ARTICLES
https://doi.org/10.1038/s41559-022-01744-9
nature ecology & evolution

Check for updates

Evolution of coastal forests based on a full set of mangrove genomes

Ziwen He[1,9], Xiao Feng[1,9], Qipian Chen[1,9], Liangwei Li[2,3,9], Sen Li[1], Kai Han[2], Zixiao Guo[1], Jiayan Wang[1], Min Liu[1], Chengcheng Shi[2], Shaohua Xu[1], Shao Shao[1], Xin Liu[2], Xiaomeng Mao[1], Wei Xie[1], Xinfeng Wang[1], Rufan Zhang[1], Guohong Li[1], Weihong Wu[1], Zheng Zheng[1], Cairong Zhong[4], Norman C. Duke[5], David E. Boufford[6], Guangyi Fan[2,7], Chung-I Wu[1], Robert E. Ricklefs[8] and Suhua Shi[1]✉

Genomic studies are now poised to explore whole communities of species. The ~70 species of woody plants that anchor the coastal ecosystems of the tropics, collectively referred to as mangroves, are particularly suited to this exploration. In this study, we de novo sequenced the whole genomes of 32 mangroves, which we combined with other sequences of 30 additional species, comprising almost all mangroves globally. These community-wide genomic data will be valuable for ecology, evolution and biodiversity research. While the data revealed 27 independent origins of mangroves, the total phylogeny shows only modest increases in species number, even in coastal areas of active speciation, suggesting that mangrove extinction is common. A possible explanation for common extinction is the frequent sea-level rises and falls (SLRs and SLFs) documented in the geological record. Indeed, near-extinctions of species with extremely small population size (*N*) often happened during periods of rapid SLR, as revealed by the genome-wide heterozygosity of almost all mangroves. Reduction in *N* has possibly been further compounded by population fragmentation and the subsequent accumulation of deleterious mutations, thus pushing mangroves even closer to extinction. Crucially, the impact of the next SLR will be exacerbated by human encroachment into these mangrove habitats, potentially altering the ecosystems of tropical coasts irreversibly.

Genomic sequences of one or a few closely related species have been informative about various aspects of biology. It seems time to explore the benefits of sequencing the majority of species that perform complementary functions in a community or an ecological guild. Woody plants that anchor the ecological communities on the tropical coasts, collectively referred to as mangroves, are particularly suited to such a pursuit. Mangroves, with ~70 species of diverse origins globally, are quite modest in species number after ~50 million years (Myr) of evolution. They have nevertheless provided insights into such fundamental issues as speciation[1,2], molecular convergence[3,4] and adaptive genic evolution in marginal habitats[5–8].

The full set of mangrove genomes would link evolutionary and ecological studies via molecular functions. In this first report of a series of genomic studies, we address a most urgent issue concerning the health and sustainability of the tropical ecosystems, with mangroves at the centre. Present-day tropical coasts are often sites of vibrant communities that perform important ecological functions[9–15]. For example, it has been estimated that mangroves account for 14% of carbon sequestration in the world's oceans[16]. We now examine their past as a guide to the future by analysing multiple mangrove genomes.

Obviously, the genomic evolution in any taxa would depend strongly on their habitats which, for mangroves, are the junctions between the land and the sea. An advantage of studying mangroves is that the changes in these habitats through historical time can usually be gleaned from the geological records. There are many alternative explanations for how the geological and geographical changes of these habitats may affect the mangrove ecosystems. These explanations, upon careful analyses, are often mutually compatible. In particular, sea-level rises (SLRs) appear to be a unifying force in the last 100 thousand years (kyr).

In the coming century, the projected climate change and the resultant SLR could severely impact global coasts. The severity of such events could be gleaned from the past, as sea-level changes (SLCs) have been common over geological history. For example, a mere 20 thousand years ago (ka), the sea surface was 120 m below its current level and SLRs of ~1 m per century were sustained for 12 kyr (ref. [17]). In this study, we evaluate impacts of past SLRs by examining the genomes of a large fraction of the mangrove community.

Mangrove trees are useful for assessing the impacts of past SLRs on coastal ecosystems for two important reasons. First, the small number of mangrove species makes it possible to cover their evolutionary history over >50 Myr, during which we will be able to track the invasion, proliferation and extinction of multiple mangrove taxa. Second, the many sequenced genomes can also be fully informative about the changes in the population sizes of these species over the past 150 kyr (refs. [6,18–21]). In particular, extant species may show signs of drastic population decline over periods of rapid SLR.

[1]State Key Laboratory of Biocontrol, Guangdong Key Lab of Plant Resources, School of Life Sciences, Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Sun Yat-sen University, Guangzhou, China. [2]BGI-Qingdao, BGI-Shenzhen, Qingdao, China. [3]College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China. [4]Hainan Academy of Forestry (Hainan Academy of Mangrove), Haikou, China. [5]Centre for Tropical Water and Aquatic Ecosystem Research, James Cook University, Townsville, Queensland, Australia. [6]Harvard University Herbaria, Cambridge, MA, USA. [7]State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen, China. [8]Department of Biology, University of Missouri-St. Louis, St. Louis, MO, USA. [9]These authors contributed equally: Ziwen He, Xiao Feng, Qipian Chen, Liangwei Li. ✉e-mail: lssssh@mail.sysu.edu.cn

This assessment of the impacts of past SLRs would inform us about the resilience and fragility of the coastal ecosystems in their natural state. We may then infer the impacts of the projected SLRs in the coming century by incorporating human perturbations into these baseline assessments of the past.

Results

Genomes inform evolutionary histories of mangroves. In this section, the genomes of mangroves are analysed to reveal their evolutionary history. We wish to explain the long-standing puzzle: why are there so few species of mangroves, only 70 or so across a large phylogenetic framework and over the global tropic coasts? We start with two alternative explanations: (1) few opportunities for woody plants to invade the coasts in the historical past; and (2) low rate of species proliferation after the invasion. The multiple independent origins and their timings will provide a clear resolution of the two explanations.

Genome sequencing and de novo assemblies. We used PacBio Single-Molecule Real-Time (SMRT) sequencing, 10X Genomics, single tube long fragment read (stLFR) and high-throughput chromosome conformation capture (Hi-C) technologies to generate ~11.61 trillion base pairs (bp) of raw reads (Supplementary Table 1) and then de novo assembled genomes for 48 plant species, namely, 32 species of mangroves (counting 3 subspecies of *Avicennia marina*), 10 species of mangrove associates (plants that grow mostly in regions adjoining the tidal periphery of mangrove habitats and extend into terrestrial communities) and 6 relatives of mangroves (Table 1). They are phylogenetically widely dispersed with 46 dicots, one monocot and one magnoliid, placed in 15 families (Supplementary Table 2).

The assembled genomes range in size from 198 to 1,865 megabases (Mb) (Supplementary Fig. 1), consistent with the *k*-mer-based and flow cytometric estimates of genome size (Supplementary Table 3). The N50 values of the genomes range from 0.20 to 137.88 Mb (median 12.27 Mb) (Table 1). To further evaluate the quality of the assemblies, we performed a Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis on each genome and observed high completeness with an average of 96.75% (range 89.03–99.01%; Table 1 and Supplementary Table 4). These results indicate the high quality of the genomes.

Independent origins and subsequent speciation of mangroves. The 67 species of mangroves[14,22] are distributed among 18 family-level taxa (Supplementary Table 5). Fossil records of the first occurrence of different mangrove lineages extend from the Late Cretaceous period to the Miocene epoch[23–29]. We now construct their phylogeny using whole-genome sequences at three levels. Each higher level is a denser coverage of taxa than the previous one (more and more taxa within phylogenetic frameworks of the same scale and size). We should note that the sparser phylogeny is consistent with the subsequent denser ones but permits a more straightforward inspection of features of interest. Hence, each of the three levels represents a different purpose of phylogenetic reconstruction as explained below.

At the first level, genome sequences of 31 mangroves (30 from this study and one from ref. [30]), 10 mangrove associates (this study) and 35 relatives (6 from this study) were compared to identify 729 orthologous low-copy nuclear genes (Table 1 and Supplementary Table 6). Using these orthologues, the phylogenetic tree was inferred and dated using MCMCTREE with 14 fossil calibrations (Fig. 1). In this least-dense dataset, we identify the most reliable orthologous genes that yield the three levels of phylogeny with consistency.

At the second level, we further constructed a phylogeny to obtain a comprehensive view of the origin of mangrove species. The genome-scale phylogenetic tree (backbone tree) was further expanded through a hierarchical analysis of each subsection phylogeny, comprising 30 more species of mangroves, 13 more mangrove associates and 29 terrestrial relatives (Methods and Supplementary Fig. 2). The expanded phylogeny (Extended Data Fig. 1 and Supplementary Table 7) covers >90% of the 67 known mangrove species[14,22]. The detailed reconstructed phylogeny in Extended Data Fig. 1 provides the dating of mangrove origin events across the angiosperm phylogeny in subsequent analysis.

At the third level, we put the 67 species of mangroves in corresponding positions in the large molecular phylogeny[31], containing 74,531 species and all families of extant vascular plants in Fig. 2a (Methods). This large phylogeny permits the comparison of rates of speciation and extinction between mangroves and other plants. Figure 2a shows where mangroves stand among vascular plants while Fig. 1 and Extended Data Fig. 1 are only about mangroves in relation to the close species. The 67 species of mangroves have 27 independent origins, including 25 origins in dicots, one origin in monocots and one origin in ferns. The order Myrtales has the largest number of mangrove origins (6), while the Malpighiales contains the largest number of species (19).

The origination times of the 21 lineages of dicot mangroves are shown in Fig. 2b, representing 57 extant mangroves. These dated events are distributed nearly continuously over the geological time span of the Cenozoic period, suggesting that the opportunities for invading the intertidal habitat were continually available. The other four undated origins of dicot mangroves (*Brownlowia tersa*, *Crenea patentinervis*, *Diospyros littorea* and *Mora oleifera*) may all be relatively recent since their closely related congeneric species are non-mangroves.

The earliest period, spanning 39–50 Mya (period I, see ref. [4]), is of particular importance. The three most species-rich clades, that is, Rhizophoraceae, *Sonneratia* and *Avicennia*, originated within this period, which also included the origin of *Pemphis*. This wave of mangrove origination was preceded by the Palaeocene–Eocene Thermal Maximum (PETM, ~55 million years ago, Ma), during which the global average temperature rose dramatically, by ~6 °C, within ~20 kyr (ref. [32]). The PETM was followed by the Eocene Thermal Maximum 2 (ETM2) and the 2 Myr-long Early Eocene Climatic Optimum[32]. The habitats of the ancestral mangroves may have been inundated several times by the rising sea levels, during which these plants eventually become adapted to the intertidal environment.

Species proliferation after mangroves invading the coasts. Although mangroves are the dominant woody plants on the tropical coasts, they comprise only ~70 species[14,22]. There may be two main explanations: (1) the lack of opportunities for woody plants to invade the tropical coasts and evolve into mangroves; and (2) the low rate of species proliferation after they invade the coasts.

