the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2022) **111,** 1411–1424 doi: 10.1111/tpj.15899

Adaptation to a new environment with pre-adaptive genomic features – Evidence from woody plants colonizing the land–sea interface

Zixiao Guo[1,*], Shaohua Xu[1], Wei Xie[1], Shao Shao[1], Xiao Feng[1], Ziwen He[1], Cairong Zhong[2], Kaichi Huang[3], Chung-I Wu[1] and Suhua Shi[1,*]

[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, 510275, Guangzhou, Guangdong, China,*
[2]*Hainan Academy of Forestry (Hainan Academy of Mangrove), 571199, Haikou, Hainan, China, and*
[3]*Department of Botany and Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, Canada*

Received 25 June 2021; revised 12 June 2022; accepted 5 July 2022; published online 7 July 2022.
*For correspondence (e-mail lssssh@mail.sysu.edu.cn [Suhua Shi]; guozx8@mail.sysu.edu.cn [Zixiao Guo]).

SUMMARY

Adaptation to new environments is a key evolutionary process which presumably involves complex genomic changes. Mangroves, a collection of approximately 80 woody plants that have independently invaded intertidal zones >20 times, are ideal for studying this process. We assembled near-chromosome-scale genomes of three *Xylocarpus* species as well as an outgroup species using single-molecule real-time sequencing. Phylogenomic analysis reveals two separate lineages, one with the mangrove *Xylocarpus granatum* and the other comprising a mangrove *Xylocarpus moluccensis* and a terrestrial *Xylocarpus rumphii*. In conjunction with previous studies, we identified several genomic features associated with mangroves: (i) signals of positive selection in genes related to salt tolerance and root development; (ii) genome-wide elevated ratios of non-synonymous to synonymous substitution relative to terrestrial relatives; and (iii) active elimination of long terminal repeats. These features are found in the terrestrial *X. rumphii* in addition to the two mangroves. These genomic features, not being strictly mangrove-specific, are hence considered pre-adaptive. We infer that the coastal but non-intertidal habitat of *X. rumphii* may have predisposed the common ancestor to invasion of true mangrove habitats. Other features including the preferential retention of duplicated genes and intolerance to pseudogenization are not found in *X. rumphii* and are likely true adaptive features in mangroves. In conclusion, by studying adaptive shift and partial shifts among closely related species, we set up a framework to study genomic features that are acquired at different stages of the pre-adaptation and adaptation to new environments.

Keywords: comparative genomics, intertidal zone, mangrove, new environment, pre-adaptation, *Xylocarpus*.

INTRODUCTION

Adaptation to a new environment is a most challenging question in biological evolution. It is usually a long and complex process involving multiple genetic changes. One important yet unresolved question is how such a complex process can be completed. This is akin to the blind watchmaker scenario whereby each step of the complex process has to be adaptive (Dawkins, 1996; Wu, Wen, et al., 2021). We shall now use mangroves' adaptation to the tropical intertidal zones as a window to investigate the adaptation to new environments.

Mangroves are woody plants inhabiting intertidal zones. These environments are at the interface of land and sea and, therefore, are extremely harsh for woody plants, with high salinity, daily fluctuating tides, strong ultra-violet (UV) light, and muddy anaerobic soils (Tomlinson, 2016). Although different from seagrasses, the angiosperms that have adapted to the submarine environments (Olsen et al., 2016), mangrove species have also adapted to many marine environmental conditions. Despite the wide geographical distribution of their habitats, there are fewer than 80 mangrove species scattered across the vascular plant

phylogeny (Duke, 2017). The general view is that it is extremely difficult for woody plants to invade these intertidal habitats. Nevertheless, the few that succeeded have been highly successful.

Many of the mangrove species have evolved highly specialized characters such as salt tolerance, aerial roots, and viviparous embryonic development (Tomlinson, 2016). Several mangrove lineages show convergent genomic features, including more frequent non-synonymous nucleotide substitutions in mangroves than in their relatives, elimination of transposable elements, and unusual amino acid usage and substitutions (Xu, He, Zhang, et al., 2017; Xu, He, Guo, et al., 2017; Lyu et al., 2018; He et al., 2020). A few other features appear to be mangrove-specific but the signals are not strong enough to be conclusive, such as positive selection in genes related to viviparity and the red bark (Xu, He, Zhang, et al., 2017). These patterns require extensive modifications of the genome, rather than a few point mutations in a specific site or gene. It is unclear how these genomic adaptations have occurred.

In this context, we wish to bring in the concept of pre-adaptation. Historically, pre-adaptation referred to evolution of structures that are not immediately beneficial, but would become useful in a new environment (Bock, 1958; Cadotte et al., 2018). In the modern literature, pre-adaptation has come to mean a process that involves stepwise changes that are useful in intermediate habitats that only partially resemble the eventual new environment (Dieterich & Sommer, 2009). Hence, extant species inhabiting these intermediate environments may appear 'pre-adapted' to a new habitat conquered by their relatives. By this narrower definition, pre-adaptive characters are mechanistically connected to the eventual adaptation, while according to the classical definition pre-adaptive traits simply happen to be adaptive by chance (see also Supplementary Note S1 in the supporting information).

In this study, we adopt the modern definition of pre-adaptation. Applied to mangrove evolution, this process would involve initial invasions of aquatic environments or coastal areas close to the intertidal zones. Adaptation during these intermediate steps would then be crucial for final adaptation to the true mangrove habitats. Traits beneficial in these intermediate environments would appear 'pre-adapted' to the intertidal habitats. Alternatively, mangrove species may have evolved directly from completely terrestrial ancestors, with the traits adapted to the intertidal lifestyle having evolved *de novo* directly from terrestrial plants. Previous studies could not distinguish between these scenarios because they compared monophyletic mangrove clades with terrestrial relatives. In those cases, all evolutionary changes are interpreted as adaptation (He et al., 2020; Lyu et al., 2018; Xu, He, Zhang, et al., 2017; Xu, He, Guo, et al., 2017; Xu et al., 2020). Testing these hypotheses would require multiple species of mangroves and non-mangroves within a clade that have independently colonized the new environment. The *Xylocarpus* genus uniquely provides such an opportunity.

The genus *Xylocarpus* comprises two mangrove species, *Xylocarpus granatum* and *Xylocarpus moluccensis*, and one non-mangrove species, *Xylocarpus rumphii*. The two mangroves occupy intertidal environments, whereas *X. rumphii* only occurs on cliffs, rocks, and sandy lands on coasts, beyond the high levels of sea tides (Figures 1b and S1). Furthermore, *X. granatum* and *X. moluccensis* have developed special root systems (Figure 1a), while *X. rumphii* has not (Duke, 2006; Duke, 2014). Most importantly, we have previously shown that *X. rumphii* is phylogenetically nested with *X. moluccensis*, with *X. granatum* representing the outer branch (Guo, Guo, et al., 2018). This unique phylogenetic pattern permits the identification of shared, potentially pre-adaptive, characters between mangroves and non-mangroves. We generated single-molecule real-time (SMRT) long reads and assembled high-quality genomes of the three *Xylocarpus* species, together with a close relative, *Swietenia macrophylla*. Genomic analyses were performed to address (i) whether an intermediate stage of pre-adaptation has been involved in mangroves' transition from the terrestrial to the mangrove lifestyle and (ii) what genomic changes underlie the lifestyle shift of the two mangroves.

RESULTS

I. Genome assemblies and phylogenomic analyses

Using PacBio SMRT sequencing technology, we obtained 77.27 Gb of long reads for *X. granatum* (resulting draft genome size of 300.0 Mb), 148.69 Gb for *X. moluccensis* (assembled size 280.9 Mb), 84.03 Gb for *X. rumphii* (assembled size 218.3 Mb), and 139.27 Gb for *S. macrophylla* (assembled size 210.8 Mb; we refer to them as 'the four species' hereafter; Tables 1, S1, and S2). All four genomes were assembled with high quality and completeness, as indicated by the low number of assembled contigs (Table 1), high N50 values (Table 1), high fractions of Illumina reads mapped to the assemblies (Langmead & Salzberg, 2012) (Table S3), and high rates of benchmarking universal single-copy orthologs mapped (Table S4). The assemblies of all three *Xylocarpus* species are in good synteny with that of *S. macrophylla* (Figure S2). We predicted 31 704, 30 130, 27 968, and 33 069 protein-coding genes in the four genomes. The total length of mRNA, the total length of coding sequence, and the exon length are comparable in the four species, but *X. rumphii* has a larger average number of exons per gene (6.14 relative to 5.33–5.53) and average mRNA length (Figure S3, Tables S5 and S6). Repeats accounted for 54.34%, 41.69%, 22.34%, and 27.98% of the four genomes (Table S7).

