four models	Model v1	Model v2	Model v3	Model v4	
	0.0515	0.9333	0.0118	0.0034	

Figure 1 *Avicennia marina* distribution range and sampling locations. Ranges of the three varieties are shown in colors as indicated in the legend. Sampling locations are indicated by black circles with numbers. Location name and population abbreviation are as follows: 1, Meed Creek, maMC; 2, Laemson, MaLS; 3, Penang, maPN; 4, Thongnian, maTN; 5, Samut Sakon, maSS; 6, Ban Kunsha, maBK; 7, Sanya, maSY; 8, Wenchang, maWC; 9, Cebu, maCB; 10, Sabah, maSB; 11, Bali, maBL, 12, Bunbury, maBB; 13, Darwin, euDW; 14, Cairns, euCA; 15, Brisbane, auBS; 16, Auckland, auAK. Detailed information for these sampling locations is in Table 1. Morphological differences between the three subspecies in leaf, flower and fruit are presented on the right.

Figure 2 Genetic divergence and differentiation among *Avicennia marina* populations. (a) Clustering of the *A.marina* populations using principal component analysis (PCA). PCA was performed on the SNP frequency matrix. Colors indicate subspecies. (b) Multi-dimensional scaling analysis of the $F_{ST}$and D_{XY} matrices of 16 *A. marina*populations. (e-d) boxplots of D_{XY} andF_{ST} values. "au", "ma" and "eu" indicate*A. m. australasica* , *A. m. marina* and *A. m. eucalyptifolia* respectively. "maWest" and "maEast" refer to the two recognized groups of *A. m. marina* populations west and east of the Malay Peninsula (see the Results section). "BB" refers to the maBB population from Bunbury, Australia.

Figure 3 The subspecies evolved independently. a) Boxplots of ϑ computed for each gene in each population (upper graph) and barplots of mean ϑ and π values computed by pooling all SNPs in a population (lower graph). (b) Simulations reconstructing demographic history of*Avicennia marina* populations. Graphical presentation of the ten models of the three subspecies. N stands for effective size and T stands for time of split. Black arrows in models 8-10 indicate gene flow.

Figure 4 Gene flow between subspecies. (a) Graphical presentation of the four models to investigate the contrast between morphological and genetic characters of the maBB population in western Australia. vT_0 and vT_1 indicate divergence time points and N_{eu}, N_{bb}, and N_{ma} indicated effective population size. The constant bi-directional migration rates are denoted by m_a and m_b. (b) TreeMix to capture gene flow events on a population splitting graph. On the Maximum likelihood tree, each yellow line indicates a gene flow event between branches it links, with color indicating migration weight. Horizontal branch lengths of the tree are proportional to the amount of genetic drift that has occurred on the branch. The triangle matrix on the top-right indicates residual fit from the maximum likelihood tree. Residuals above zero imply candidate admixture events.

Figure 5 Networks and geographical distribution of haplotypes inferred in eight *Avicennia marina* populations. Haplotypes are indicated by different colours. Lines linking haplotypes reflect mutations, with mutations exceeding a single step marked. The geographic distribution of haplotypes is also indicated. The presented a to f cases are six typical ones to represent six types of haplotype networks. Among the 231 segments, 134, 66, 14, 11, 5 and 1 segments are classified to each type of a to f respectively.

SUPPORTING INFORMATION

The online file *supplementary information* includes supplementary tables S1-S4 and supplementary figures S1-S2.