The first explanation is not compatible with the evidence. While it has been suggested that the window of opportunity for invading the intertidal zone is narrow, largely confined to the brief PETM period[33], Fig. 2a,b show that multiple taxa have independently invaded the intertidal zones 27 times over an extended period of 40–50 Myr.

For the second explanation, a comparison between the number of dicot mangroves and other dicot lineages through time is displayed in Fig. 2c. Since there is barely a case whereby mangrove speciation gave rise to non-mangrove species, the line of mangrove proliferation can indeed be compared to that of other plants. The low rate of species proliferation among mangroves is visually evident across the entire time span. Figure 2c shows that the rate of species proliferation is ~0.02 Myr^{-1} for mangroves, much lower than the rate of 0.05–0.07 Myr^{-1} for non-mangroves in the last 15–60 Myr.

Low speciation and/or high extinction rate. The previous section shows that the invasion by woody plants into the mangrove habitats is continual with invasions taking place as early as >50 Ma. Hence,

Table 1 | Information of genome assemblies for mangroves, mangrove associates and relatives

No.	Organism	Type	Sequencing method	Total assembled size (Mb)	N50 (Mb)	BUSCO completeness (%)	
						Genome	Protein
1	*Rhizophora stylosa*	Mangrove	10X + Hi-C	247.16	12.51	96.96	93.18
2	*Rhizophora mucronata*	Mangrove	10X + Hi-C	234.16	12.03	95.72	90.95
3	*Rhizophora apiculata*	Mangrove	PacBio + Hi-C	232.04	13.31	98.45	97.65
4	*Rhizophora mangle*	Mangrove	PacBio	281.12	1.93	97.71	95.54
5	*Ceriops tagal*	Mangrove	10X + Hi-C	244.25	11.73	96.22	90.52
6	*Ceriops zippeliana*	Mangrove	PacBio	320.38	3.46	96.53	93.43
7	*Kandelia candel*	Mangrove	PacBio	224.71	9.45	98.39	98.33
8	*Bruguiera sexangula*	Mangrove	PacBio + Hi-C	241.30	12.79	97.89	96.90
9	*Bruguiera gymnorhiza*	Mangrove	PacBio + Hi-C	270.60	13.02	97.77	94.24
10	*Carallia pectinifolia*	Non-mangrove	PacBio + Hi-C	359.74	31.58	98.33	94.73
11	*Xylocarpus rumphii*	Mangrove associate	PacBio	218.26	2.63	98.64	92.13
12	*Xylocarpus moluccensis*	Mangrove	PacBio	280.87	1.27	97.03	93.93
13	*Xylocarpus granatum*	Mangrove	PacBio + Hi-C	300.08	9.56	99.01	94.86
14	*Swietenia macrophylla*	Non-mangrove	PacBio	270.80	2.42	96.34	90.52
15	*Heritiera littoralis*	Mangrove	PacBio	1,021.59	4.00	98.20	92.07
16	*Sonneratia ovata*	Mangrove	10X + Hi-C	217.58	15.37	97.46	89.65
17	*Sonneratia apetala*	Mangrove	10X + Hi-C	249.05	15.08	97.09	93.74
18	*Sonneratia alba*	Mangrove	PacBio + Hi-C	204.46	15.69	97.21	91.95
19	*Sonneratia caseolaris*	Mangrove	Illumina + Hi-C	197.75	14.51	97.83	92.44
20	*Trapa natans*	Non-mangrove	PacBio	497.24	11.66	96.41	93.12
21	*Duabanga grandiflora*	Non-mangrove	stLFR	419.86	5.42	98.27	96.53
22	*Lagerstroemia speciosa*	Non-mangrove	PacBio + Hi-C	319.66	12.74	96.28	92.50
23	*Pemphis acidula*	Mangrove	10X + Hi-C	495.42	28.96	93.43	90.02
24	*Lumnitzera littorea*	Mangrove	10X + Hi-C	1,444.51	114.49	97.89	93.37
25	*Lumnitzera racemosa*	Mangrove	PacBio + Hi-C	1,317.41	106.37	96.03	94.86
26	*Laguncularia racemosa*	Mangrove	PacBio + Hi-C	1,145.61	3.41	94.98	91.39
27	*Conocarpus erectus*	Mangrove	PacBio	1,536.32	4.08	97.15	93.49
28	*Combretum micranthum*	Non-mangrove	PacBio	1,865.11	0.72	95.04	87.05
29	*Pongamia pinnata*	Mangrove associate	PacBio + Hi-C	1,074.15	88.49	96.78	95.66
30	*Avicennia marina* subsp. *marina*	Mangrove	PacBio + Hi-C	457.80	14.15	97.40	92.44
31	*Avicennia marina* subsp. *australasica*	Mangrove	10X + Hi-C	493.61	13.60	94.86	87.61
32	*Avicennia marina* subsp. *eucalyptifolia*	Mangrove	10X + Hi-C	468.09	14.17	96.34	91.39
33	*Avicennia officinalis*	Mangrove	stLFR	447.06	0.20	96.78	92.50
34	*Avicennia alba*	Mangrove	PacBio	457.14	7.63	97.96	96.65
35	*Avicennia rumphiana*	Mangrove	PacBio	470.61	7.04	97.71	97.15
36	*Avicennia germinans*	Mangrove	10X + Hi-C	656.67	16.43	94.55	88.85
37	*Premna obtusifolia*	Mangrove associate	PacBio	770.51	2.75	97.58	91.39
38	*Clerodendrum inerme*	Mangrove associate	stLFR + Hi-C	867.36	23.82	89.03	86.12
39	*Scyphiphora hydrophyllacea*	Mangrove	10X + Hi-C	1,241.80	99.19	95.97	90.83
40	*Morinda citrifolia*	Mangrove associate	PacBio	890.83	0.73	98.27	96.16
41	*Pluchea indica*	Mangrove associate	PacBio	1,776.74	1.26	97.46	71.07
42	*Scaevola sericea*	Mangrove associate	10X + Hi-C	1,126.42	119.65	94.49	88.17
43	*Scaevola hainanensis*	Mangrove associate	10X + Hi-C	1,260.96	137.88	93.80	86.43
44	*Aegiceras corniculatum*	Mangrove	PacBio	826.70	3.87	94.55	95.29
45	*Barringtonia racemosa*	Mangrove	PacBio	1,263.22	9.85	98.64	98.08
46	*Barringtonia asiatica*	Mangrove associate	PacBio	787.47	9.73	97.71	95.42
47	*Annona glabra*	Mangrove associate	PacBio + Hi-C	1,027.31	75.19	97.65	97.40
48	*Nypa fruticans*	Mangrove	PacBio + Hi-C	489.48	27.90	98.39	90.02

Short names used under 'Sequencing method' are: PacBio, PacBio SMRT sequencing; 10X, 10X Genomics; stLFR, single tube long fragment read; Hi-C, high-throughput chromosome conformation capture; Illumina, paired-end and mate-paired sequencing.

the increase in species number in mangroves may be slow after the invasion into the new habitats. In this context, there are two alternative hypotheses. Hypothesis I is that the speciation rate of mangroves is low and hypothesis II is that the extinction rate of mangroves is unusually high. It is important to note that these two hypotheses are mutually compatible. While many attempts have been made to simultaneously estimate the two rates, it has been shown recently that such estimates are mathematically flawed[34]. Both should be tested on the basis of different types of information.

This current study will focus on hypothesis II as hypothesis I has been extensively addressed in previous studies[1], which do not support this hypothesis. We shall discuss hypothesis I briefly mainly because the biological and geological features that are unique to mangrove habitats may affect both hypotheses, on speciation and species extinction, respectively. Two of the main features are: (1) the geographical distribution of mangroves, being long and narrow (or '1D, one-dimensional' according to ref. [1]), that restricts gene flow; and (2) frequent SLCs, either rises (SLRs) or falls (SLFs).

The two explanations have turned out to be the two sides of the same coin. While the 1D distribution may influence speciation in the mode of isolation-by-distance, the pattern of population differentiation shows the divergence to be driven by ocean currents, rather than by the geographical distance (Fig. 2 of ref. [1]). Nevertheless, the 1D distribution, compounded by frequent SLCs in the geological past, indeed drives speciation on the Indo-Malayan coasts in a manner referred to as the MIM (for mixing–isolation–mixing) model of speciation[1]. In the MIM model, diverging populations experience episodes of gene flow during speciation that punctuate longer periods of geographical isolation. The mangrove species on the Pacific coasts versus the Indian Ocean coasts are separated by the Strait of Malacca. In the last 5 Myr, the two ocean coasts are connected or disconnected via the Strait as the sea level rises and falls, due to the shallowness of the Strait (~25 m deep). On average, the cycle consists of 10 kyr with gene flow and 100 kyr in isolation (hence, mixing–isolation and mixing again).

The MIM mechanism is efficient in generating new species in an exponential manner. He et al.[1] shows that if each MI cycle leads to m species ($1 < m < 2$; $[m-1]$ being the probability of speciation), then there would be m^n species after n cycles[1]. For mangrove species found on both ocean coasts connected by the Strait of Malacca, m is about 1.03–1.05. Hence, the speciation time is ~1.5–2.5 Myr. Therefore, mangroves in the Indo-Malayan region in the last 5 Myr could have an above-average (or at least average) speciation rate among woody plants.

Past population reductions bordering on extinctions. In this section, we test hypothesis II by observing the frequency of extant mangrove species becoming endangered in their evolutionary past, in particular, in the last 120 kyr. During this period, SLRs were common, rapid and sustained. We therefore expect to see signs of near extinction in past population sizes (N). There are indeed hints that mangroves may be constantly 'endangered'. The heterozygosity values of mangroves are generally lower than their non-mangrove relatives (Fig. 2d). The low heterozygosity of mangroves suggests that their effective population sizes (N_e) were much smaller than those of the present-day populations.

Changes in mangrove N_e inferred from their genomes. We estimate the past population sizes of mangrove species using the pairwise sequentially Markovian coalescent (PSMC) model applied to whole genomes[19]. The divergence between a pair of orthologue genes reflects their coalescence time and, hence, the effective population size (N_e) of the gene region. The distribution of the coalescence time between all pairs of loci in the diploid genome can be used to estimate the historical population size changes. In short, the PSMC method uses the local density of heterozygous sites across the genome, partitioned by the inferred recombination events, to estimate the demographic history of a population.

Figure 3 shows that N_e has fluctuated substantially over the last 300 kyr, which span three glacial cycles. During this time span, N_e has varied 10- to 100-fold in these species. For most mangrove species, N_e declined after the end of the last glaciation when the sea level began to rise rapidly (Fig. 3 and Supplementary Fig. 3). In particular, during the period 20 to 5 kyr before present, N_e often decreased to hundreds of individuals. As shown recently[35], the rise in sea level, >0.6 m per century, is likely to have nullified usable habitats for mangroves. It is therefore plausible that almost all mangrove species experienced population size reduction during SLRs. This is depicted in the model of Fig. 4a, which shows the reduction in N in a hypothetical mangrove population.