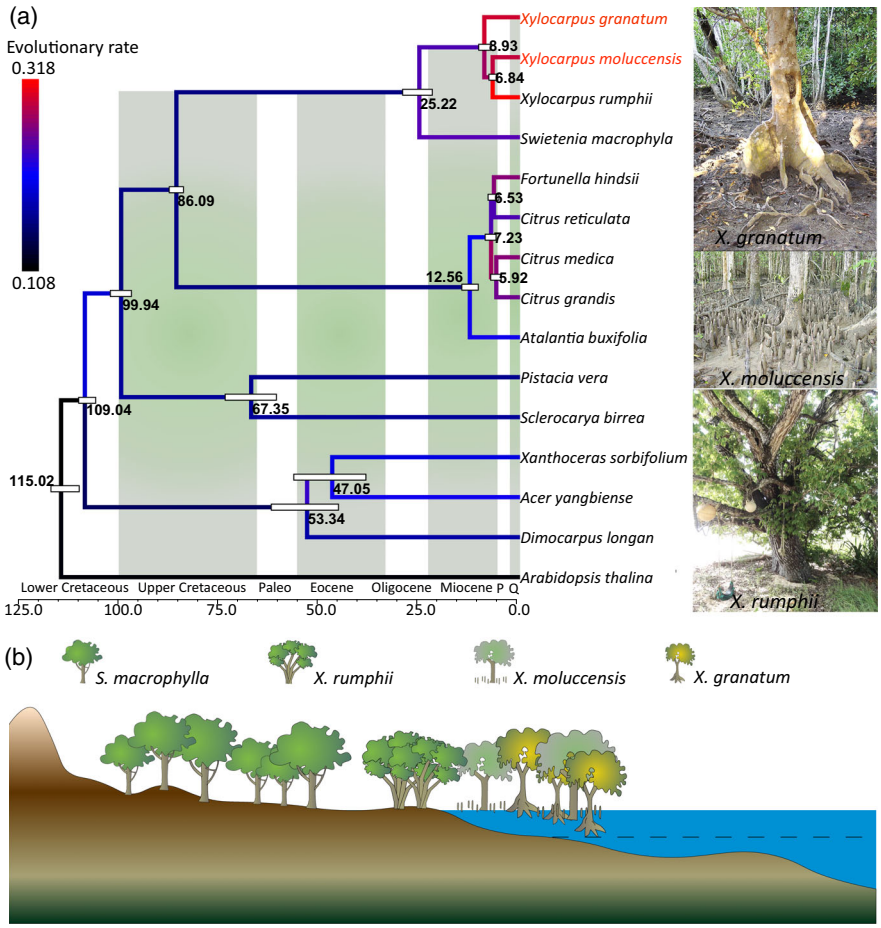

Figure 1. Phylogeny and habitat differentiation of *Xylocarpus* species.
(a) Dated phylogeny of 14 Sapindales species, with *Arabidopsis thaliana* as the outgroup. Numbers on nodes indicate node age, while white error bars are 95% posterior probability ranges. Branch color indicates the ratio of non-synonymous to synonymous amino acid substitutions (evolutionary rate). Species names in red indicate mangroves. Geological periods are indicated by the alternative background coloring. 'P' is short for 'Pleistocene' and 'Q' is short for 'Quaternary'. Photos show the root systems of the three *Xylocarpus* species on the right. Photos are from the Mangrove iD (Duke, 2014). (b) Habitats of *Xylocarpus* species and *Swietenia macrophylla* from inland to intertidal zones. The dashed line indicates low tide levels. The illustrations of trees in the bottom panel are contributed by the IAN Image Library (https://ian.umces.edu/media-library/).

Table 1 Features of the assembled genomes

Statistics	*X. granatum*	*X. moluccensis*	*X. rumphii*	*S. macrophylla*
Assembled genome size (bp)	300 049 238	280 868 920	218 264 657	270 800 498
Number of contigs	498	2982	1433	1364
Longest contig (bp)	9 335 078	6 078 191	6 097 399	7 877 045
Contig N50 (bp) (i-th contig)	3 428 067 (33)	1 396 120 (58)	2 631 781 (31)	2 418 160 (34)
Contig N90 (bp) (i-th contig)	328 928 (124)	29 173 (1064)	107 396 (145)	152 978 (157)
Shortest contig (bp)	4698	364	1058	456
GC content	38.5%	37.6%	36.2%	32.9%
Number of gene models	31 704	30 130	27 968	33 068
Number of pseudogenes	6282	6627	8873	7954
Retro-pseudogenes	962	917	1324	1131
Duplicated pseudogenes	1820	1933	2291	2380
Fragments of pseudogenes	3500	3777	5258	4443

Genomic phylogenies constructed using several methods consistently show that *Xylocarpus* is separated into two lineages: (i) the mangrove *X. granatum* and (ii) the other lineage consisting of the terrestrial species *X. rumphii* and the mangrove *X. moluccensis* (Figure 1a). We also constructed phylogenies for each of the 2582 groups of

orthologs using both maximum likelihood and Bayesian inference methods, and then we inferred consensus species trees from these gene trees. The consensus species trees given by both methods are exactly consistent with the previous phylogeny constructed from concatenated sequences (Figure 2a). When constructing gene trees for the orthologs identified for the four Meliaceae species, we found that 95.8% of these single-copy genes (2249) showed three of the 15 possible topologies (I, II, and III) with the three *Xylocarpus* species forming a monophyletic group (Figure 2b). In particular, 62.8% (1473) of the genes showed topology I, which is consistent with the species tree (Figure 2b). These findings strongly support *X. rumphii* grouping with *X. moluccensis*. We estimated that *Xylocarpus* diverged from *S. macrophylla* at approximately 25.22 million years ago (Mya) (95% highest posterior density [HPD]: 21.97–29.39), *X. granatum* diverged from the other two sister species at approximately 8.93 Mya (95% HPD: 7.42–10.35), and *X. moluccensis* diverged from *X. rumphii* at approximately 6.84 Mya (95% HPD: 5.63–7.80) (Figure 1a).

Having determined the phylogeny, we can now ask how the two mangroves became mangroves. They might have *de novo* adapted to the intertidal environment in parallel from an ancestor completely living on land without encountering intertidal environments (*de novo* adaptation model, Figure 3a). Under this model, the genomic features associated with intertidal adaptations should be private to *X. granatum* and *X. moluccensis*. Alternatively, before the final adaptation to intertidal environments, the ancestors might have evolved in an intermediate habitat such as coastal areas close to the intertidal zone or aquatic freshwater habitats (Figure 3a). These adaptations in the intermediate habitats might have pre-adapted the ancestor to invade the true intertidal environments. In this case, some genomic features associated with mangroves may be common among all three *Xylocarpus* species. We may distinguish the two hypotheses by uncovering genomic changes of the species complex.

II. Pre-adaptation indicated by genomic changes

Inhabiting a coastal environment usually adjacent to mangrove forests, *X. rumphii* does not show morphological characters typical for mangroves, such as the plank roots of *X. granatum* or pneumatophores of *X. moluccensis*

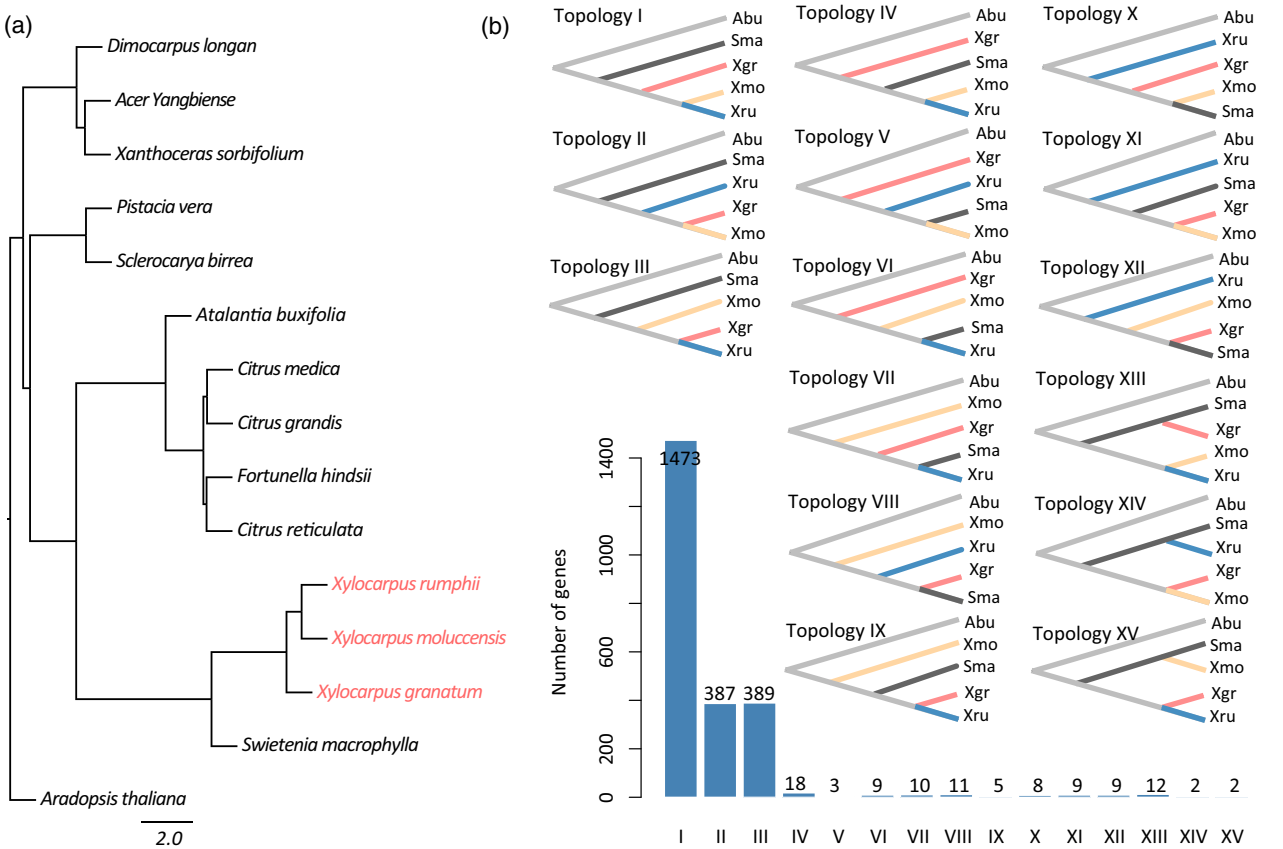

Figure 2. Phylogenomic analyses of *Xylocarpus* and their related species.
(a) Consensus species tree based on 2582 gene trees. (b) Fifteen possible topologies and the number of genes showing each topology. Abu, Sma, Xgr, Xmo, and Xru are short for *Atalantia buxifolia*, *Swietenia macrophylla*, *Xylocarpus granatum*, *Xylocarpus moluccensis*, and *Xylocarpus rumphii*, respectively.