The pattern of a long steady population decline could explain the low genetic diversity in contemporary populations. Compared to the shrinkage N_e of mangrove species, the N_e of mangrove associates and non-mangroves in the same period are variable with no clear pattern (Supplementary Figs. 4 and 5).

Accumulation of deleterious mutations when N decreases. As shown in Fig. 3, the decline of N_e is well correlated with the speed of SLCs, which presumable affects the population size N. It would be desirable to have a different line of evidence corroborating the PSMC determination of N_e. To that end, we use the fixation of slightly deleterious mutations for corroboration.

A slightly deleterious mutation is one that has a selective disadvantage close to, or smaller than, the strength of genetic drift[36–39]. Here, we use the A/S ratio (where A is the number of non-synonymous substitutions and S is that of synonymous ones) to measure the accumulation of deleterious mutations by assuming that many non-synonymous changes are slightly deleterious. Deleterious mutations during periods of low N_e should be fixed more readily than in other time periods. Low N_e usually also means small N_{ALL}.

As shown in Supplementary Fig. 6, the A/S ratio of each genomic region is negatively correlated with the 'cumulative' effective population size associated with that region. If we consider deleterious mutations that reduce fitness less than $1/N_e$ to be effectively neutral, then the number of effectively neutral mutations would increase as N_e decreases and the A/S ratio would be negatively correlated with N_e. Small population size and the resultant accumulation of slightly deleterious mutations may also signal the decline in population fitness. In particular, in the last 5 kyr, N_{ALL} has been large, resulting in many deleterious mutations being kept at low frequencies in the population. A drastic decrease in mangrove population sizes in the coming century may allow many of these deleterious mutations to increase in frequency due to more rapid genetic drift. These deleterious mutations may further depress the fitness of mangrove species.

Will mangroves be endangered as the sea level rises? A main objective of analysing the past is to project the health of the global mangrove ecosystems in the next century. If the decline will take centuries before it would impact the ecosystems, the urgency may not be justified. Indeed, the across-the-board declines in N_e among mangroves depicted in Fig. 3 show the decline to be slow and gradual. In this last section, we present an alternative and probably more realistic, prediction of N_e decline due to population subdivision that is far more rapid than by the conventional PSMC analysis.

We should note that the effects of SLRs on the Indo-Malayan coasts are not restricted to a single population of each mangrove species (Fig. 4a). Instead, as local populations migrate upland during SLR, they may become disconnected from others, resulting in the total population becoming fragmented. Each subpopulation also becomes smaller in the process (Fig. 4b). One line of evidence for population subdivision comes from the analysis of speciation in mangroves[1]. As the sea level rises and falls, different

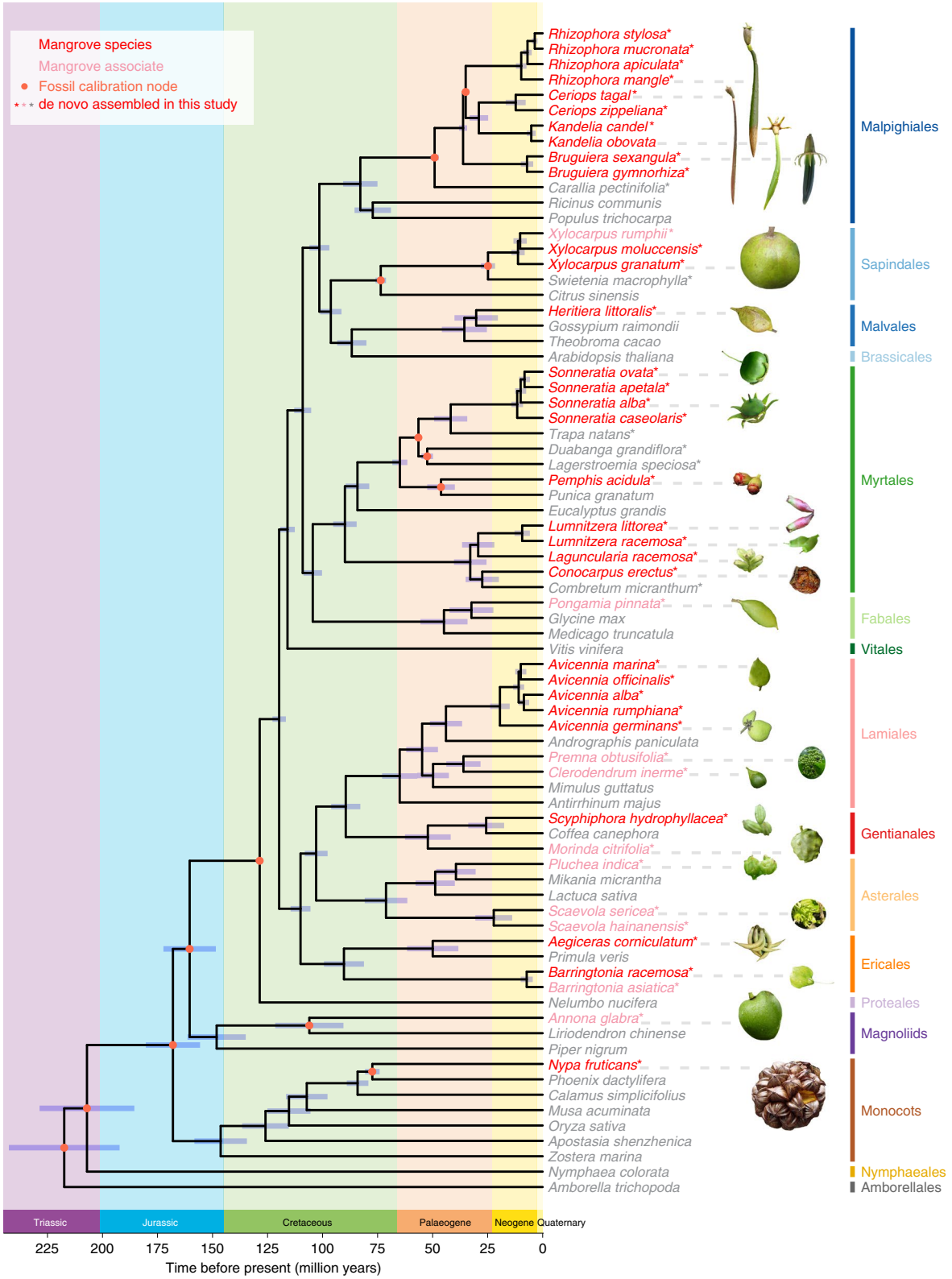


Fig. 1 | Genome-scale phylogeny of mangroves. The maximum likelihood phylogenetic tree contains 31 species of mangroves, 10 mangrove associates and 35 relatives. The node bars are 95% confidence intervals. The 14 red nodes indicate fossil calibration nodes. Mangroves are indicated by red typeface, mangrove associates by pink typeface and mangrove relatives by grey typeface. The de novo genomes that we assembled are marked with asterisks. The sources of the fruit photos are listed in Supplementary Table 8. Photo credits: *Sonneratia ovata*, Ron Yeo; *Lumnitzera littorea*, Roger Fryer.

subpopulations may become connected or isolated at different times, leading to MIM cycles. He et al.[1] suggested that MIM cycles drive speciation and biodiversity along the Indo-Malayan coasts[1].

With two subpopulations, or more, the coalescence time (thus, the estimated N_e) would depend on two aspects of population structure. Figure 4b shows such a scenario where X_t are the coalescence

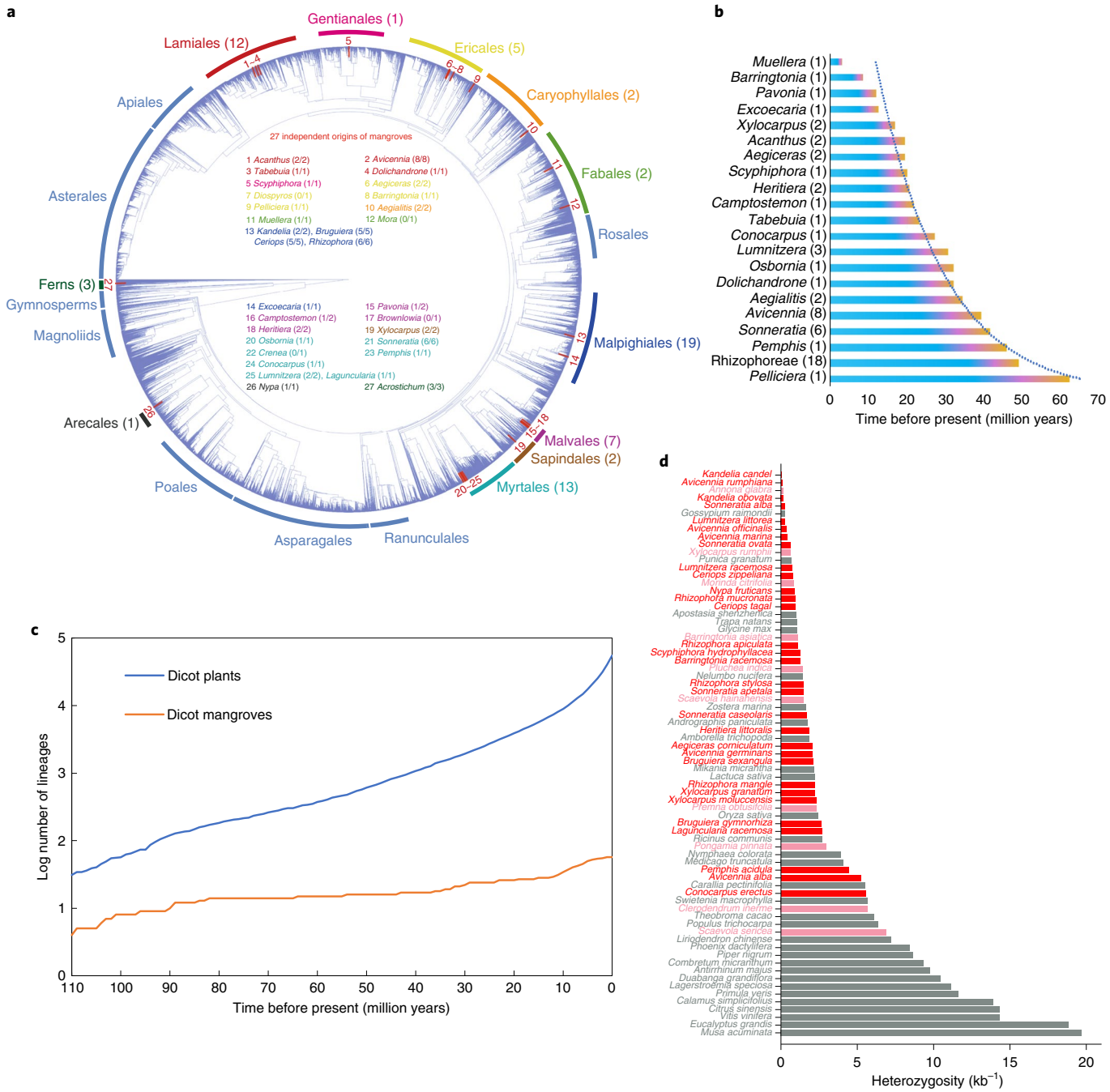


Fig. 2 | Independent origins of mangrove lineages. a, The 27 independent origins of mangroves in the plant phylogeny. The plant phylogenetic tree contains 74,531 species and all families of extant vascular plants[31]. The numbers following each genus name denote the number of mangrove species used in Extended Data Fig. 1 and the total number of species of mangroves in the genus. **b**, The time of origin for 21 lineages of dicot mangroves. The dotted line indicates a smooth curve, in disagreement with an earlier claim of a narrow time window for mangrove origin. The number displayed in parentheses following each genus name indicates the number of species of mangroves in that clade. The colour gradient in each bar represents the process that mangrove species separated from their terrestrial relatives and then fully adapted to the intertidal zone. **c**, The number of dicot mangrove and dicot plant lineages through time. **d**, Average genome-wide heterozygosity values of plant species. The *x* axis represents the average proportion of sites within the autosomes that are heterozygous. Red bars, mangroves; pink bars, mangrove associates; dark grey bars, non-mangroves.