(Figure 1a). It appears that these morphological characters have *de novo* originated in the two mangroves. However, some genomic features previously found to have convergently evolved in mangrove species suggest a different history.

High evolutionary rates. Mangrove adaptation is accompanied by elevated ratios of non-synonymous to synonymous amino acid substitutions (dN/dS) (Xu, He, Zhang, et al., 2017). We observed a dN/dS ratio that is higher in the whole *Xylocarpus* clade (0.291–0.318) than on other branches (0.108–0.267), likely related to the transition to an intertidal environment (Figures 1a and S4). Notably, the dN/dS ratio on the *X. rumphii* branch was even higher than that of the two mangroves (0.318 versus 0.291/0.299). The higher dN/dS values in the *Xylocarpus* clade than in other terrestrial species indicate that their genomes have preserved more radical amino acid substitutions, providing a molecular basis for functional or character innovation.

Active elimination of transposons. Plant genomes feature high proportions of transposable elements. In particular, long terminal repeat retrotransposons (LTR-RTs) are usually the most dominant. However, mangrove species typically carry a reduced transposable element load, which is considered a burden (Lyu et al., 2018). LTRs occupy 14.51% of the genome in *X. rumphii* and 15.73% in *S. macrophylla*, falling in the second-smallest quartile among angiosperms (Lyu et al., 2018). The LTR-RT proportion is 31.54% in *X. granatum* and 28.18% in *X. moluccensis*. We identified 4372, 1832, 796, and 1456 LTR-RTs with LTRs on both ends for the four species. Interestingly, *X. rumphii* and *S. macrophylla* carry fewer LTR-RTs than the other two taxa.

Solo LTRs or truncated LTR-RTs are usually relics of eliminated intact transposons. The rate of LTR elimination can thus be estimated by comparing the number of intact LTR-RTs to that of solo LTRs and truncated LTR-RTs. We identified 389, 222, 115, and 381 intact LTRs in the four species. All three *Xylocarpus* species show much higher solo LTR/intact LTR (S/I) ratios than *S. macrophylla* (5.00, 4.15, and 3.51 versus 0.79), indicating that elimination of LTRs was much more active in the *Xylocarpus* ancestors than in *S. macrophylla* (Figure 3c). Similarly, the ratios of truncated to intact LTR-RTs (T/I) are higher in *Xylocarpus* species (1.02, 1.73, and 2.21 versus 0.87) than in *S. macrophylla* (Figure 3d).

We identified many LTR-RTs as old as approximately 27–30 Mya preserved in the four genomes. However, *X. granatum* harbors an excess of LTR-RTs younger than 20 My (Figure 3b) compared to the other species in this study. Many of these *X. granatum* LTR-RTs are older than the divergence of *X. granatum* from the other two *Xylocarpus* taxa (approximately 9 Mya, Figure 1a). Hence, LTR-RTs have been actively inserted into the genome of the *Xylocarpus* ancestor. However, higher proportions of the inserted LTR-RTs have been eliminated from the genomes of *X. moluccensis* and *X. rumphii*, resulting in the absence of old LTR-RTs in these two species. In addition, an increase in LTR-RT number was observed in both *X. granatum* and *X. moluccensis* very recently (Figure 3b).

Many of the genes involved in salinity tolerance and root development are under positive selection. Mangrove species were shown to exhibit pan-exome modifications of amino acid usage and unusual amino acid substitutions (He et al., 2020; Xu, He, Zhang, et al., 2017). Among the 2582 orthologs identified in 15 species, 116 were determined to be under positive selection in *X. granatum*, 75 in *X. moluccensis*, and 146 in *X. rumphii* (Figure 3e). Genes positively selected in *X. granatum* are enriched for 10 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, including steroid biosynthesis, glycerophospholipid metabolism, and carotenoid biosynthesis (Table S8). Carotenoids function as antioxidants (Tomášek et al., 2016). Similarly, the four most enriched KEGG pathways of *X. moluccensis* include betalain biosynthesis (Table S8). Betalain is also an antioxidant (Kanner et al., 2001).

Mangrove species tolerate high salinity of intertidal zones. Genes likely involved in salt tolerance were found to be positively selected not only in the two *Xylocarpus* mangroves but also in *X. rumphii* (Tables S9–S11). Some examples are described below. Gene function annotations were extracted from TAIR (https://www.arabidopsis.org/index.jsp). *Site-1 Protease* (*S1P*) selected in *X. granatum* helps protect seedlings against salt and osmotic stress. The roots of the *s1p-3* mutant are hypersensitive to NaCl, KCl, LiCl, and mannitol. The *Tetratricopeptide-repeat Thioredoxin-Like 1* (*TTL1*) gene encodes one of the 36 carboxylate clamp (CC)-tetratricopeptide repeat proteins which are required for osmotic stress tolerance. The genes under selection in *X. moluccensis* also respond to salt stress (e.g., *CES1* and *S6K2*). Expression of the *Serine/Threonine Protein Kinase 2* (*S6K2*) gene, which encodes a ribosomal-protein S6 kinase, is induced by cold and salt. Notably, many genes positively selected in *X. rumphii* are involved in salt tolerance. The *Defense No Death 1* (*DND1*) gene encodes a mutated cyclic nucleotide-gated cation channel that conducts K^+ and other monovalent cations but excludes Na^+. The *Plant U-Box 19* (*PUB19*) gene is involved in salt inhibition of germination. *Salt Induced Zinc finger protein 1* (*SIZ1*) encodes a C2H2-type zinc finger protein that is involved in salt tolerance. *Salt Tolerance Zinc Finger* (*STZ*) increases salt tolerance in a manner partially dependent on ENA1/PMR2, a P-type ATPase required for Li^+ and Na^+ efflux. *Hal2-Like* (*HL*) also encodes a 3′-phosphoadenosine-5′-phosphate (PAP) phosphatase that is sensitive to physiological concentrations of Na^+. Similarly,

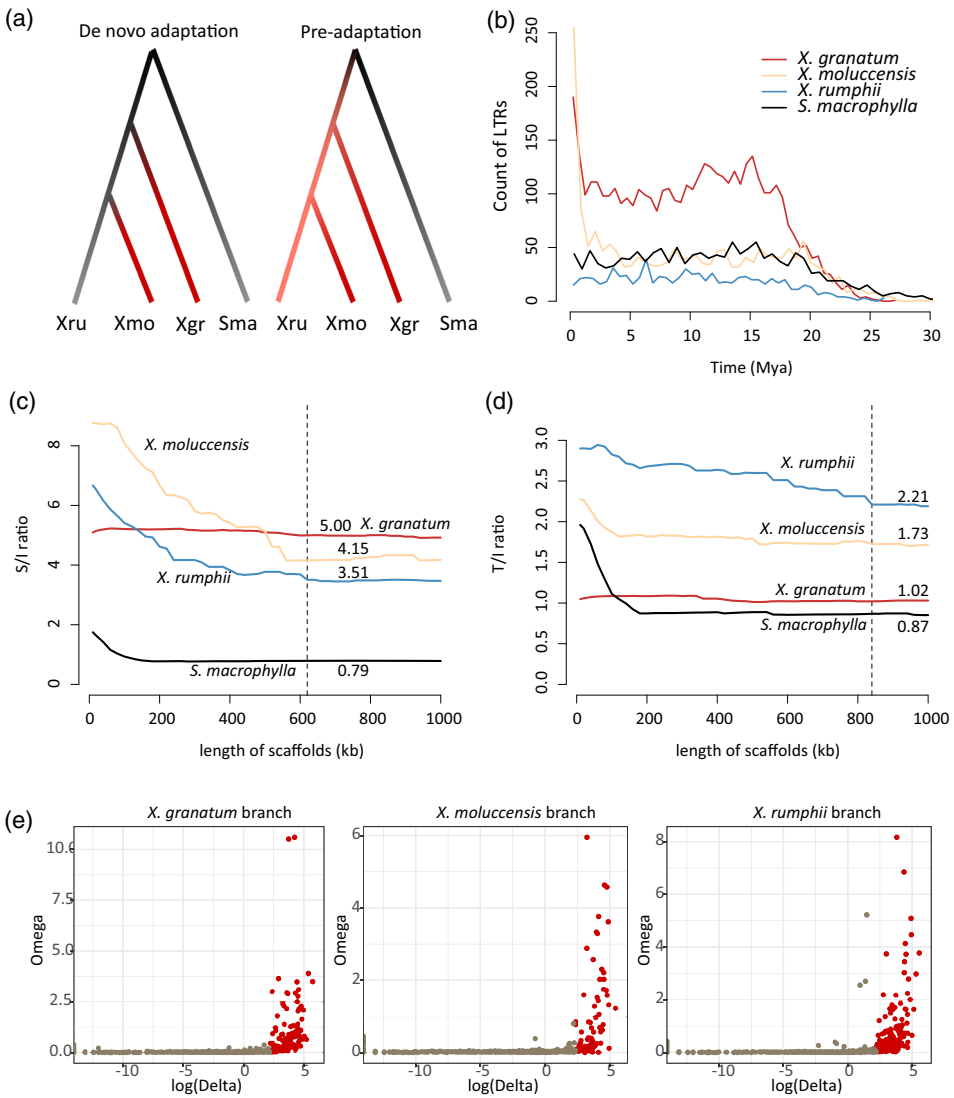

Figure 3. Genomic signals of pre-adaptation to mangrove habitats in the *Xylocarpus* ancestor.
(a) Schematic diagrams of the hypothetical *Xylocarpus* evolutionary histories. The shade from black to red indicates the transition from a terrestrial to a mangrove lifestyle, while orange represents an intermediate stage of pre-adaptation. (b) Age distribution of LTR insertions. (c) Ratios of solo LTR to intact LTR (S/I) of the four species with different scaffold size cut-offs. (d) Ratios of truncated LTR-RT to intact LTR (T/I) of the four species with different scaffold size cut-offs. (e) Genes under positive selection in each of the three *Xylocarpus* species. The *x*-axis shows log values of $2\times(L_1 - L_0)$, where L_1 and L_0 are the likelihoods of the alternative and the null model of the branch-site method. The *y*-axis shows the ω values.

seagrasses have regained full salinity tolerance, likely through the strong expression of a gene encoding a salt-tolerant H^+-ATPase in vegetative tissue (Olsen et al., 2016).