times within each subpopulation and Y_t the coalescence time between two subpopulations ($t=0$, 1 and 2 represent three time periods). In Fig. 4b, X_t reflects the subpopulation size N_{sub} at time period t. The total number of mangrove trees, N_{ALL}, in the region should be $N_{ALL}=k\times N_{sub}$, where k is the number of subpopulations. Both N_{ALL} and N_{sub} affect X_t but N_e, as obtained from PSMC, is a function of X_t and Y_t. If the divergence between subpopulations is substantial, we may obtain $N_e \gg N_{ALL}$. The interest in conservation biology is in N_{ALL}, which influences the ecological functions (such as carbon sequestration) of mangroves.

Here, we attempt to connect the N_e estimates from PSMC to N_{ALL}. In subdivided populations, PSMC analysis yields an interesting but somewhat unpredictable pattern. In fact, the gradual decline in N_e shown in Fig. 3 is often not true when the population is

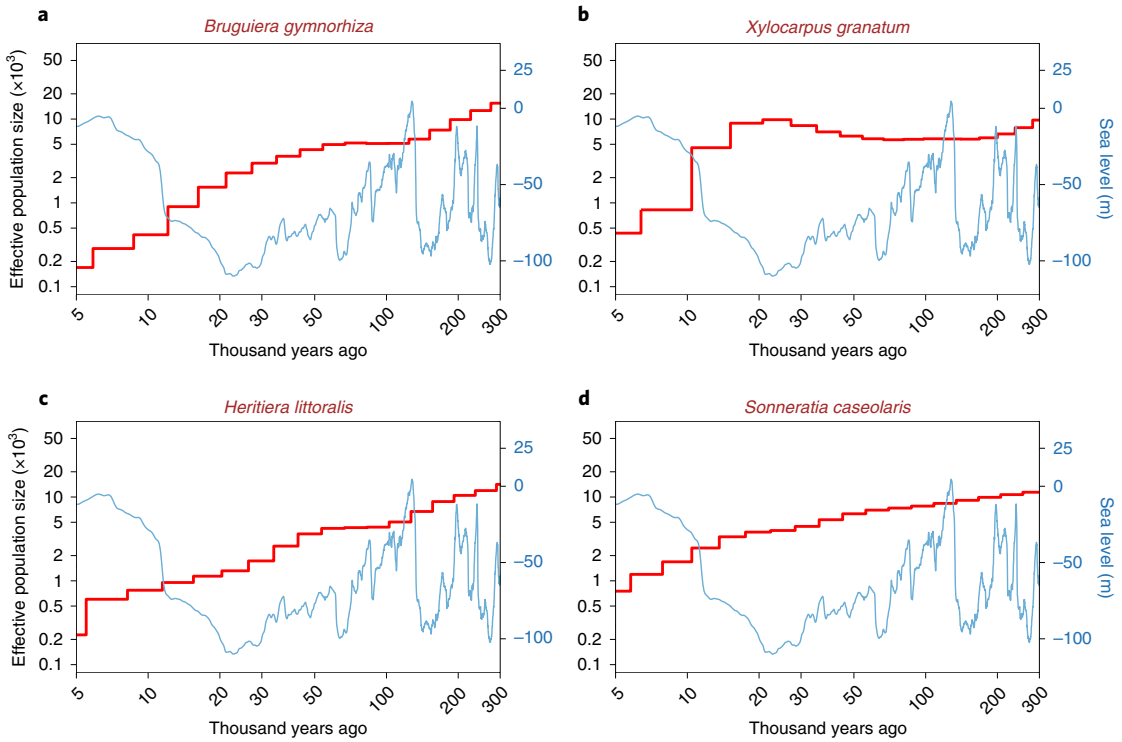


Fig. 3 | Demographic histories of four mangrove species. a–d, Inferred historical population sizes from pairwise sequentially Markovian coalescent analyses of four mangrove species. Red lines represent inferred N_e values through time. Blue lines present relative sea levels through time based on the estimates in ref. [17]. The *y* axis at the left of each graph represents $\log_{10}$-scaled effective population size and that at the right represents a linear scale of relative sea level.

subdivided. We also note that PSMC is not the only method to infer past changes in N_e. Previous work has developed a 'stairway plot' method for the same purpose[40,41]. The two methods generally yield comparable results for a single undivided population[40] but they may make very different inferences of the past for subdivided populations. In Supplementary Fig. 7, the inferred N_e values of *A. marina* experienced a bottleneck at 30–60 ka, while the PSMC estimated a continuous decline of N_e in the last 300 kyr. The differences between the two inferences, provide further support for population subdivision in mangroves.

Figure 4c-f present various scenarios of population subdivision. In these examples, $N_{ALL}=50$ (5×10 or 25×2) or $N_{ALL}=500$ (50×10 or 250×2). Note that these are very small population sizes. When a population decreases precipitously, as shown by the black line, the PSMC results, shown in the red lines, often lag behind the actual drop. Thus, when *N* increases again, the PSMC pattern has not yet reached the low equilibrium. In short, the slow decline in N_e obtained by the PSMC method, as shown in Fig. 3, could be driven by much more abrupt decreases in N_{ALL} if the total population is fragmented.

We only show some of the more extreme scenarios whereby N_{sub} becomes very small. Given the slow rate of sediment accretion during SLR[35,42,43], especially for a prolonged period (20–8 kyr before present), such reductions may not be implausible. To our surprise, these extreme scenarios are still compatible with the genomic data. Figure 4c,d are important because N_{ALL} as low as 50 could mean imminent extinction for many mangrove species. If extinction is common for mangrove species, the extant mangrove species should show signs of endangered status (low population size) in the past. This endangered status of very small population size might explain the relatively few existing species (~70) of mangroves worldwide. Most important, N_{ALL} is tied to the ecological functions that include sequestering carbon and providing habitats for intertidal organisms in the tropics.

Discussion

Mangrove forests dominating the ecosystems of the tropical coasts are a major carbon sink globally. However, the total number of mangrove species has not even tripled in the last 50 Myr. By using the large genomic data of all known mangrove species, we reconstruct the phylogenies among mangroves as well as between mangroves and terrestrial vascular plants. The phylogenetic relationships reveal 27 independent invasions by terrestrial plants into these habitats over the last 50 Myr. After each invasion, the colonizing species either fails to produce new species through speciation or the new species become extinct.

The frequent SLCs in the coastal environments may stress mangrove species to the brink of extinction when SLCs are rapid and sustained over time. While SLCs may fulfil the criteria of the MIM speciation model that allows species number to increase rapidly[1], extinction must be sufficiently prevalent to offset the formation of new species. Here, by analysing the genomic heterozygosity, we find that most mangrove species had very small effective population sizes during periods of rapid SLCs. In short, mangrove populations may have been barely able to keep up with the continually shifting habitats on the coasts.

There are a few convincing cases of extinct fossil mangrove taxa, namely *Wetherellia* and *Paleowetherellia*[27] and the *Brevitricopites* group[44]. Therefore, the phylogenetic history inferred from mangrove genomes, changes in their heterozygosity levels and fossil records all support the interpretation of high extinction rate. While today's tropical coasts are teeming with mangrove trees, this might not have been the historical norm. The sea level has been unusually stable over the last 6 kyr (ref. [45]). Before this recent stasis, the sea level had been rising and falling frequently.

The present analysis shows that the ecosystems of tropical coasts may be far more fragile than realized, even in pristine settings free of human activities. Furthermore, additional elements in the

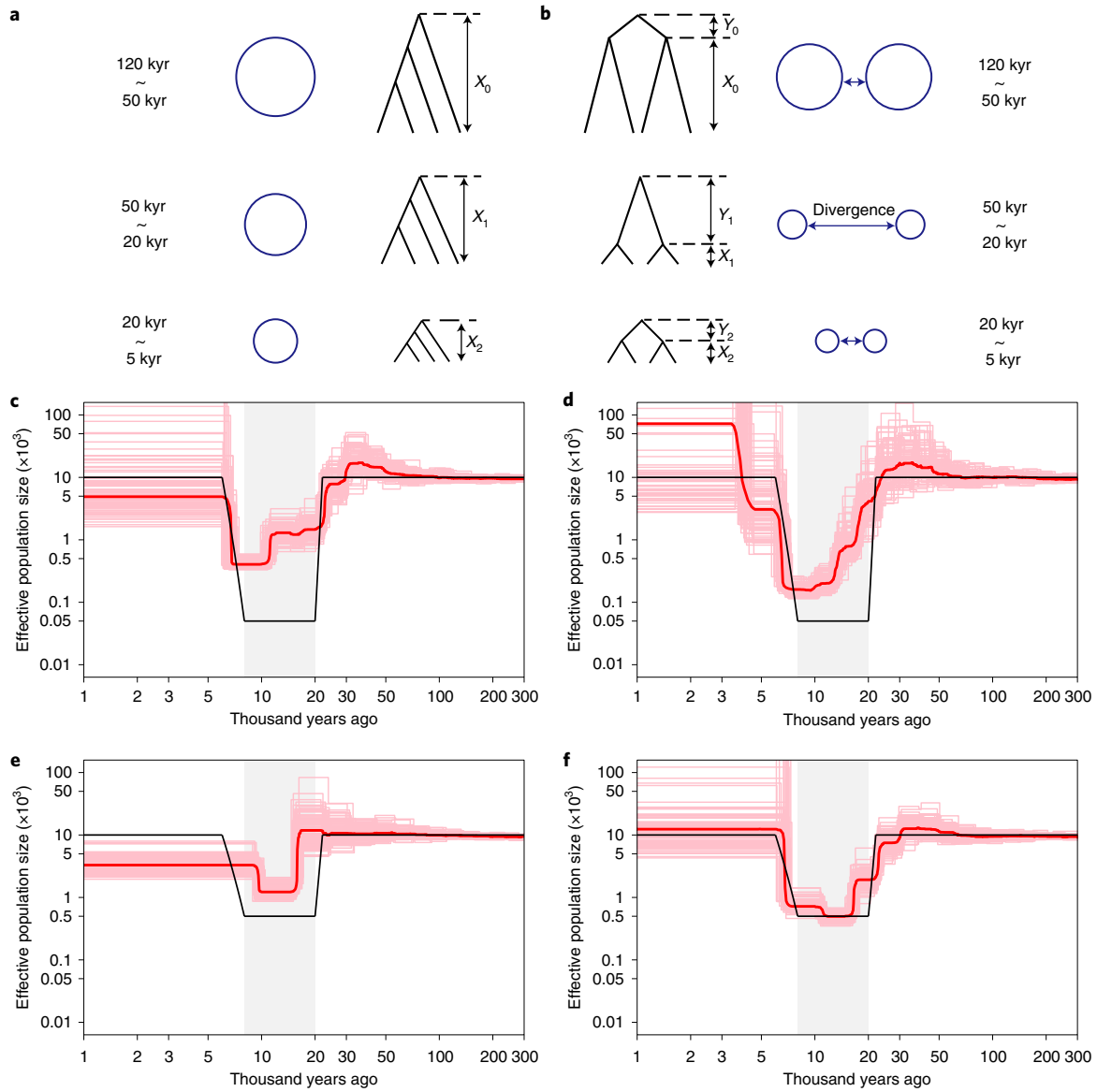


Fig. 4 | Model of population subdivision and inferred histories of the PSMC method using simulated samples. a,**b**, Model of population shrinkage (**a**) and subdivision (**b**). X_t, coalescence time within population at t ($t=0$, 1 and 2, represent three time periods); Y_t, coalescence time between population at t; size of circles, effective population size (N_e). **c**–**f**, PSMC results of simulations. A large population ($N_{ALL}=10{,}000$, 22 ka) is fragmented into k small subpopulations (N_{sub}) that are fully isolated until 8 ka. The newly merged population gradually grows to N_{ALL} in the last 8 kyr. **c**, $k=10$ and $N_{sub}=5$; **d**, $k=2$ and $N_{sub}=25$; **e**, $k=10$ and $N_{sub}=50$; and **f**, $k=2$ and $N_{sub}=250$. Pink lines, PSMC results of 100 replicates; red lines, median values of the PSMC results of 100 replicates; black lines, actual simulated population sizes.

forthcoming century might push these ecosystems beyond historical norms. The first such new element has been carbon emission from burning fossil fuels. Mangrove ecosystems are a major global carbon sink[16,46–50] but this capacity will be greatly reduced by mangrove destruction. The extent to which the loss of mangroves may exacerbate global warming has been suggested[10,51–53] and challenged[54]. The uncertainty about the consequences might not be resolved until it is too late.

The second element is the disruption of mangrove habitats by human settlements. In the past, mangroves have been able to move inland during SLRs but, in the future, many mangrove forests simply will have no place to retreat to, due to human dominance of coastal areas. This phenomenon has been well documented[43,55]. Lastly, as shown in Supplementary Fig. 6, very small populations tend to accumulate deleterious mutations, which may cause a decline in population fitness at a time when the demand for adaptability in the new upland environments will increase.

This study provides an example in which the many genomes of an ecological community offer comprehensive information about its ecology and evolution. In this first study using the dataset, we document the extreme fragility of the mangrove ecosystems during past SLRs. Many of them survived. However, compounded by new human-made disruptions, the future will probably be far worse than the past—to the point that mass extinctions become a serious possibility.

Methods

Plant materials. A total of 48 plant genomes were sequenced in this project, including 32 mangrove species (counting three subspecies of *Avicennia marina*), 10 mangrove associates and 6 relatives. Among the 48 species, five genome assemblies

(*Rhizophora apiculata*, *Bruguiera gymnorhiza*, *Sonneratia alba*, *Sonneratia caseolaris* and *A. marina* subsp. *marina*) were improved to chromosome level upon the previous version published by the authors[4,6]. All tissue samples were collected from multiple sources, listed in Supplementary Tables 9 and 10. For population genomic data, 15 individuals of *Sonneratia alba* from Ngao, Ranong, Thailand and 12 individuals of *A. marina* from Qinglan Harbour, Wenchang, China, were collected.

Genome sequencing, assembly and annotation. *Genomic DNA and RNA preparation.* Fresh and healthy leaves of 48 species (counting three subspecies of *A. marina*) (Supplementary Table 9) were harvested and immediately frozen in liquid nitrogen, followed by preservation at −80 °C in the laboratory before DNA extraction or Hi-C library construction. Fresh and healthy leaves of the other two species (Supplementary Table 9) were dried and preserved with silica gel.

High-quality genomic DNA was extracted from leaves using TIANamp Plant DNA Kits (Tiangen) or the CTAB (hexadecyltrimethylammonium bromide) method[56]. The quality and quantity of the extracted DNA were examined using a NanoDrop 2000 spectrophotometer (NanoDrop Technologies) and electrophoresis on a 0.8% agarose gel, respectively. High-quality RNA was also extracted from several tissues (Supplementary Table 10) using TRNzol Universal Reagent (Tiangen) or the modified CTAB method[57].

Library construction and sequencing. SMRT long reads sequencing was performed with a PacBio sequel II platform (Pacific Biosciences). The SMRT-bell libraries for different plant species with 20–40 kb long insert were constructed from sheared genomic DNA using a template library preparation workflow. The libraries were sequenced on PacBio SMRT cells 8 M (acquiring one video of 15 h per SMRT cell) using a PacBio Sequel II instrument.

For 10X Genomics library construction, the long genomic DNA was denatured according to the standard protocol of Genome Library Kit & Gel Bead Kit v.2 (10x Genomics). Standard circularization and DNA nanoballs were prepared. Linked reads of the 10X Genomics library were generated on the BGISEQ-500 platform.

The stLFR libraries[58] were constructed according to the standard protocol, via the MGIEasy stLFR library preparation Kit (MGI Tech) and were sequenced on the BGISEQ-500 platform.

Hi-C libraries were constructed from tender leaves. The leaves were fixed with formaldehyde and lysed and then the cross-linked DNA was digested with MboI. Restriction fragment ends were biotinylated and ligated. The purified DNA was physically sheared to a length of ~400 bp. The Hi-C libraries were then sequenced on the Illumina or BGISEQ-500 platform.

For DNA short-read sequencing, paired-end libraries (300–500 bp) were prepared for sequencing on Illumina or BGISEQ-500 platform. The RNA-seq library was sequenced on the Illumina or BGISEQ-500 platform for facilitating the prediction of protein-coding genes.

For population genomic data, short-read libraries were sequenced using the Illumina platform and yielded 98.25 gigabases (Gb) in *S. alba* and 165.24 Gb in *A. marina* (Supplementary Table 11).

Genome assembly. The quality-filtered reads were used for genome size estimation. First, we generated the 17-mer occurrence distribution of short-insert reads using the *k*-mer method[59]. Then, we estimated the genome size using genomic character estimator[60] and GenomeScope[61]. In contrast, flow cytometry analyses of nuclear DNA content in 29 plant species (Supplementary Table 3) were performed in this study, as well as in our previous studies[4,5].

We assembled the de novo genomes of 31 species (Table 1 and Supplementary Table 1) based on PacBio long reads with FALCON (ref. [62]), wtdbg2 (ref. [63]), hifiasm (ref. [64]) or MECAT2 (ref. [65]) with optimized parameters. The assemblies were further polished with Quiver (SMRT Analysis v.2.3.0)[66] or Racon v.1.3.1 using long reads[67]. To improve primary assembly accuracy, we performed several rounds of iterative error correction using Pilon v.1.22 based on Illumina short reads[68].

We assembled 13 plant genomes with 10X Genomics libraries and three plant genomes with stLFR libraries using Supernova v.2.0.1 (ref. [69]) (Table 1 and Supplementary Table 1). We first transformed stLFR barcodes to 10X Genomics using scripts available on GitHub (https://github.com/BGI-Qingdao/stlfr2supernova_pipeline). We then assembled de novo genomes based on stLFR clean reads after barcodes transformed or 10X Genomics reads using Supernova v.2.0.1 (ref. [69]). The gaps in the assemblies were filled using GapCloser v.1.12 with paired-end reads[70].

On the basis of Hi-C data and genome assemblies of 28 species (Table 1 and Supplementary Table 1) (including five species genome assemblies published earlier by the authors[4,6]), we improved them to generate pseudochromosome-level genomes. The Hi-C data were first evaluated and qualified using HiC-Pro[71] and then used to generate Hi-C maps by Juicer[72]. The scaffolds were roughly split by Juicebox[73] and anchored to chromosomes using 3D-DNA[74].

Genome annotations. We identified repetitive sequences in each of the whole genomes by combining homology-based and de novo approaches. For homology-based prediction, we identified the known transposable elements (TEs) within the genome using RepeatMasker (open-4.0.9, https://www.repeatmasker.org) with the Repbase TE library[75,76]. Then, RepeatProteinMask searches were conducted using the TE protein database as a query library. For de novo prediction, we constructed a de novo repeat library of the genome using the RepeatModeler pipeline[77]. We performed a de novo search for long terminal repeat retrotransposons using LTR_FINDER v.1.0.7 (ref. [78]). We also identified tandem repeats using Tandem Repeat Finder package[79] and we determined the non-interspersed repeat sequences, including low-complexity repeats, satellites and simple repeats, using RepeatMasker. Finally, we merged the identifications of the two methods and used RepeatMasker. The repeat contents of the de novo assembled genomes are shown in Supplementary Fig. 1 and Supplementary Table 12.

We conducted annotation of protein-coding genes in the assembled genome by integrating de novo prediction, homology-based prediction and RNA-seq-based prediction. For de novo prediction, we adopted Augustus v.3.3.1 (ref. [80]), GENSCAN (ref. [81]) and GlimmerHMM (ref. [82]) to analyse the masked genome. The protein sets were collected for homology-based prediction and chosen as homology-based evidence from sequenced related plant species and model plant species. The GeneWise[83] was applied on the basis of aligned output via BLAT[84]. The clean RNA-seq reads were used as RNA-seq-based evidence. The gene structure prediction was performed using Tophat2 v.2.1.1 and Cufflinks v.2.2.1 (refs. [85–87]). Finally, all predictions were integrated to produce non-redundant and consensus gene sets using Maker v.3 (ref. [88]) or EVidenceModeler[89].

To further evaluate the qualities of genome assemblies and gene predictions, we used a BUSCO analysis on each genome based on BUSCO v.4 with the embryophyta_odb10 database (1,614 single-copy genes)[90].

Phylogeny reconstruction. To obtain a comprehensive phylogeny of mangrove species, we constructed the phylogenetic tree using whole-genome sequences and expanded the tree based on sequences of the transcriptome, chloroplast genome and DNA fragments through a hierarchical analysis (Supplementary Fig. 2).