The three *Xylocarpus* species can be distinguished by their root systems. Plank roots help *X. granatum* trees settle on the hard mudflats of intertidal zones and pneumatophores help *X. moluccensis* exchange oxygen under tidal inundation. Genes likely involved in root development were also identified as adaptively evolving (Tables S9–S11) in all three *Xylocarpus* species. The *Arg1-Like 1* (*ARL1*) gene is positively selected in *X. granatum* and *X. rumphii*. It encodes a DnaJ-like protein similar to ARG1 and ARL2, which are involved in root and hypocotyl gravitropism responses. The *AGCVIII Kinase 1–12* (*AGC1-12*) gene under selection in *X. moluccensis* is also involved in first positive phototropism and gravitropism. Genes encoding MIZU-KUSSEI-like proteins (one gene in *X. moluccensis* and two in *X. rumphii*) are under positive selection and play a role in lateral root development by negatively regulating cytokinin sensitivity during root development and positively regulating hydrotropism in roots. A gene encoding shoot gravitropism 9, which modulates the interaction between statoliths and F-actin in gravity sensing, and a gene encoding a GRAS transcription factor, which regulates root and shoot development, are both adaptively evolving in *X. rumphii*.

III. Genomic evolution likely underlies true adaptation of *Xylocarpus* mangroves

The genomic changes featured in mangrove species described in the last section are common in all *Xylocarpus* species, regardless of their being mangroves or not. These genomic adaptations are interpreted as pre-adaptations since the ancestors had not been adapted to the true mangrove environment yet. In the following section, we attempt to identify genomic changes underlying the parallel adaptation to intertidal zones of the two mangroves in our study.

Preferential retention of duplicated genes involved in mangrove adaptation. Gene gain and loss play essential roles in plant environmental adaptations (Wang, Wang, & Paterson, 2012; Xu et al., 2021). Whole genome duplication (WGD) doubles all genes at once, with many genes lost during the subsequent re-diploidization process. These drastic gene copy number changes may result in adaptations to new environments. We examined the presence of WGD and the subsequent gene family evolution in *Xylocarpus* species.

By ascertaining collinear genomic blocks and analyzing the Ks distribution of genes that reside in them (Wang et al., 2010), we identified a WGD that occurred before *Xylocarpus* diverged from *Swietenia* (Figures 4a and S5–S9, Table S12). The WGD event was further dated at approximately 63.05 Mya (95% HPD: 54.93–73.48). The age of this WGD is close to that of the Cretaceous-Tertiary Extinction Event (K/T event) and that of the crown node of the Meliaceae family (Koenen et al., 2015). The WGD may have helped the Meliaceae ancestor survive in the K/T event and subsequently diversify.

Genes duplicated during a WGD can retain two, one, or no copies in their current genomes (Table S13). With *S. macrophylla* as control, we found that the two mangroves are similar to each other in their gene retention patterns and distinct from *X. rumphii* (Figure 4b). Particularly, the mangrove taxa have preferentially retained more genes with 2:2 and 1:1 ratios (for example, *X. granatum*:*S. macrophylla* = 2:2), while *X. rumphii* has more genes with a 2:0 ratio (Figure 4b, Table S13). *Xylocarpus granatum* retained 24 128 genes following the WGD event (including the loci retained with all ratios), *X. moluccensis* retained 24 013 genes, and *X. rumphii* retained 23 048 genes, implying that *X. rumphii* lost approximately 1000 more genes (Table S13) than the other two taxa. The excess genes retained in the two mangroves with ratios 2:2 and 1:1 are enriched in many Gene Ontology (GO) terms including oxidoreductase activity, response to light intensity, response to stress, and DNA mismatch repair (Tables S14 and S15). Retention of these genes likely reflects adaptations to intertidal environments with strong UV and high salinity. In contrast, the GO terms related to macromolecule metabolic processes and phosphorylation are overrepresented in the excess genes retained in *X. rumphii* (Xru:Sma = 2:0, Table S16).

We further compared *X. moluccensis* and *X. rumphii* to *X. granatum*. We found that more genes are retained in *X. moluccensis* in parallel with *X. granatum* (2:2 and 1:1, Figure 4c) than in *X. rumphii*. Many of these genes are functionally enriched in response to stress, DNA mismatch repair, and hydrolase activity (Table S17). However, *X. rumphii* preferentially retained more genes completely lost in *X. granatum* (2:0, Figure 4c), and these genes are functionally enriched in macromolecule biosynthetic processes (Table S17).

We also found that gene family contraction overwhelmed expansion in the *Xylocarpus* genus, with *X. rumphii* showing the most drastic contraction (Figure S10). Comparing the four species, we found that the number of unique gene families (present only in one species) is largest in *X. rumphii* (2301) and the number of absent gene families (absent in one species but present in all other species) is also largest in *X. rumphii* (1434) (Figure 4e). These genes absent in *X. rumphii* were enriched in 68 to 81 GO terms ($P < 0.01$), many of which are related to mangrove adaptation to the salty, rhythmically flooded intertidal zone, indicated by GO terms such as response to water deprivation, regulation of circadian rhythm, cell wall macromolecule biosynthetic process, and xylan biosynthetic process (Figure 4d, Table S18). Since rhythmical tides are unique to intertidal zones, these genes may have been important for the two mangroves to adjust to this new circadian rhythm. The cell walls of the two mangroves living in seawater may be distinct. Likewise, the seagrass *Zostera marina* has re-evolved new combinations of structural traits related to cell walls by regaining the ability to produce sulfated polysaccharides and low-methylated zosterin in cell walls (Olsen et al., 2016). In contrast, genes unique to *X. rumphii* are enriched in only one GO term (oxidation–reduction process; $P < 0.01$). These genes are enriched in 10 KEGG (Kanehisa & Goto, 2000) pathways ($P < 0.05$), including sesquiterpenoid and triterpenoid biosynthesis (Table S19).

Fewer genes are pseudogenized in the mangroves. *Xylocarpus granatum* has 3738 and *X. moluccensis* has 2162 more functional genes than *X. rumphii* (Table 1). In addition to gene loss during the post-WGD re-diploidization, *X. rumphii* may have lost genes through pseudogenization. We predicted 6282, 6627, 8873, and 7954 pseudogenes in the four species. Consistently, *X. rumphii* indeed has 2591 more pseudogenes than *X. granatum* and 2246 more than *X. moluccensis*. The increase in *X. rumphii* is significant, using all genes of each species as control (Fisher exact test, $P < 2.2\times10^{-16}$ in both comparisons with *X.*

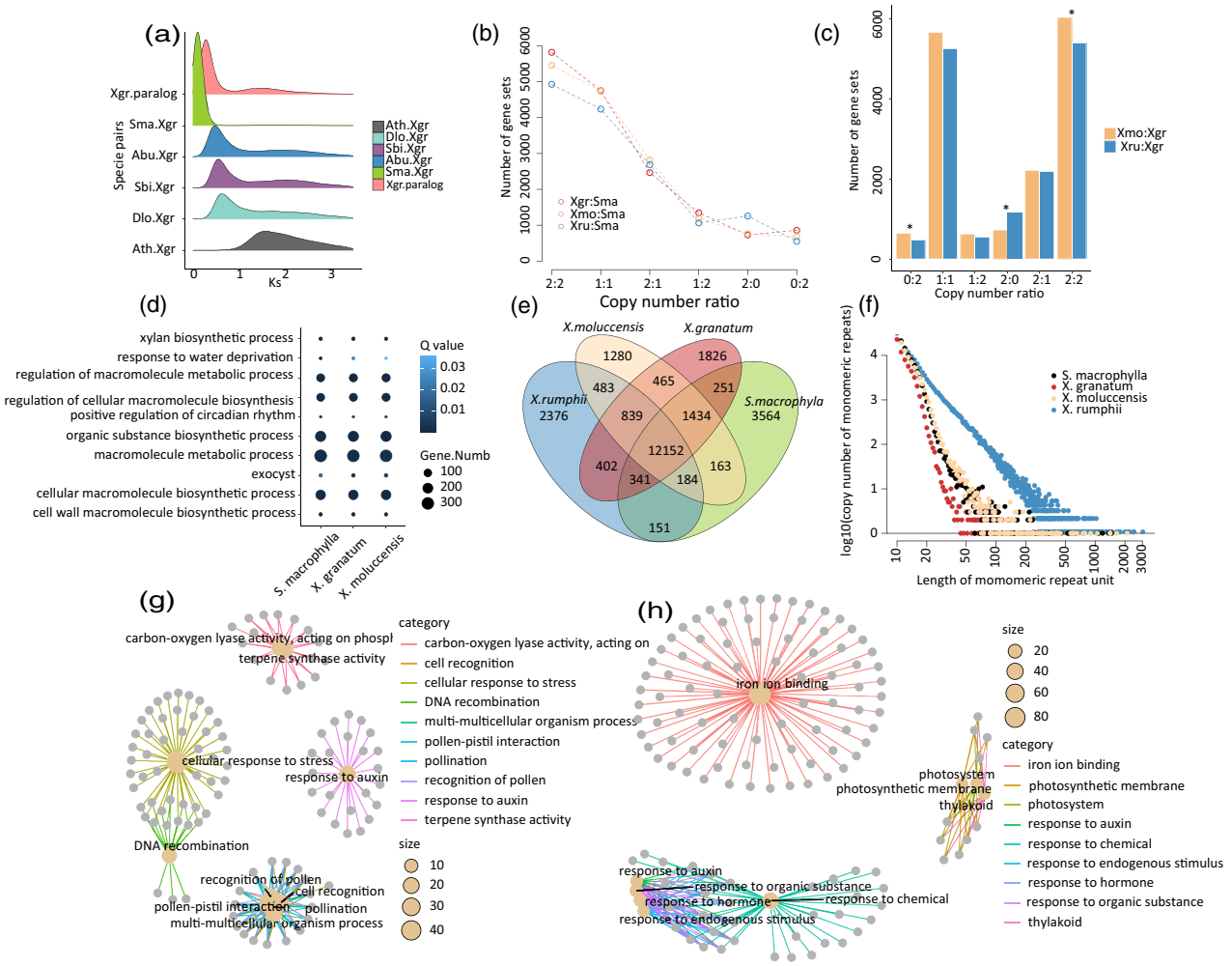

Figure 4. Genomic changes underlying ecological differentiation between *Xylocarpus rumphii* and the two mangroves.
(a) Distribution of Ks values of collinear genes between blocks within the genome and between orthologous genes of closely related species or paralogs within the same species. 'Xgr.paralog' indicates paralogs of *Xylocarpus granatum*, and the others indicate pairs of species. (b) Retention ratios inferred from comparisons of each *Xylocarpus* species against *Swietenia macrophylla*. (c) Retention ratios inferred from comparing *Xylocarpus moluccensis* or *X. rumphii* against *X. granatum*. Asterisks above bar plots indicate $P < 0.01$. Notably, the *P*-value for '1:1' is 0.06. (d) Genes that are lost in *X. rumphii* but retained in the other three species are enriched for GO terms related to mangrove adaptation. Bonferroni correction was used for multiple testing correction. (e) Venn diagram showing gene clustering of the four species. (f) Numbers of simple sequence repeats with various lengths in the four species. (g) Enriched GO terms of *X. granatum* pseudogenes. (h) Enriched GO terms of *X. rumphii* pseudogenes. The abbreviations for species names are: 'Xgr' = *X. granatum*, 'Xmo' = *X. moluccensis*, 'Xru' = *X. rumphii*, 'Sma' = *S. macrophylla*, 'Ath' = *Arabidopsis thaliana*, 'Dlo' = *Dimocarpus longan*, 'Sbi' = *Sclerocarya birrea*, and 'Abu' = *Atalantia buxifolia*.

granatum and *X. moluccensis*). Interestingly, the total gene loss from both preferential WGD retention and pseudogenization matches well with the shortage in functional genes in *X. rumphii* (3671 versus 3738 in comparison with *X. granatum*, and 3211 versus 2162 in comparison with *X. moluccensis*). The pseudogenes are classified into retro- and duplicated pseudogenes. The former originated from WGD while the latter originated from LTR activities. Both retro-pseudogenes (1933 and 1820 versus 2291) and duplicated pseudogenes (962 and 917 versus 1324) have occurred at a higher rate in *X. rumphii* than in the two mangroves (Table 1). Notably, retro-pseudogenes have higher GC content than duplicated genes did (Table S20), because retrogenes usually lack introns (Kaessmann et al., 2009). The GC content of pseudogene fragments in *X. rumphii* was higher than that in the other three species, indicating that fragments of *X. rumphii* probably contain more retro-pseudogene relicts (Table S20). GC content in pseudo- and functional genes is usually higher than in the whole genome (Tables S20 and 21). In short, the two mangroves minimized the rate of pseudogenization (pseudogene/functional gene ratios are 0.20 and 0.22, compared with 0.32 in *X. rumphii*).

Pseudogenes of *X. rumphii* are enriched in GO terms such as iron ion binding, photosystem, and responses to

various signals (such as auxin, chemical, endogenous stimulus, hormone, and organic substance; $P < 0.05$) (Figure 4g). These genes may be important in the two mangroves and thus were retained in these species. Compared to the highly dynamic intertidal habitats, environmental conditions on the cliffs, rocks, and sandy land habitats of *X. rumphii* may be more stable. *Xylocarpus granatum* pseudogenes are preferentially associated with GO terms related to pollination, DNA recombination, terpene synthase, response to auxin, and response to stress ($P < 0.05$, Figure 4h). The pseudogenes of both *X. moluccensis* and *S. macrophylla* are enriched in GO terms hydrolase activity and iron ion binding (Figures S11 and 12). Notably, *X. moluccensis* pseudogenes are also enriched in potassium ion transmembrane transporter activity (Figure S11).

Intolerance of simple sequence repeats. In addition to the changes in functional genes and pseudogenes, repeat sequences may have also evolved differently as the *Xylocarpus* species invaded different environments. As described above, LTR insertions and eliminations were active in all three species in ancient times, but the insertions were repressed in *X. rumphii* recently (Figure 3b). In contrast, the occurrence of simple sequence repeats in the *X. rumphii* genome is twofold more frequent than in the other two species (2.66% against 1.06% and 1.68%, Table S7). As long simple sequence repeats are more likely to have an evolutionary influence on the genome, we compared the number of simple repeats of various lengths among the four species. The two mangroves have many fewer simple repeats with lengths from 20 to 1000 bp than *X. rumphii* (Figure 4f, Table S22). In particular, the two mangroves have nearly no (0 in *X. granatum* and five in *X. moluccensis*) long simple repeats (>1000 bp), but *X. rumphii* has 36 in its genome (Figure 4f, Table S22). However, the majority of these simple sequence repeats are inserted in intergenic regions and intronic regions, and very few of them occur in exons (Table S23).

DISCUSSION

How a species adapts to a new environment is one of the most intriguing questions in evolutionary biology. We used mangrove plants, which have adapted to the extreme conditions of intertidal zones, to disentangle the complex processes of genomic adaptation underlying the shift from a terrestrial to an intertidal lifestyle. By investigating the genomes of three *Xylocarpus* species and their close relative *S. macrophylla*, we show that the two mangroves have shifted from a terrestrial to an intertidal lifestyle through a long and gradual process, with stages of pre-adaptation and final adaptation.

The ancestor of *Xylocarpus* had adapted to intermediate habitats before the final shift to true intertidal habitats. *Xylocarpus* ancestors might have lived close to coasts, with a preliminary ability to tolerate salinity and primitive root systems. Some genomic features, such as high dN/dS ratios, active elimination of LTRs, and positive selection on genes related to adaptive morphological and physiological traits, have pre-adapted the genomes for the final shift to the true intertidal habitats. From such pre-adapted ancestors, the two mangroves subsequently adapted to the true intertidal habitat, facilitated by genomic changes including preferential retention of duplicated genes and intolerance of pseudogenization. It is notable that the so-called mangrove associate species, which are plants inhabiting both intertidal and terrestrial environments, have been considered somewhat preliminary mangroves (Ricklefs et al., 2006). There are more than 20 species of mangrove associates, belonging to various plant families (Duke, 2014; Mao et al., 2021). The ancestors of *Xylocarpus* mangroves and other true mangrove taxa have perhaps had life histories similar to the current mangrove associate species. One may also hypothesize that all three *Xylocarpus* taxa lived in intertidal zones but *X. rumphii* changed to its current habitat later, assuming the *Xylocarpus* common ancestor adapted to the intertidal environments and *X. rumphii* reverted to living along terrestrial coasts. However, this hypothesis is much less likely, because (i) this assumption is much more complicated than our 'pre-adaption hypothesis' and (ii) the genomic changes of *X. rumphii* (intensified pseudogenization and simple sequence repeat insertion) do not appear to be beneficial for living in coastal rock and cliff habitats.

Using the mangroves as a model, we set up a framework for distinguishing the stages of the complex adaptive process to a new environment. Genome-wide adaptations to lifestyle shifts have long been investigated. Several genomic mechanisms were proposed to have played essential roles in many taxa returning to the seas, such as coding gene expansion, rapid evolution of protein-coding genes, and evolution of transposon elements in sea snakes (Peng et al., 2020), positive selection on specific genes and pseudogenization in sea otters (Beichman et al., 2019), and genomic gain and loss of specific genes in seagrasses (Olsen et al., 2016). Interestingly, streptophyte green algae developed multiple traits that represent the first step on the long road of plant terrestrialization (Wang et al., 2020). Although these genomic changes presumably accrued gradually during long-term processes, the previous studies have not attempted to distinguish stages of evolution. By investigating the mangrove transition to intertidal environments, we made one more step toward understanding the details of adaptation to a new environment.

Our study identified evolutions on multiple levels during the mangrove transition to intertidal environments. These evolutionary mechanisms are also generally found in other

taxa. Evolutions of protein-coding genes, including gene gain and loss and amino acid substitutions in coding sequences, are at the center of adaptations to new environments. Transition from terrestrial to marine life in Shaw's sea snakes (Peng et al., 2020), sea colonization by seagrasses (Olsen et al., 2016), scorpions surviving extreme climate changes (Cao et al., 2013), desiccation tolerance of desert mosses (Silva et al., 2021), and Shanputao cold tolerance (Wang et al., 2021) are all associated with an increase of coding gene number. Gene loss also contributes to adaptation to new environments, such as pseudogenization of olfactory receptor genes in sea otters (Beichman et al., 2019) and loss of some angiosperm innovations in seagrasses (Olsen et al., 2016). Evolution of amino acid substitutions is usually interpreted as positive selection, such as in polar bear adaptation to the high arctic (Liu et al., 2014), Tibetan wild boar adaptation to high altitudes (Li et al., 2013), and aquatic adaptation of sea otters (Beichman et al., 2019). Evolution of transposon elements has also contributed to adaptation to a new environments, for example, in Shaw's sea snakes (Peng et al., 2020). Our study of mangroves revealed that amino acids (higher dN/dS ratios and genes under positive selection) and transposon element evolution seem to have played most important roles during intermediate stage adaptation, i.e., pre-adaptation, and gene gain (preferential retention of duplicated genes) and loss (intolerance of pseudogenization) likely contributed more to adaptation to the true intertidal environments. We suggest this stepwise evolution in a new environment is general in other taxa, with diverse genomic mechanisms involved in the process.

Analysis of evolutionary history provides insights useful for conservation. Mangrove plants bear the brunt of sea-level changes as they are strictly constrained in very narrow intertidal habitats along seacoasts. The rising sea levels necessitate careful conservation for mangrove ecosystems to sustain their ecological services in protecting coasts, supporting intertidal biodiversity, mitigating the effects of typhoons, and carbon sequestration (Barbier et al., 2011; del Valle et al., 2020; Guo, Li, et al., 2018). Despite the many adaptive features, mangrove species are commonly low in genetic diversity (Guo, Li, et al., 2018). Consistently, we also found decreasing effective population sizes in *Xylocarpus* species in the past 100 000 years using the pairwise sequential Markov coalescent method (Li & Durbin, 2011) (Figure S13). This low genetic diversity indicates that the individuals in current mangrove populations are highly homogenous, reducing the possibility of evolving new ecotypes from standing variations. The current mangrove species have adapted to intertidal environments through long-term evolution, with multiple genomic changes during the pre- and final adaptation. The projected sea-level rise can reach one meter before the end of this century. It is unlikely that mangrove species will be able to adjust to living in a much more flooded environment or colonize inland habitats in response to the contemporary sea-level rise. Therefore, slowing the rate of sea-level elevation is essential to protect mangrove ecosystems.

EXPERIMENTAL PROCEDURES

Sampling and genome sequencing

We collected *X. granatum* and *X. moluccensis* tissue samples from the Dongzhai Harbor Mangrove Reserve, Haikou, China (19.9°N, 110.6°E), *X. rumphii* samples were from Singapore (1.4°N, 103.7°E), and *S. macrophylla* samples were from Sun Yat-Sen University, Guangzhou, China (23.1°N, 113.3°E). Fresh leaves from adult individuals were collected for DNA sequencing, while leaves, flowers, or roots were collected for RNA sequencing. We used TIANamp Plant DNA Kits (Tiangen, Beijing, China) and the modified CTAB method (Doyle & Doyle, 1987) to extract DNA. TRIzol Universal Reagent (TIANGEN, Beijing, China) was used to extract RNA from the sampled tissues. *Xylocarpus granatum*, *X. moluccensis*, *X. rumphii*, and *S. macrophylla* genomes were generated using SMRT sequencing, supplemented with Illumina reads. Methods used for genome sequencing, assembly, and annotation are detailed in Supplementary Note S1 in the supporting information. These genomes were sequenced as part of a project aiming to sequence most mangrove species worldwide (He et al., 2022).

Phylogenomic analyses

Ortholog groups were inferred for the four species together with additional 11 closely related taxa (Table S24) using OrthoFinder (Emms & Kelly, 2019). In the all-vs-all blast, the e-value cut-off was set to 1.0×10^{-3}. We found 54 single-copy ortholog groups with only one member in each of the 15 species. Coding sequences from each single-copy ortholog group were aligned using a combination of MAFFT (Katoh & Standley, 2013) and PAL2NAL (Suyama et al., 2006). To reduce false-positive variants, we further refined the alignments using GBLOCK (Talavera & Castresana, 2007) to filter out hypervariable regions. All 54 alignments were longer than 150 bp. We concatenated the sequences of these 54 genes to construct the phylogeny for the 15 species. We used maximum likelihood methods to construct this phylogeny with PhyML (Yang, 2007) and RAxML (Stamatakis, 2014). The GTR + I + G model was selected using ModelTest (Posada & Crandall, 1998). One hundred bootstrap replicates were used to estimate posterior probabilities.

As described above, we identified only 54 strictly single-copy ortholog groups in the 15 species, which may be attributed to the fact that many genes in our four focal taxa retained paralogous copies from a recent WGD. To reduce paralog bias, we used Notung 2.9 (Chen et al., 2000) to reconcile gene trees constructed from ortholog groups inferred by OrthoFinder (Emms & Kelly, 2019). Facilitated by the reconciled gene trees, orthologs of the four species were first inferred from low-copy ortholog groups with *Atalantia buxifolia* as the outgroup. Using these orthologs as seeds, we identified orthologs for each of the other 11 species by choosing the candidate with the highest sequence similarity. As a result, we obtained 2582 high-confidence single-copy orthologous groups. We reconstructed the 15-species phylogeny using these 2582 orthologs following the methods described above. We also constructed trees for each of the 2582 orthologous groups using both RAxML (Stamatakis, 2014) and exaBayes (Aberer

et al., 2014). The gene trees were used to generate a consensus species tree using ASTRAL (Mirarab et al., 2014).

To further check the relationships among the four species, we also used OrthoFinder (Emms & Kelly, 2019) to identify 2347 single-copy genes from orthologous gene groups of the three *Xylocarpus* species, *S. macrophylla*, and *Citrus grandis* (outgroup). For each ortholog, we constructed a gene tree with RAxML. We counted the number of gene trees with different topologies.

The reconstructed phylogeny was further dated using MCMCTREE implemented in PAML (Yang, 2007). Several reliable pieces of fossil evidence were placed on the phylogenetic tree for calibration (see Supplementary Note S1). We sampled one chain every five steps, retaining 1 000 000 samples. Before sampling, 1 250 000 iterations were discarded as burn-in. The ratio of nonsynonymous to synonymous amino acid substitutions (dN/dS) on each branch of the 15-species phylogeny was estimated using *codeml* implemented in PAML (Yang, 2007). Gene family (ortholog group) size evolution was inferred for the 15-species phylogeny using CAFE (De Bie et al., 2006).

LTR analyses

We used LTRharvest (Ellinghaus et al., 2008) for *de novo* prediction of LTR-RT elements in the genomes of the four species. An LTR-RT was identified if it has an internal domain of 1 to 15 kb flanked by a pair of putative LTRs with lengths 100 to 1000 bp and sequence similarity over 90%. We estimated the ages of the LTR-RT insertions based on the divergence between the 5′- and 3′-LTRs within each LTR-RT (SanMiguel et al., 1998). We aligned the sequences of each pair of LTRs with MUSCLE and calculated genetic distances using the Kimura two-parameter method (Kimura, 1980). Insertion times were calculated by scaling genetic distances by LTR mutation rates, which were assumed to be twice the genomic average (Ma & Bennetzen, 2004). We estimated average genomic mutation rates as 5.02×10^{-9}, 4.18×10^{-9}, 4.54×10^{-9}, and 3.79×10^{-9} per site per year in each of the four species (Table S25).

Next, LTRdigest (Steinbiss et al., 2009) was used to annotate the internal sequences of candidate LTR-RTs. The candidates with complete *Gag-Pol* were retained as intact LTR-RTs. We identified unpaired LTRs (solo LTRs and truncated LTR-RTs) by identifying sequences with high sequence similarity to the LTRs of intact LTR-RTs. The intact LTR-RTs from all four species were pooled and blasted against each genome. We saved the paralogous sequences (blast hits) with >90% overlap, >90% identity, and e-value $< 1\times10^{-10}$ with any intact LTR-RTs as candidate paralogs. We then searched the flanking 3000-bp sequences (both upstream and downstream) of each candidate paralog against the GyDB2.0 (Llorens et al., 2011) database with tblastn. A candidate paralog was accepted as a truncated LTR-RT if its flanking sequences covered >50% of any Gal-Pol sequence with identity > 30% and e-value $< 1\times10^{-8}$. Candidate paralogs that were not accepted were identified as solo LTRs.

We followed Lyu et al. (2018) to estimate the removal rates of LTR-RTs in each species. SiLiX (Miele et al., 2011) was used to assign LTR-RTs into a cluster if their 5′-LTR sequences overlapped with coverage > 70% and identity > 60%. Each of the solo LTRs and truncated LTR-RTs identified above was assigned to the cluster with its most similar intact LTR-RT. In each LTR cluster, we counted the number of solo LTRs (S), truncated LTR-RTs (T), and intact LTR-RTs (I), allowing us to estimate T:I and S:I ratios. We repeated the estimation process with increasing scaffold size cutoffs to avoid false-positive findings of solo LTRs and truncated LTR-RTs due to the discontinuous assembly of scaffolds. Once values stopped changing with scaffold size, we retained the final estimates.

Gene evolution analyses

Using the method detailed in Supplementary Note S1 in the supporting information, we inferred that the common ancestor of the four species had doubled its whole genome (Figures S5–S9). We then explored the gain and loss of genes duplicated in the WGD by comparing the four species in pairs. For each species pair, we identified collinear genes within species and between species using MCScanX (Wang, Tang, et al., 2012). Gene pairs with Ks values significantly deviating from the WGD peak, which are less likely to have originated from the WGD event, were removed. We used the package igraph (https://igraph.org/redirect.html) to infer retained gene ratios (specifically 1:1, 2:2, 1:2, 2:1, 2:0, and 0:2) in the two compared species.

We predicted pseudogenes in each species using PseudoPipe (Zhang et al., 2006) with default parameters. We removed predicted pseudogenes that overlapped with functional genes. We conducted GO enrichment analysis using the R package clusterProfiler 4.0 (Wu, Hu, et al., 2021).

The branch-site model implemented in PAML (Yang, 2007) was used to detect genes under positive selection. This method allows dN/dS (ω) to vary among both branches of the phylogeny and sites in the protein, aiming to detect positive selection affecting a few sites on particular lineages (foreground branches). We compared the two models with fixed (null model) or variable ω (alternative model) using a likelihood ratio test. Low-copy orthologs in the 15 species mentioned above were used for this analysis. Protein sequences of these orthologs were aligned using MUSCLE (Edgar, 2004) and transformed to codon matrices using PAL2NAL after excluding proteins with fewer than 50 amino acids. We obtained 2582 alignments of orthologs for subsequent analysis. For comparison, *X. granatum*, *X. moluccensis*, and *X. rumphii* were in turn set as the foreground branch. The likelihood ratio test was used to compare the likelihoods of the two models, and the *P*-values were transformed to Q-values using Benjamini–Hochberg correction (Benjamini & Hochberg, 1995) for multiple tests. Functional annotation of the positively selected genes was obtained from TAIR (https://www.arabidopsis.org/).

GO terms enriched among genes on different lists, including positively selected genes, species-specific genes, and WGD retained genes, were inferred using the R package topGO (Alexa & Rahnenfuhrer, 2020). We also identified enriched KEGG terms and used Fisher's exact test to estimate significance levels.

ACKNOWLEDGMENTS

We thank Dr. Kai Zeng at the University of Sheffield for his comments on the manuscript and we thank Dr. Zhang Zhang, Dr. Wuxia Guo, Dr. Haomin Lyu, and Dr. Xinfeng Wang from Sun Yat-Sen University for their advice on experiments and data analyses. We thank Dr. Jean W. H. Yong at the Swedish University of Agricultural Sciences for providing the tissues of *X. rumphii*. This study was supported by the National Natural Science Foundation of China (grant numbers 31830005 and 31900200), the Guangdong Basic and Applied Basic Research Foundation (grant number 2021A1515011160), the National Key Research and Development Plan (grant number 2017FY100705), and the Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (No. 311021006). The authors have no conflict of interest to declare.

AUTHOR CONTRIBUTION

SShi and ZG conceived the study. ZG collected the data, performed the analyses, and wrote the manuscript. SX, WX, SShao, and ZH helped in the data analyses. XF and CZ helped in collecting samples and sequencing. SShi, CW, KH, and SX helped in the preparation of the manuscript.

CONSENT FOR PUBLICATION

The tree images in the bottom panel of Figure 1 were kindly provided by Diana Kleine from the Marine Botany UQ (ian.umces.edu/media-library). This reproduction is authorized by the license of Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) (https://creativecommons.org/licenses/by-sa/4.0/#).

DATA AVAILABILITY STATEMENT

The genome sequences and annotations are deposited into the National Genomics Data Center database under accession numbers GWHBCIL00000000, GWHBCIK00000000, GWHBCIJ00000000, and GWHBCIM00000000.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Supplementary Note S1. Details of genome sequencing, assembly, and annotation, and phylogenomic analyses.

Table S1. Statistics of raw reads of DNA sequencing.

Table S2. Statistics of raw reads of RNA sequencing.

Table S3. Statistics of genome size estimation and mapping Illumina raw reads to the genome assemblies.

Table S4. Assessment of genome sequence assembly and predicted genes using BUSCO.

Table S5. Summary of each predicted protein-coding gene for *Xylocarpus* and *Swietenia macrophylla* genomes.

Table S6. The number of genes with certain numbers of exons.

Table S7. Summary of repeats in the four genomes.

Table S8. KEGG pathway enrichment analyses of genes under positive selection.

Table S9. Genes positively selected in *Xylocarpus granatum*.

Table S10. Genes positively selected in *Xylocarpus moluccensis*.

Table S11. Genes positively selected in *Xylocarpus rumphii*.

Table S12. Summary of whole genome collinearity.

Table S13. Genes retained with different ratios after whole genome duplication.

Table S14. GO terms enriched in conserved genes (2:2 and 1:1) retained in *Xylocarpus granatum*.

Table S15. GO terms enriched in conserved genes (2:2 and 1:1) retained in *Xylocarpus moluccensis*.

Table S16. GO terms enriched in genes retained with a ratio of 2:0 in *Xylocarpus rumphii*.

Table S17. GO terms enriched in retained genes comparing *Xylocarpus moluccensis* and *Xylocarpus rumphii* to *Xylocarpus granatum*.

Table S18. Enriched GO terms of genes only missing or private in *Xylocarpus rumphii*.

Table S19. Enriched KEGG pathways of genes only missing or private in *Xylocarpus rumphii*.

Table S20. GC content of each type of pseudogene in the *Xylocarpus* species and *Swietenia macrophylla* genomes.

Table S21. Statistics of nucleic content of *Xylocarpus* genomes.

Table S22. Number and lengths of simple sequence repeats.

Table S23. Genomic locations of simple sequence repeats.

Table S24. Genomes used in the phylogeny analysis.

Table S25. Mutation rates of the four species.

Table S26. Length information of DNA fragments in PacBio library construction.

Figure S1. Photos showing the habitat of the three *Xylocarpus* species and *Swietenia macrophylla*.

Figure S2. Genome synteny between each of the three *Xylocarpus* species with *Swietenia macrophylla*.

Figure S3. Boxplots and barplots showing features of gene structure and gene families.

Figure S4. dN/dS on each branch of the 15 species phylogeny.

Figure S5. Graphical representation of the genomic features of *Xylocarpus granatum*.

Figure S6. Graphical representation of the genomic features of *Xylocarpus moluccensis*.

Figure S7. Graphical representation of the genomic features of *Xylocarpus rumphii*.

Figure S8. Graphical representation of the genomic features of *Swietenia macrophylla*.

Figure S9. The distributions of Ks between pairs of species or paralogs within species.

Figure S10. The evolutionary dynamics of orthologous gene families among *Xylocarpus* and related species.

Figure S11. Enriched GO terms of *Xylocarpus moluccensis* pseudogenes.

Figure S12. Enriched GO terms of *Swietenia macrophylla* pseudogenes.

Figure S13. Demographic history of *Xylocarpus granatum*, *Xylocarpus rumphii*, *Xylocarpus moluccensis*, and *Swietenia macrophylla* inferred by pairwise sequential Markov coalescence.

REFERENCES

Aberer, A.J., Kobert, K. & Stamatakis, A. (2014) Exabayes: massively parallel bayesian tree inference for the whole-genome era. *Molecular Biology and Evolution*, **31**, 2553–2556.

Alexa, A. and Rahnenfuhrer, J. (2020. topGO: enrichment analysis for gene ontology. R package version 2.42.0.

Barbier, E.B., Hacker, S.D., Kennedy, C., Kock, E.W., Stier, A.C. & Sillman, B.R. (2011) The value of estuarine and coastal ecosystem services. *Ecological Monographs*, **81**, 169–193.

Beichman, A.C., Koepfli, K.P., Li, G., Murphy, W., Dobrynin, P., Kliver, S. *et al.* (2019) Aquatic adaptation and depleted diversity: a deep dive into the genomes of the sea otter and giant otter. *Molecular Biology and Evolution*, **36**, 2631–2655.

Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, **57**, 289–300.

Bock, W.J. (1958) Preadaptation and multiple evolutionary pathways. *Evolution (N. Y)*, **13**, 194–211.

Cadotte, M.W., Campbell, S.E., Li, S.P., Sodhi, D.S. & Mandrak, N.E. (2018) Preadaptation and naturalization of nonnative species: darwin's two fundamental insights into species invasion. *Annual Review of Plant Biology*, **69**, 661–684.

Cao, Z., Yu, Y., Wu, Y., Hao, P., di, Z., He, Y. *et al.* (2013) The genome of Mesobuthus martensii reveals a unique adaptation model of arthropods. *Nature Communications*, **4**, 2602. https://doi.org/10.1038/ncomms3602

Chen, K., Durand, D. & Farach-Colton, M. (2000) NOTUNG: a program for dating gene duplications and optimizing gene family trees. *Journal of Computational Biology*, **7**, 429–447. https://pubmed.ncbi.nlm.nih.gov/11108472/ [Accessed November 30, 2020].

Dawkins, R. (1996) Why the evidence of Evolution Reveals a Universe without design. In: *The Blind Watchmaker*. New York: WW Norton & Company.

De Bie, T., Cristianini, N., Demuth, J.P. & Hahn, M.W. (2006) CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*, **22**, 1269–1271.

del Valle, A., Eriksson, M., Ishizawa, O.A. & Miranda, J.J. (2020) Mangroves protect coastal economic activity from hurricanes. *Proceedings of the National Academy of Sciences of the United States of America*, **117**, 265–270.

Dieterich, C. & Sommer, R.J. (2009) How to become a parasite - lessons from the genomes of nematodes. *Trends in Genetics*, **25**, 203–209.

Doyle, J.J. & Doyle, J.L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11–15.

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

Duke, N.C. (2014) *'World Mangrove iD: expert information at your fingertips' Version 1.1 for Android*. Australia: MangroveWatch Pubication.

Duke, N.C. (2017) Mangrove Floristics and Biogeography Revisited: Further Deductions from Biodiversity Hot Spots, Ancestral Discontinuities, and Common Evolutionary Processes. In: Rivera-Monroy, V.H., Lee, S.Y., Kristensen, E. , and Twilley, R.R. (Eds.) *Mangrove Ecosystems: A Global Biogeographic Perspective*, Switzerland: Springer, pp. 17–53.

Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5**, 113.

Ellinghaus, D., Kurtz, S. & Willhoeft, U. (2008) LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*, **9**, 18. http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-18 [Accessed December 8, 2020].

Emms, D.M. & Kelly, S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, **20**, 238. https://pubmed.ncbi.nlm.nih.gov/31727128

Guo, Z., Guo, W., Wu, H., Fang, X., Ng, W.L., Shi, X. *et al.* (2018) Differing phylogeographic patterns within the Indo-West Pacific mangrove genus Xylocarpus (Meliaceae). *Journal of Biogeography*, **45**, 676–689.

Guo, Z., Li, X., He, Z., Yang, Y., Wang, W., Zhong, C. *et al.* (2018) Extremely low genetic diversity across mangrove taxa reflects past sea level changes and hints at poor future responses. *Global Change Biology*, **24**, 1741–1748.

He, Z., Feng, X., Chen, Q. *et al.* (2022) Evolution of coastal forests based on a full set of mangrove genomes. *Nature Ecology & Evolution*, **6**, 738–749.

He, Z., Xu, S., Zhang, Z., Guo, W., Lyu, H., Zhong, C. *et al.* (2020) Convergent adaptation of the genomes of woody plants at the land-sea interface. *National Science Review*, **7**, 978–993.

Kaessmann, H., Vinckenbosch, N. & Long, M. (2009) RNA-based gene duplication: mechanistic and evolutionary insights. *Nature Reviews. Genetics*, **10**, 19–31.

Kanehisa, M. & Goto, S. (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, **28**, 27–30.

Kanner, J., Harel, S. & Granit, R. (2001) Betalains - a new class of dietary cationized antioxidants. *Journal of Agricultural and Food Chemistry*, **49**, 5178–5185. https://pubmed.ncbi.nlm.nih.gov/11714300/ [Accessed January 6, 2021].

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772–780. https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mst010 [Accessed December 3, 2020].

Kimura, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.

Koenen, E.J.M., Clarkson, J.J., Pennington, T.D. & Chatrou, L.W. (2015) Recently evolved diversity and convergent radiations of rainforest mahoganies (Meliaceae) shed new light on the origins of rainforest hyperdiversity. *The New Phytologist*, **207**, 327–339.

Langmead, B. & Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359. https://pubmed.ncbi.nlm.nih.gov/22388286/ [Accessed December 1, 2020].

Li, H. & Durbin, R. (2011) Inference of human population history from individual whole-genome sequences. *Nature*, **475**, 493–496. doi:10.1038/nature10231

Li, M., Tian, S., Jin, L., Zhou, G., Li, Y., Zhang, Y. *et al.* (2013) Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nature Genetics*, **45**, 1431–1438.

Liu, S., Lorenzen, E.D., Fumagalli, M., Li, B., Harris, K., Xiong, Z. *et al.* (2014) Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell*, **157**, 785–794.

Llorens, C., Futami, R., Covelli, L., Dominguez-Escriba, L., Viu, J.M., Tamarit, D. *et al.* (2011) The Gypsy Database (GyDB) of mobile Genetic elements: release 2.0. *Nucleic Acids Research*, **39**, D70–D74. https://academic.oup.com/nar/article/39/suppl_1/D70/2506273 [Accessed December 8, 2020].

Lyu, H., He, Z., Wu, C.I. & Shi, S. (2018) Convergent adaptive evolution in marginal environments: unloading transposable elements as a common strategy among mangrove genomes. *The New Phytologist*, **217**, 428–438.

Ma, J. & Bennetzen, J.L. (2004) Rapid recent growth and divergence of rice nuclear genomes. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 12404–12410. www.phrap.orgphrap.docsgeneral.html [Accessed December 10, 2020].

Mao, X., Xie, W., Li, X., Shi, S. & Guo, Z. (2021) Establishing community-wide DNA barcode references for conserving mangrove forests in China. *BMC Plant Biology*, **21**, 571. https://doi.org/10.1186/s12870-021-03349-z

Miele, V., Penel, S. & Duret, L. (2011) Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC Bioinformatics*, **12**, 116. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-116 [Accessed December 8, 2020].

Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, S.M. & Warnow, T. (2014) ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*, **30**, 541–548.

Olsen, J.L., Rouzé, P., Verhelst, B., Lin, Y.C., Bayer, T., Collen, J. *et al.* (2016) The genome of the seagrass Zostera marina reveals angiosperm adaptation to the sea. *Nature*, **530**, 331–335. https://doi.org/10.1038/nature16548

Peng, C., Ren, J.L., Deng, C., Jiang, D., Wang, J., Qu, J. *et al.* (2020) The genome of shaw's sea snake (hydrophis curtus) reveals secondary adaptation to its marine environment. *Molecular Biology and Evolution*, **37**, 1744–1760.

Posada, D. & Crandall, K.A. (1998) Modeltest:testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Ricklefs, R.E., Schwarzbach, A.E. & Renner, S.S. (2006) Rate of lineage origin explains the diversity anomaly in the world's mangrove vegetation. *The American Naturalist*, **168**, 805–810.

SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y. & Bennetzen, J.L. (1998) The paleontology of intergene retrotransposons of maize. *Nature Genetics*, **20**, 43–45.

Silva, A.T., Gao, B., Fisher, K.M., Mishler, B.D., Ekwealor, J.T.B., Stark, L.R. *et al.* (2021) To dry perchance to live: Insights from the genome of the desiccation-tolerant biocrust moss Syntrichia caninervis. *The Plant Journal*, **105**, 1339–1356.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Steinbiss, S., Willhoeft, U., Gremme, G. & Kurtz, S. (2009) Fine-grained annotation and classification of de novo predicted LTR retrotransposons. *Nucleic Acids Research*, **37**, 7002–7013. http://www.repeatmasker.org [Accessed December 8, 2020].

Suyama, M., Torrents, D. & Bork, P. (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, **34**, 609–612.

Talavera, G. & Castresana, J. (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments K. Kjer, R. Page, and J. Sullivan, eds. *Systematic Biology*, **56**, 564–577. https://academic.oup.com/sysbio/article/56/4/564/1682121 [Accessed December 3, 2020].

Tomášek, O., Gabrielová, B., Kačer, P., Maršík, P., Svobodová, J., Syslová, K. *et al.* (2016) Opposing effects of oxidative challenge and carotenoids on antioxidant status and condition-dependent sexual signalling. *Scientific Reports*, **6**, 23546.

Tomlinson, P.B. (2016) The botany of mangroves, second edition.

Wang, D., Zhang, Y., Zhang, Z., Zhu, J. & Yu, J. (2010) KaKs_calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics, Proteomics Bioinforma*, **8**, 77–80.

Wang, S., Li, L., Li, H., Sahu, S.K., Wang, H., Xu, Y. *et al.* (2020) Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. *Nature plants*, **6**, 95–106. https://doi.org/10.1038/s41477-019-0560-3

Wang, Y., Tang, H., Debarry, J.D. *et al.* (2012) MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, **40**, e49.

Wang, Y., Wang, X. & Paterson, A.H. (2012) Genome and gene duplications and gene expression divergence: a view from plants. *Annals of the New York Academy of Sciences*, **1256**, 1–14.

Wang, Y., Xin, H., Fan, P., Zhang, J., Liu, Y., Dong, Y. *et al.* (2021) The genome of Shanputao (Vitis amurensis) provides a new insight into cold tolerance of grapevine. *The Plant Journal*, **105**, 1495–1506.

Wu, C.-I., Wen, H., Lu, J., Su, X.D., Hughes, A.C., Zhai, W. *et al.* (2021) On the origin of SARS-CoV-2—The blind watchmaker argument. *Science China Life Sciences*, **64**, 1560–1563. https://doi.org/10.1007/s11427-021-1972-1

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z. *et al.* (2021) clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *The Innovation*, **2**, 100141. https://doi.org/10.1016/j.xinn.2021.100141

Xu, S., Guo, Z., Feng, X., Shao, S., Yang, Y., Li, J., Zhong, C., He, Z. and Shi, S. (2021) Where whole-genome duplication is most beneficial – Adaptation of mangroves to a wide salinity range between land and sea. *Molecular Ecology*, n/a. Available at: https://doi.org/10.1111/mec.16320

Xu, S., He, Z., Guo, Z., Zhang, Z., Wyckoff, G.J., Greenberg, A. *et al.* (2017) Genome-wide convergence during evolution of mangroves from woody plants. *Molecular Biology and Evolution*, **34**, 1008–1015.

Xu, S., He, Z., Zhang, Z., Guo, Z., Guo, W., Lyu, H. *et al.* (2017) The origin, diversification and adaptation of a major mangrove clade (Rhizophoreae) revealed by whole-genome sequencing. *National Science Review*, **4**, 721–734.

Xu, S., Wang, J., Guo, Z., He, Z. & Shi, S. (2020) Genomic convergence in the adaptation to extreme environments. *Plant Communications*, **1**, 100117. https://doi.org/10.1016/j.xplc.2020.100117

Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586–1591.

Zhang, Z., Carriero, N., Zheng, D., Karro, J., Harrison, P.M. & Gerstein, M. (2006) PseudoPipe: an automated pseudogene identification pipeline. *Bioinformatics*, **22**, 1437–1439. https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btl116 [Accessed December 8, 2020].