We first constructed the phylogenetic tree and estimated the divergence time using whole-genome sequences of 31 mangrove species, 10 mangrove associates and 35 relatives (Table 1 and Supplementary Table 6). Low-copy nuclear (LCN) genes in 76 angiosperms as well as the gymnosperm *Gnetum montanum* (outgroup) were identified on the basis of the embryophyta_odb10 lineage ancestral variant dataset (containing a consensus sequence and variants of the extant sequences) in BUSCO v.4 (ref. [90]), which contains genes present as single-copy orthologues in at least 90% of 50 land plants. We gradually aligned the low-copy orthologous proteins using MAFFT v.7.429 with the merge method[91,92]. Briefly, we first used L-INS-i (a highly accurate option in MAFFT) to generate a sub-multiple sequence alignment (sub-MSA) of each LCN orthogroup in the same-order plant species. Then we merged these sub-MSAs into a single MSA by L-INS-i. The aligned protein sequences among 77 plant species were used to generate codon alignments with PAL2NAL[93]. We further trimmed alignments using Gblocks 0.91b, with the gap treating parameter 'half'[94] and discarded alignments shorter than 150 bp. Finally, we identified 729 LCN orthologous groups. On the basis of these alignments, we inferred a phylogenetic tree by RAxML-NG[95] with the GTR + GAMMA + I model and 1,000 bootstrap replicates with the gymnosperm *G. montanum* (order Gnetales) as the outgroup.

To estimate the evolutionary timescale for mangrove species, we dated the tree using MCMCTREE from the PAML v.4.9j package[96] with approximate likelihood calculation, which we chose due to its speed and efficiency in Bayesian dating analysis using large datasets[97]. Fourteen reliable fossil calibrations were incorporated to constrain the tree nodes (Supplementary Table 13). The MCMC analyses were run for 10 million generations, sampling every 500 generations after a burn-in of 1,000,000 iterations. The MCMC analyses were run twice independently to ensure convergence. The phylogenetic tree was visualized using the R package GGTREE[98].

To obtain a comprehensive view of the origin of mangrove species, we further constructed a phylogeny, covering >90% of mangrove species, using a hierarchical analysis based on the genome-scale phylogenetic tree (backbone tree). In this hierarchical approach, the subsections of the backbone tree were reconstructed separately (Supplementary Fig. 2). First, a phylogenetic tree for each subsection was inferred on the basis of sequences of the transcriptome, chloroplast genome or DNA fragments (Supplementary Table 7) by RAxML-NG with the GTR + GAMMA + I model[95]. The corresponding node ages from the backbone tree were fixed to estimate divergence time in each subsection with r8s v.1.81 software[99]. According to the corresponding node ages, all the subsection phylogenies reconstructed were integrated into a single phylogeny. Finally, we constructed a comprehensive phylogeny containing 61 mangrove species, 23 mangrove associates and 64 relatives.

To further detect independent origination events of mangrove species, we first summarized the taxonomic status and distribution of all mangrove species on the basis of previous studies[14,22] and the global biodiversity information facility (GBIF) database (https://www.gbif.org) (Supplementary Table 5). The origination events among all mangrove species were identified on the basis of the well-known phylogenetic (only topological) relationships. We then labelled the resulting 27 origination events in corresponding positions in the large molecular phylogeny, which is a mega-tree derived from two recently published plant phylogenies[100,101] and contains 74,531 species and all families of extant vascular plants.

Speciation rate analysis. We compared lineages through time between dicot mangroves and 55,018 dicot plants[31] using count_lineages_through_time in the castor package[102]. We used fit_hbd_psr_on_grid in the castor package to estimate the effective speciation rate for dicot mangroves and other dicot plants in different time intervals. The overall effective speciation rate of mangroves was estimated as 0.037 Myr^{-1} after correcting the origin times of 21 lineages, as displayed in Fig. 2b.

Heterozygosity estimation. We estimated genome-wide heterozygosity across the genomes of 72 plant species and obtained the heterozygosity of three plant species (*Andrographis paniculata*[103], *Nymphaea colorata*[104] and *Oryza sativa*[105]) from previous studies. We first aligned clean short-insert paired-end reads of 47 species (including 46 plant species (Table 1) sequenced in this study and *Kandelia obovata* sequenced in ref. [30]) and 25 species downloaded from other available studies (Supplementary Table 14) to the assembled genomes using Burrows–Wheeler Aligner (BWA)[106]. Then, potential PCR-duplicated, single-end mapped and improperly paired reads in alignments were removed using SAMtools v.1.9 (ref. [107]) and Picard v.2.18.20 (http://broadinstitute.github.io/picard). Reads encompassing indels were realigned to increase the indel call accuracy using Genome Analysis Toolkit (GATK) v.3.7 (refs. [108,109]). To ensure the accuracy of SNP calling, only sites with adequate sequencing depth were identified as effective sites and sites with minor variant allele frequency >0.15 were called heterozygous sites. Heterozygosity was estimated as the number of identified heterozygous sites divided by the total number of effective sites.

Demographic history reconstruction. We inferred the demographic history for 31 species of mangroves (30 species of mangroves sequenced in this study plus *K. obovata* sequenced in ref. [30]), 10 mangrove associates (all sequenced in this study) and 31 mangrove relatives (6 sequenced in this study and 25 sequenced in previous studies). Details are given in Supplementary Table 14. We first generated the diploid consensus genome sequence of each species on the basis of high-quality sites. We then inferred population size histories applying the pairwise sequentially Markovian coalescent (PSMC)[19] with the parameter '-p 4+25×2+4+6' as well as optimized generation times and mutation rates (Supplementary Table 14). In Supplementary Figs. 3–5, the estimated population sizes were displayed in the background of SLCs[17,110]. The different generation times of different plant species were set according to their features (woody plants, 20 yr; shrubs and vines, 5 yr; perennial herbs, 2 yr; annual herbs, 1 yr). The mutation rate for each genus was estimated on the basis of the phylogenetic analyses.

For each genus with multiple whole-genome sequenced species (*Rhizophora*, *Ceriops*, *Kandelia*, *Bruguiera*, *Xylocarpus*, *Sonneratia*, *Avicennia*), we first estimated the mutation rate of CDS region (μ_{exon}) with the equation μ_{exon} = branch_length/divergence_time. Then we approximately inferred μ_{intron} with the equation $\mu_{intron}/\mu_{exon} = K_{intron}/K_{exon}$ and $\mu_{intergenic}$ with the equation $\mu_{intergenic}/\mu_{exon} = K_{intergenic}/K_{exon}$, where K is the number of substitutions per site between two sequences from sibling species and calculated from the whole-genome alignment of the two species generated by Mummer v.4.0.0 (ref. [111]). We calculated the μ of the whole genome with equation $\mu_{total} = (\mu_{exon} \times Length_{exon} + \mu_{intron} \times Length_{intron} + \mu_{intergenic} \times Length_{intergenic})/(Length_{exon} + Length_{intron} + Length_{intergenic})$. Moreover, we used the average mutation rate for each genus, respectively.

For the other genera that lack of congeneric species pairs, we estimated the μ'_{total} of the whole genome by multiplying the μ'_{exon} with a coefficient α, $\mu'_{total} = \mu'_{exon} \times \alpha$. The coefficient α can be approximately estimated by the nearby branch that used the method described above, $\alpha = (\mu_{exon} \times Length_{exon} + \mu_{intron} \times Length_{intron} + \mu_{intergenic} \times Length_{intergenic})/(Length_{exon} + Length_{intron} + Length_{intergenic})/\mu_{exon}$.

Population demographic history may also be inferred from genomic sequences that incorporate polymorphism data. We first aligned reads from 15 individuals of *S. alba* to the reference genome and discarded potential PCR-duplicated, single-end mapped, improperly paired reads as described in the subsection Heterozygosity estimation of the Methods. We then used mpileup in BCFtools with an option '-Q 20 -a DP, AD' and call in BCFtools with an option '-cvO' for variant calling[112]. Multi-allelic sites and indels were excluded using VCFtools v.0.1.17 (ref. [113]) with an option '--remove-indels --min-alleles 2 --max-alleles 2'. To infer the ancestral state of *S. alba* polymorphisms, we mapped reads of congeneric species *S. apetala* against the *S. alba* genome and generated genotype calls. We then obtained the unfolded SNP frequency spectrum (SFS) of *S. alba*. We also used the same pipeline to generate the unfolded SFS of *A. marina*. The reads of *A. alba* were used to infer the ancestral state of *A. marina* polymorphisms. We applied Stairway Plot v.2 software[40,41] with the unfolded SFS using 200 bootstraps to infer the historical changes in N_e over time in *A. marina* and *S. alba*, respectively.

We also estimated the correlation of A/S ratios with N_e. The standard PSMC result (using parameter '-p 4+25*2+4+6') includes 28 time intervals (T_1 to T_{28}) and the population size N_i for each time interval (N_1 to N_{28}). We also used option -d in PSMC to infer the specific time interval of the TMRCA (time to the most recent common ancestor) for each 100 bp segment. Then we combined the segments with the same TMRCA T_i, denoted as G_i. The effective population size N_e of G_i can be estimated as the weighted harmonic mean of N_2 to N_i (N_1 was discarded since it may be not accurately estimated by PSMC). To calculate A/S ratios, we used SnpEff[114] to annotate synonymous and non-synonymous heterozygous sites on the basis of the reference genome assembly. G_1, G_{28} and G_i with less than five synonymous heterozygous sites were discarded.

Simulations of population subdivision. To determine the impact of population subdivision on the results of PSMC, we designed and simulated several demographic models with different numbers of subpopulations and different population sizes. We assumed a large population with 10,000 individuals up to 22 ka. The generation time was set to 20 yr. The population decreased to several very small subpopulations that were completely isolated until 8 ka, when all subpopulations merged into one population. The newly merged population expanded exponentially to 10,000 individuals within 100 generations and then remained constant. We used the program ms[115] to generate simulated sequences and then used PSMC to infer the changes in historical population sizes. We used a mutation rate (μ) of 4.0×10^{-8} per bp per generation and a recombination rate of $\rho = \theta/2$ per bp per generation. For each model, 200 sequences of 1 Mb in length were simulated with 100 replications.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The sequences of this study have been deposited in National Genomics Data Center (NGDC), China National Center for Bioinformation. The 48 genome assembly sequences have been deposited in the Genome Warehouse (https://ngdc.cncb.ac.cn/gwh) in NGDC, under BioProject ID PRJCA004930. Detailed accession numbers of assemblies are listed in Supplementary Table 1. RNA-seq data from multiple species have been deposited in the Genome Sequence Archive (https://ngdc.cncb.ac.cn/gsa) in NGDC, under accession number CRA004363 with BioProject ID PRJCA005451. The genomic and RNA sequences were also deposited in National Center for Biotechnology Information under BioProject ID PRJNA817364.

Received: 12 October 2021; Accepted: 22 March 2022;
Published online: 28 April 2022

References

1. He, Z. et al. Speciation with gene flow via cycles of isolation and migration: insights from multiple mangrove taxa. *Natl Sci. Rev.* **6**, 275–288 (2019).
2. Zhou, R. et al. Population genetics of speciation in nonmodel organisms: I. Ancestral polymorphism in mangroves. *Mol. Biol. Evol.* **24**, 2746–2754 (2007).
3. Xu, S. et al. Genome-wide convergence during evolution of mangroves from woody plants. *Mol. Biol. Evol.* **34**, 1008–1015 (2017).
4. He, Z. et al. Convergent adaptation of the genomes of woody plants at the land–sea interface. *Natl Sci. Rev.* **7**, 978–993 (2020).
5. Lyu, H., He, Z., Wu, C.-I. & Shi, S. Convergent adaptive evolution in marginal environments: unloading transposable elements as a common strategy among mangrove genomes. *New Phytol.* **217**, 428–438 (2018).
6. Xu, S. et al. The origin, diversification and adaptation of a major mangrove clade (Rhizophoreae) revealed by whole-genome sequencing. *Natl Sci. Rev.* **4**, 721–734 (2017).
7. Feng, X. et al. Molecular adaptation to salinity fluctuation in tropical intertidal environments of a mangrove tree *Sonneratia alba*. *BMC Plant Biol.* **20**, 178 (2020).
8. Feng, X. et al. Genomic insights into molecular adaptation to intertidal environments in the mangrove *Aegiceras corniculatum*. *New Phytol.* **231**, 2346–2358 (2021).
9. Angelini, C. et al. A keystone mutualism underpins resilience of a coastal ecosystem to drought. *Nat. Commun.* **7**, 12473 (2016).
10. Atwood, T. B. et al. Global patterns in mangrove soil carbon stocks and losses. *Nat. Clim. Change* **7**, 523–528 (2017).
11. Barbier, E. B. et al. Coastal ecosystem-based management with nonlinear ecological functions and values. *Science* **319**, 321–323 (2008).
12. Barbier, E. B. et al. The value of estuarine and coastal ecosystem services. *Ecol. Monogr.* **81**, 169–193 (2011).
13. Hensel, M. J. S. & Silliman, B. R. Consumer diversity across kingdoms supports multiple functions in a coastal ecosystem. *Proc. Natl Acad. Sci. USA* **110**, 20621–20626 (2013).
14. Tomlinson, P. B. *The Botany of Mangroves* 2nd edn (Cambridge Univ. Press, 2016).
15. Rovai, A. S. et al. Global controls on carbon storage in mangrove soils. *Nat. Clim. Change* **8**, 534–538 (2018).
16. Alongi, D. M. Carbon sequestration in mangrove forests. *Carbon Manag.* **3**, 313–322 (2012).
17. Grant, K. M. et al. Sea-level variability over five glacial cycles. *Nat. Commun.* **5**, 5076 (2014).
18. Guo, Z. et al. Extremely low genetic diversity across mangrove taxa reflects past sea level changes and hints at poor future responses. *Glob. Change Biol.* **24**, 1741–1748 (2018).
19. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).

20. Sollars, E. S. A. et al. Genome sequence and genetic diversity of European ash trees. *Nature* **541**, 212–216 (2017).
21. Zhao, S. et al. Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nat. Genet.* **45**, 67–71 (2013).
22. Duke, N. C. in *Mangrove Ecosystems: A Global Biogeographic Perspective* (eds Rivera-Monroy, V. H. et al.) 17–53 (Springer, 2017).
23. Ellison, A. M., Farnsworth, E. J. & Merkt, R. E. Origins of mangrove ecosystems and the mangrove biodiversity anomaly. *Glob. Ecol. Biogeogr.* **8**, 95–115 (1999).
24. Gee, C. T. The mangrove palm *Nypa* in the geologic past of the new world. *Wetl. Ecol. Manag.* **9**, 181–203 (2001).
25. Germeraad, J. H., Hopping, C. A. & Muller, J. Palynology of tertiary sediments from tropical areas. *Rev. Palaeobot. Palynol.* **6**, 189–348 (1968).
26. Graham, A. Paleobotanical evidence and molecular data in reconstructing the historical phytogeography of Rhizophoraceae. *Ann. Missouri Bot. Gard.* **93**, 325–334 (2006).
27. Mazer, S. J. & Tiffney, B. H. Fruits of *Wetherellia* and *Palaeowetherellia* (?Euphorbiaceae) from Eocene sediments in Virginia and Maryland. *Brittonia* **34**, 300–333 (1982).
28. Muller, J. Fossil pollen records of extant angiosperms. *Bot. Rev.* **47**, 1–142 (1981).
29. Srivastava, J. & Prasad, V. Evolution and paleobiogeography of mangroves. *Mar. Ecol.* **40**, e12571 (2019).
30. Hu, M.-J. et al. Chromosome-scale assembly of the *Kandelia obovata* genome. *Hortic. Res.* **7**, 75 (2020).
31. Jin, Y. & Qian, H. V.PhyloMaker: an R package that can generate very large phylogenies for vascular plants. *Ecography* **42**, 1353–1359 (2019).
32. Zachos, J. C., Dickens, G. R. & Zeebe, R. E. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* **451**, 279–283 (2008).
33. Handley, L., Crouch, E. M. & Pancost, R. D. A New Zealand record of sea level rise and environmental change during the Paleocene–Eocene Thermal Maximum. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **305**, 185–200 (2011).
34. Louca, S. & Pennell, M. W. Extant timetrees are consistent with a myriad of diversification histories. *Nature* **580**, 502–505 (2020).
35. Saintilan, N. et al. Thresholds of mangrove survival under rapid sea level rise. *Science* **368**, 1118–1121 (2020).
36. Lu, J. & Wu, C.-I. Weak selection revealed by the whole-genome comparison of the X chromosome and autosomes of human and chimpanzee. *Proc. Natl Acad. Sci. USA* **102**, 4063–4067 (2005).
37. Lynch, M. et al. Perspective: spontaneous deleterious mutation. *Evolution* **53**, 645–663 (1999).
38. Ohta, T. Slightly deleterious mutant substitutions in evolution. *Nature* **246**, 96–98 (1973).
39. Ohta, T. The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* **23**, 263–286 (1992).
40. Liu, X. & Fu, Y. X. Exploring population size changes using SNP frequency spectra. *Nat. Genet.* **47**, 555–559 (2015).
41. Liu, X. & Fu, Y.-X. Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biol.* **21**, 280 (2020).
42. Krauss, K. W. et al. How mangrove forests adjust to rising sea level. *New Phytol.* **202**, 19–34 (2014).
43. Lovelock, C. E. et al. The vulnerability of Indo-Pacific mangrove forests to sea-level rise. *Nature* **526**, 559–563 (2015).
44. Frederiksen, N. O. *Review of Early Tertiary Sporomorph Paleoecology* (American Association of Stratigraphic Palynologists Foundation, 1985).
45. Smith, D. E., Harrison, S., Firth, C. R. & Jordan, J. T. The early Holocene sea level rise. *Quat. Sci. Rev.* **30**, 1846–1860 (2011).
46. Bouillon, S. et al. Mangrove production and carbon sinks: a revision of global budget estimates. *Glob. Biogeochem. Cycles* **22**, GB2013 (2008).
47. Donato, D. C. et al. Mangroves among the most carbon-rich forests in the tropics. *Nat. Geosci.* **4**, 293–297 (2011).
48. Hamilton, S. E. & Friess, D. A. Global carbon stocks and potential emissions due to mangrove deforestation from 2000 to 2012. *Nat. Clim. Change* **8**, 240–244 (2018).
49. Hutchison, J., Manica, A., Swetnam, R., Balmford, A. & Spalding, M. Predicting global patterns in mangrove forest biomass. *Conserv. Lett.* **7**, 233–240 (2014).
50. Ouyang, X. & Lee, S. Y. Improved estimates on global carbon stock and carbon pools in tidal wetlands. *Nat. Commun.* **11**, 317 (2020).
51. Bauer, J. E. et al. The changing carbon cycle of the coastal ocean. *Nature* **504**, 61–70 (2013).
52. Richards, D. R., Thompson, B. S. & Wijedasa, L. Quantifying net loss of global mangrove carbon stocks from 20 years of land cover change. *Nat. Commun.* **11**, 4260 (2020).
53. Sanders, C. J. et al. Are global mangrove carbon stocks driven by rainfall? *J. Geophys. Res. Biogeosci.* **121**, 2600–2609 (2016).
54. Alongi, D. M. Carbon cycling and storage in mangrove forests. *Ann. Rev. Mar. Sci.* **6**, 195–219 (2014).
55. Valiela, I., Bowen, J. L. & York, J. K. Mangrove forests: one of the world's threatened major tropical environments. *Bioscience* **51**, 807–815 (2001).
56. Doyle, J. J. & Doyle, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**, 11–15 (1987).
57. Yang, G., Zhou, R., Tang, T. & Shi, S. Simple and efficient isolation of high-quality total RNA from *Hibiscus tiliaceus*, a mangrove associate and its relatives. *Prep. Biochem. Biotechnol.* **38**, 257–264 (2008).
58. Wang, O. et al. Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome Res.* **29**, 798–808 (2019).
59. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764–770 (2011).
60. Liu, B. et al. Estimation of genomic characteristics by analyzing *k*-mer frequency in de novo genome projects. Preprint at https://arxiv.org/abs/1308.2012v2 (2013).
61. Vurture, G. W. et al. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
62. Chin, C.-S. et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **13**, 1050–1054 (2016).
63. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* **17**, 155–158 (2020).
64. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
65. Xiao, C.-L. et al. MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *Nat. Methods* **14**, 1072–1074 (2017).
66. Chin, C.-S. et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569 (2013).
67. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).
68. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
69. Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of diploid genome sequences. *Genome Res.* **27**, 757–767 (2017).
70. Luo, R. et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **4**, 30 (2015).
71. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
72. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
73. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
74. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
75. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
76. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **25**, 4.10.1–4.10.14 (2009).
77. Flynn, J. M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl Acad. Sci. USA* **117**, 9451–9457 (2020).
78. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
79. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
80. Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
81. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).
82. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
83. Birney, E. Genewise and genomewise. *Genome Res.* **14**, 988–995 (2004).
84. Kent, W. J. BLAT—The BLAST-Like Alignment Tool. *Genome Res.* **12**, 656–664 (2002).
85. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
86. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

87. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
88. Cantarel, B. L. et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2007).
89. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, R7 (2008).
90. Seppey, M., Manni, M. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol. Biol.* **1962**, 227–245 (2019).
91. Katoh, K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
92. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
93. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–W612 (2006).
94. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
95. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
96. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
97. Reis, M. Dos & Yang, Z. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Mol. Biol. Evol.* **28**, 2161–2172 (2011).
98. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. GGTREE: an package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
99. Sanderson, M. J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302 (2003).
100. Smith, S. A. & Brown, J. W. Constructing a broadly inclusive seed plant phylogeny. *Am. J. Bot.* **105**, 302–314 (2018).
101. Zanne, A. E. et al. Three keys to the radiation of angiosperms into freezing environments. *Nature* **506**, 89–92 (2014).
102. Louca, S. & Doebeli, M. Efficient comparative phylogenetics on large trees. *Bioinformatics* **34**, 1053–1055 (2018).
103. Liang, Y. et al. Chromosome level genome assembly of *Andrographis paniculata*. *Front. Genet.* **11**, 701 (2020).
104. Zhang, L. et al. The water lily genome and the early evolution of flowering plants. *Nature* **577**, 79–84 (2020).
105. Huang, X. et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **42**, 961–967 (2010).
106. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
107. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
108. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
109. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
110. Miller, K. G. et al. The Phanerozoic record of global sea-level change. *Science* **310**, 1293–1298 (2005).
111. Marçais, G. et al. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput. Biol.* **14**, e1005944 (2018).
112. Narasimhan, V. et al. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* **32**, 1749–1751 (2016).
113. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
114. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**, 80–92 (2012).
115. Hudson, R. R. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337–338 (2002).

Acknowledgements

We thank P. H. Raven, X. He, W. Wang, Q. Qiu, J. Lu, J. Liu, K. Zeng, T. Tang and K. Huang for insightful comments. We also thank J. W. H. Yong, L. Chen, S. Jian, Y. Zhang, W. Lun Ng, M. Tracy, Y. Chen and S. Li for sample collection and Y. Sun and J. Ruan for technical support. This project was supported by the National Natural Science Foundation of China (grant nos. 31830005 to S. Shi and 31971540 to Z.H.); the National Key Research and Development Plan (grant no. 2017FY100705 to S. Shi); the Guangdong Basic and Applied Basic Research Foundation (grant no. 2019A1515010752 to Z.H.); the Science and Technology Project of Guangzhou (grant no. 202102020483 to Z.H.); and the Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (grant no. 311021006 to S. Shi).

Author contributions

S. Shi conceived of the study, while Z.H., C.-I.W. and S. Shi designed and conceptualized it. S. Shi, Z.H., X.F., Q.C., S.L., Z.G., M.L., S. Shao, X.M., W.X., X.W., R.Z., G.L., W.W., C.Z. and N.C.D. collected samples. X.F., S.L., M.L., S. Shao, W.W. and Z.Z. performed the experiments. Z.H., X.F., L.L., S.L., K.H., C.S., X.L. and G.F. sequenced and assembled the genomes. Z.H., X.F. and Q.C. performed the data analysis with assistance from S.L., J.W., S.X., S. Shao and W.W. Z.H., C.-I.W., R.E.R and S. Shi made the data interpretation and presentation. Z.H., C.-I.W. and S. Shi wrote the manuscript with input from X.F., Q.C. and Z.G. D.E.B and R.E.R revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41559-022-01744-9.

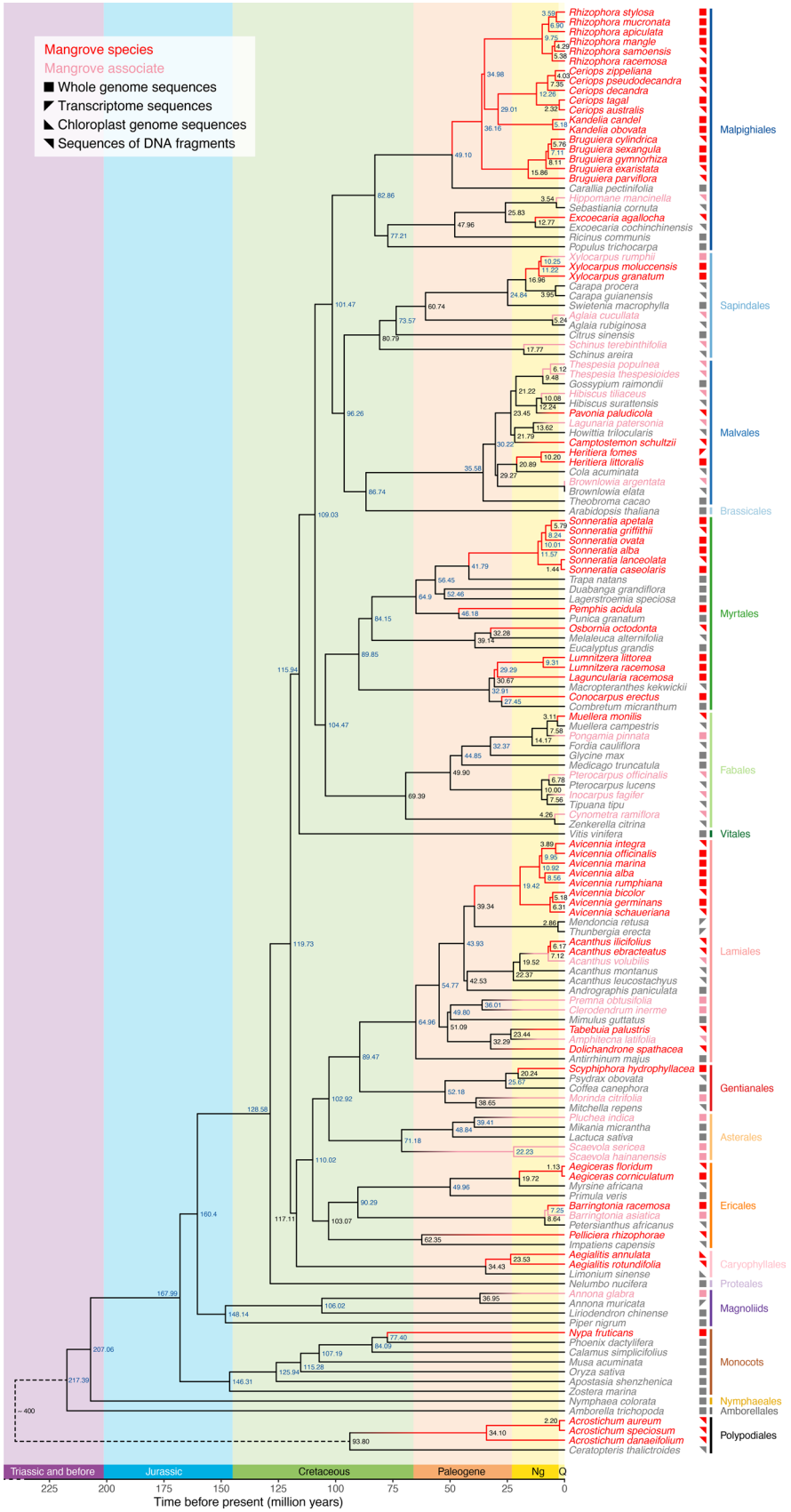
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-022-01744-9.

Correspondence and requests for materials should be addressed to Suhua Shi.

Peer review information *Nature Ecology & Evolution* thanks Richard Buggs, Abraham Morales-Cruz and Maheshi Dassanayake for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Extended Data Fig. 1 | Phylogenetic tree of 61 species of mangroves, 23 mangrove associates, and 64 relatives of mangroves. The types of sequences used to generate the phylogenetic tree are denoted to the right of the name of each species.

nature portfolio

Corresponding author(s): Suhua Shi

Last updated by author(s): Mar 18, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	PacBio SMRT reads were generated on the PacBio sequel II platform; Linked reads of the 10X Genomics library were generated on the BGISEQ-500 platform; Single tube long fragment reads (stLFRs) were generated on the BGISEQ-500 platform; Hi-C data and paired-end short-reads were generated on the Illumina or BGISEQ-500 platform.
Data analysis	Genome survey and assembly: jellyfish (v2.3.0), GCE (v1.0.2), GenomeScope, FALCON (v2.2.4), wtdbg2 (v2.5), hifiasm (v0.15.2), MECAT2 (v20190314), Quiver (SMRT Analysis v2.3.0), Racon (v.1.3.1), Pilon (v1.22), Supernova (v2.0.1), GapCloser (v1.12), HiC-Pro (v3.0.0), Juicer, Juicebox, 3D-DNA; Genome annotation and assessment: RepeatMasker (open-4.0.9), RepeatModeler, LTR_FINDER (v1.0.7), Tandem Repeat Finder (TRF), Augustus (v3.3.1), GENSCAN, GlimmerHMM, GeneWise, BLAT, Bowtie2 (v2.3.4.3), Tophat2 (v2.1.1), Cufflinks (v2.2.1), Maker (v3), EVidenceModeler, BUSCO (v4); Phylogenomic analysis: MAFFT (v7.429), PAL2NAL (v14.0), Gblocks (v0.91b), RAxML-NG (v0.9.0), PAML (v4.9j), GGTREE (v2.2.4), r8s (v1.81), castor; Population analysis: BWA (v0.7.17-r1188), SAMtools (v1.9), Picard (v2.18.20), GATK (v3.7), PSMC, BCFtools (v1.9), VCFtools (v0.1.17), Stairway Plot (v2), SnpEff (v4.3t), ms.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The sequences of this study have been deposited in National Genomics Data Center (NGDC), China National Center for Bioinformation. The 48 genome assembly sequences have been deposited in the Genome Warehouse (GWH, https://ngdc.cncb.ac.cn/gwh) in NGDC, under BioProject ID PRJCA004930. Detailed accession numbers of assemblies are listed in Supplementary Table 1. RNA-seq data from multiple species have been deposited in the Genome Sequence Archive (GSA, https://ngdc.cncb.ac.cn/gsa) in NGDC, under accession number CRA004363 with BioProject ID PRJCA005451. The genomic and RNA sequences were also deposited in National Center for Biotechnology Information (NCBI) under BioProject ID PRJNA817364.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We de novo sequenced the whole-genomes of 32 mangroves, which we combined with other sequences of 30 additional species, comprising almost all mangroves globally. We also de novo sequenced the whole-genomes of 10 mangrove associates and 6 relatives. Based on these genomic and genetic data, we inferred the origins of mangroves, examined the effects of past sea level changes, and attempted to forecast what may happen in the coming century. These community genomic data will be valuable for ecology, evolution, and biodiversity research.
Research sample	A total of 48 plant genomes were sequenced in this project, including 32 mangrove species (counting three subspecies of Avicennia marina), 10 mangrove associates, and 6 relatives. Among the 48 species, five genome assemblies (Rhizophora apiculata, Bruguiera gymnorhiza, Sonneratia alba, Sonneratia caseolaris, Avicennia marina subsp. marina) were improved to chromosome-level upon the previous version published by our laboratory group. All tissue samples were collected from multiple sources, listed in Supplementary Tables 9-10. For population genomic data, 15 individuals of Sonneratia alba from Ngao, Ranong, Thailand, and 12 individuals of Avicennia marina from Qinglan Harbor, Wenchang, China, were collected.
Sampling strategy	For de novo genome data, fresh and healthy tissues were harvested and immediately frozen in liquid nitrogen, followed by preservation at -80°C in the laboratory before DNA or RNA extraction, with two exceptions. Fresh and healthy leaves of Xylocarpus rumphii and Avicennia officinalis were dried and preserved with silica gel. For population data, fresh and healthy leaves were dried and preserved with silica gel. The sample size of Sonneratia alba and Avicennia marina were 15 and 12, respectively.
Data collection	Plant samples information has been deposited in the BioSample database in National Genomics Data Center. High-throughput sequencing data were generated on different platforms, including PacBio, Illumina, and BGISEQ.
Timing and spatial scale	This is not relevant to our study.
Data exclusions	No data were excluded.
Reproducibility	The read coverages for each genome are from dozens to hundreds.
Randomization	The mature plant samples for whole genome sequencing were randomly sampled with permission.
Blinding	The investigators were blinded to group allocation during the data collection.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging