

# Genomic evidence for rediploidization and adaptive evolution following the whole-genome triplication

Received: 1 August 2023

Accepted: 13 February 2024

Published online: 22 February 2024

 Check for updates

Xiao Feng <sup>1,4</sup>, Qipian Chen <sup>1,2,4</sup>, Weihong Wu<sup>1</sup>, Jiexin Wang<sup>1</sup>, Guohong Li<sup>1</sup>, Shaohua Xu<sup>1</sup>, Shao Shao<sup>1</sup>, Min Liu<sup>1</sup>, Cairong Zhong<sup>3</sup>, Chung-I Wu<sup>1</sup>, Suhua Shi ✉ & Ziwen He ✉

Whole-genome duplication (WGD), or polyploidy, events are widespread and significant in the evolutionary history of angiosperms. However, empirical evidence for rediploidization, the major process where polyploids give rise to diploid descendants, is still lacking at the genomic level. Here we present chromosome-scale genomes of the mangrove tree *Sonneratia alba* and the related inland plant *Lagerstroemia speciosa*. Their common ancestor has experienced a whole-genome triplication (WGT) approximately 64 million years ago coinciding with a period of dramatic global climate change. *Sonneratia*, adapting mangrove habitats, experienced extensive chromosome rearrangements post-WGT. We observe the WGT retentions display sequence and expression divergence, suggesting potential neo- and sub-functionalization. Strong selection acting on three-copy retentions indicates adaptive value in response to new environments. To elucidate the role of ploidy changes in genome evolution, we improve a model of the polyploidization–rediploidization process based on genomic evidence, contributing to the understanding of adaptive evolution during climate change.

The origin and radiation of flowering plants (angiosperms) in the mid-Cretaceous was famously referred to by Charles Darwin as “an abominable mystery”<sup>1,2</sup>. Presently, angiosperms encompass over 90% of all living plant species, with approximately 350,000 known species, making them the most successful land plants on Earth ([www.theplantlist.org](http://www.theplantlist.org)). There is growing consensus that whole-genome duplication (WGD) events, also known as polyploidy, have played a widespread and significant role in the evolutionary history of angiosperms<sup>3–9</sup>. Early WGD events in plants can be traced back to the common ancestors of extant seed plants and angiosperms, respectively<sup>10</sup>. Furthermore, core eudicots, a major clade within angiosperms, experienced a well-known paleo-hexaploidization

event<sup>11</sup>. WGDs have also occurred in various lineages, even recurrently, like *Arabidopsis*<sup>12,13</sup>, soybean<sup>14</sup>, carrot<sup>15</sup>, and *Utricularia*<sup>16</sup>. One specific type of WGD, known as whole-genome triplication (WGT) or hexaploidy, originated through hybridization between tetraploid and diploid species<sup>17–19</sup>. Genomic data have revealed at least 18 independent WGT events in eudicots<sup>11,17,20–34</sup>, indicating a prevalence higher than previously assumed (Supplementary Fig. 1). Despite the challenges that have emerged since the WGD, such as stable chromosome segregation, detrimental ecological interactions with diploid progenitors, and minority cytotype exclusion<sup>35,36</sup>, the polyploidy events observed in plants highlight their evolutionary potential. Experimental and simulation studies have supported the adaptive potential of

<sup>1</sup>State Key Laboratory of Biocontrol and Guangdong Provincial Key Laboratory of Plant Resources, School of Life Sciences, Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Sun Yat-sen University, 510275 Guangzhou, China. <sup>2</sup>Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, 518120 Shenzhen, China. <sup>3</sup>Hainan Academy of Forestry (Hainan Academy of Mangrove), 571100 Haikou, China. <sup>4</sup>These authors contributed equally: Xiao Feng, Qipian Chen. ✉e-mail: [lssssh@mail.sysu.edu.cn](mailto:lssssh@mail.sysu.edu.cn); [hezizwen@mail.sysu.edu.cn](mailto:hezizwen@mail.sysu.edu.cn)

polyploidy, especially in the face of dramatic and fluctuating environmental conditions<sup>37–39</sup>. Overall, polyploidy has been recognized as a major driving force behind evolutionary adaptation and diversification<sup>4,5</sup>.

Plants have experienced periods of global climate change, and genomic resources offer an opportunity to better understand the dynamics of plant evolution during such global climate changes<sup>40–42</sup>. The relationship between WGD and the success of plant lineages is an intriguing topic<sup>43–46</sup>. Previous studies have revealed several instances of WGD occurring independently during three periods of climatic instability and environmental perturbations: the Early Cretaceous around 120 million years ago (Mya)<sup>47</sup>, the K-Pg boundary around 65 Mya<sup>48</sup>, and the Miocene–Pliocene (<20 Mya)<sup>9,49</sup>. These WGD events may have provided a buffer for plants and facilitated their survival and adaptation to rapidly changing environments by increasing genomic plasticity and genotypic combinations.

Mangroves have successfully adapted to extreme intertidal zones, bridging terrestrial and marine ecosystems, evolving a series of adaptive traits, such as salt tolerance, aerial roots, and viviparous embryos<sup>50–52</sup>. They are attractive ecological model systems to investigate adaptive evolution. Prior to colonizing their new habitat, several mangrove species independently experienced WGD events<sup>53–57</sup>. Nevertheless, almost all mangrove species are currently considered diploids (Supplementary Data 1), indicating the importance of the rediploidization process in ancient polyploids. Rediploidization involves redundancy reduction, coordination of subgenomic function, and chromosome fractionations, ultimately leading to the establishment of modern diploid descendants cytogenetically and potentially contributing to plant adaptation<sup>58,59</sup>. Despite its significance, the rediploidization process in ancient polyploid plants remains poorly understood. With advancements in genome sequencing and assembly technologies, high-quality chromosome-scale genomes provided an opportunity to reconstruct ancestral genomes and infer the trajectory of plant genome evolution<sup>60,61</sup>. We can now explore the process of rediploidization following polyploidization on a genomic scale.

In this work, we present two chromosome-scale genomes of Lythraceae plants: the mangrove tree *Sonneratia alba* (Supplementary Fig. 2) and related inland plant *Lagerstroemia speciosa* (Supplementary Fig. 3), as a part of the worldwide mangrove genomes project<sup>62</sup>. Through comprehensive analyses, we trace the evolutionary history of genomes and investigate the polyploidization–rediploidization process and its implications for adaptive evolution in the face of global climate change.

## Results and discussion

### Genome sequencing, assembly, and annotation

We first utilized high-throughput chromosome conformation capture (Hi-C) technology to improve the genome of *S. alba*. This improvement builds upon our prior study utilizing PacBio Single-Molecule Real-Time (SMRT) sequencing and Illumina short reads sequencing<sup>54</sup>, resulting in a chromosome-scale assembly (Supplementary Table 1). The assembled genome derived from anchored contigs was 204.46 Mb, aligning closely with the genome size estimated through k-mer-based analysis (211.67 Mb). It comprised 12 chromosomes (97.60% of all sequences) and 40 unanchored scaffolds. The N50 value notably increased from 5.52 Mb to 15.69 Mb (Table 1). Additionally, we de novo assembled the genome of the closely related inland woody plant *L. speciosa* by incorporating high-depth PacBio SMRT sequencing, Illumina short reads sequencing, and Hi-C technologies (Supplementary Table 1). The assembled genome of *L. speciosa* was 319.66 Mb, with an N50 value reaching 12.74 Mb, consistent with the estimated genome size (361 Mb by flow cytometry and 340.46 Mb by k-mer-based analysis). It comprised 24 chromosomes, encompassing 98.08% of all sequences (Table 1). The assembled genomes of *S. alba* and *L. speciosa* both showed high congruence because of their strongest interaction

**Table 1 | Statistics for *Sonneratia alba* and *Lagerstroemia speciosa* genomes**

Genome features	<i>Sonneratia alba</i>	<i>Lagerstroemia speciosa</i>
Sequencing methods	Illumina + PacBio + HiC	Illumina + PacBio + HiC
Sequencing reads	32.99 Gb + 28.36 Gb + 103.88 Gb	41.16 Gb + 95.60 Gb + 54.19 Gb
Assembled genome size	204.46 Mb	319.66 Mb
Anchored size	199.55 Mb (97.60%)	313.51 Mb (98.08%)
Anchored gene number	25,126	30,323
GC content	41.77%	40.40%
Number of chromosomes	2n = 24	2n = 48
Number of scaffolds	52	629
N50 length	15.69 Mb	12.74 Mb
N90 length	12.96 Mb	10.54 Mb
Longest sequence length	22.93 Mb	17.34 Mb
Gap content	0.05%	0.02%

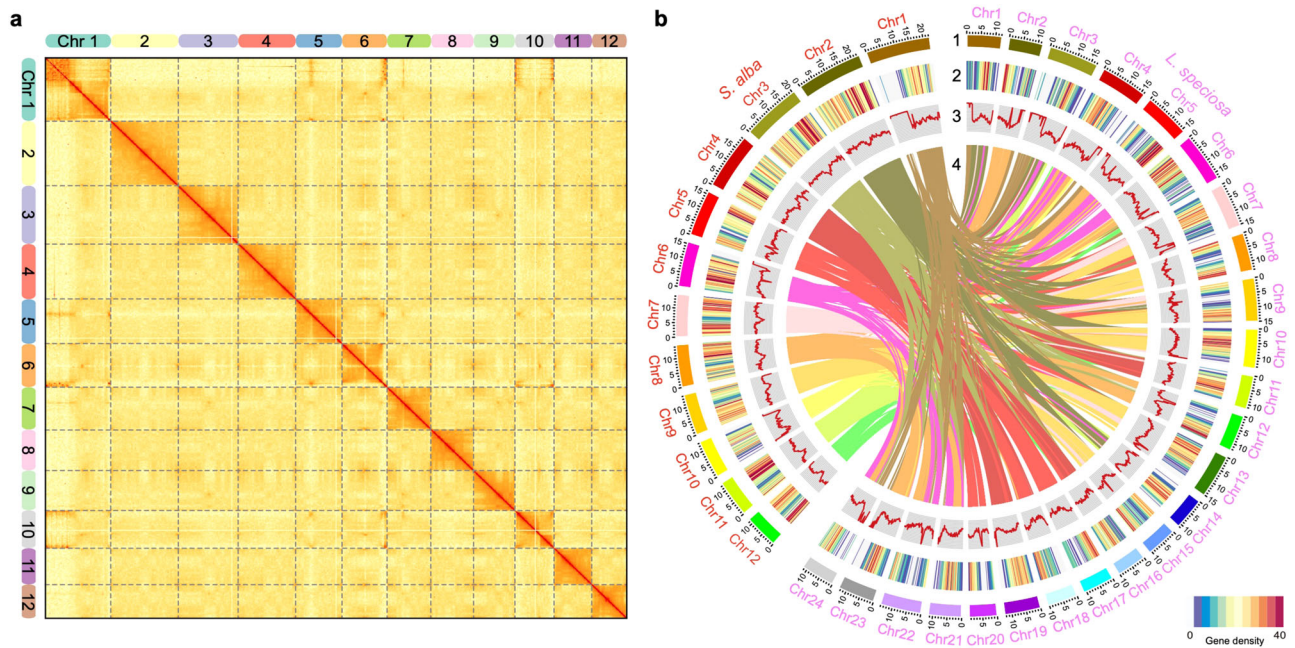
signals from the Hi-C data clustered at the expected diagonal region (Fig. 1a and Supplementary Fig. 4). The gene prediction process involved a comprehensive approach, combining ab initio, homology-based and RNA-seq-assisted strategies. The integration of these predictions through EvidenceModeler resulted in the identification of non-redundant and consensus gene models for the *S. alba* and *L. speciosa* genomes (see Methods for details). This unveiled a total of 25,284 (Supplementary Fig. 5) and 30,497 (Supplementary Fig. 6) protein-coding genes, respectively, characterized by high completeness (Supplementary Table 2). Moreover, 99.38% and 99.43% of them were categorized into chromosomes, respectively. The presence of syntenic blocks between the two genomes further supported their quality as chromosome-scale assemblies (Fig. 1b). These high-quality genomes can supplant earlier assemblies, serving as valuable references for genomic and evolutionary studies in plants (Supplementary Note 1).

### Less TE accumulation in the mangrove

The mangrove species have small genome sizes compared with inland relatives<sup>63</sup>. Repetitive sequences are the primary determinant of plant genome size, and transposable elements (TEs) are the predominant components of repetitive elements<sup>64,65</sup>. First, we observed that *S. alba* has fewer chromosomes compared to *L. speciosa* (Fig. 1b). We estimated that 20.95% (43 Mb) of the *S. alba* genome consists of TE sequences, while 36.50% (117 Mb) in the *L. speciosa* genome and higher TE contents in other relatives (Supplementary Fig. 7 and Supplementary Table 3). The long terminal repeat retrotransposons (LTR-RTs), typical class I TEs, usually have much copy number and large size in plant genomes, contributing significantly to genome size growth<sup>66</sup>. The intact LTR-RTs were further classified as *Copia* and *Gypsy* element families, and their insertion time distributions were examined. We found that *S. alba* has much lower recent LTR-RT insertion rates than relatives in Myrtales, especially in the *Copia* element family (Supplementary Fig. 8). Overall, the mangrove species *S. alba* maintains a smaller genome size, fewer chromosomes, lower accumulation of TEs, and a reduced rate of LTR-RT insertion, resulting in a more simplified genome.

### WGT coinciding with dramatic global climate change

With the availability of chromosome-scale reference genomes in Lythraceae, we revisited the origin of *Sonneratia*, the significant taxon within the mangrove ecosystem. We reconstructed the phylogeny



**Fig. 1 | Genomic architecture of *Sonneratia alba* and relatives. a** Hi-C interactive heatmap of the genome-wide organization of *S. alba*. The deeper red means the stronger interaction between the DNA regions. Chr: chromosome. **b** Circos plot of

the *S. alba* and *L. speciosa* genomes. Concentric circles, from outer to inner, show (1) pseudo-chromosome (Mb), (2) gene density, (3) GC content (34.01–57.76% per 200 Kb), and (4) syntenic block. Source data are provided as a Source Data file.

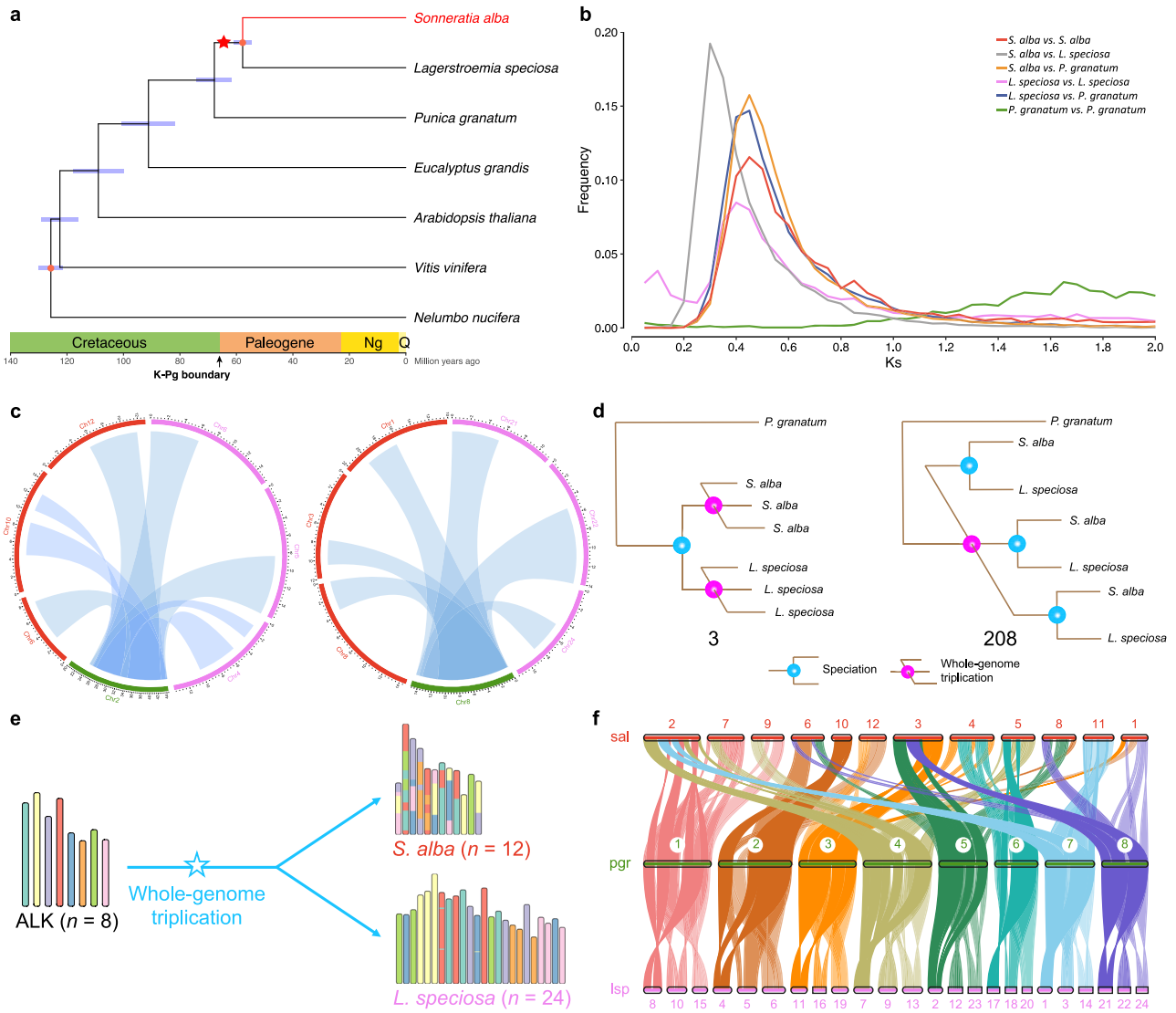
among three Lythraceae species (*S. alba*, *L. speciosa*, and *Punica granatum*) and four other species (*Eucalyptus grandis*, *Arabidopsis thaliana*, *Vitis vinifera*, and *Nelumbo nucifera*) with available pseudo-chromosome scale genome data. The tree topology was inferred using RAXML-NG with the GTR + GAMMA + I model based on 1,963 orthologous single-copy gene groups (Supplementary Fig. 9), and the divergence time was estimated using MCMCTREE from the PAML package with two reliable calibrations (see Methods for details). The divergence times were consistent with a previous study (Supplementary Table 4)<sup>67</sup>. Additionally, our estimation suggests that the mangrove *S. alba* diverged from the closely related inland woody plant *L. speciosa* around 57.79 Mya, while the common ancestor of them diverged from the same family plant *P. granatum* around 67.82 Mya (Fig. 2a). We further constructed a larger-scale phylogenetic tree, incorporating 42 sequenced angiosperms along with the gymnosperm *Gnetum montanum* (as an outgroup), to reflect the positions of these plants within Lythraceae (Supplementary Fig. 10).

Whole-genome duplication (WGD), or polyploidy, events have played a significant role in the evolutionary history of angiosperms, aiding in their survival during periods of dramatic environmental changes<sup>4,9,43</sup>. WGD events can provide a substantial amount of genetic material for adaptation. In this study, we utilized a combination of synteny, Ks-base, and phylogenetic approaches (Supplementary Fig. 11) to confirm that *S. alba* and *L. speciosa* underwent a whole-genome triplication (WGT) event prior to their divergence from a common ancestor (Fig. 2a). Initially, we scanned the genomes of three Lythraceae plants, namely *S. alba*, *L. speciosa*, and *P. granatum*, using BLASTP and MCScanX. We identified 164 syntenic block pairs comprising 5,999 gene pairs in *S. alba*; 486 syntenic block pairs comprising 12,180 gene pairs in *L. speciosa*; and 219 syntenic block pairs comprising 3,333 gene pairs in *P. granatum*. The presence of extensive syntenic block pairs indicated past polyploidy events. Subsequently, we calculated synonymous substitution rates (Ks) between paralogous genes in each genome. The Ks distribution revealed recent peaks in *S. alba* and *L. speciosa*, but not in *P. granatum* (Fig. 2b), suggesting that *P. granatum* did not experience the polyploidy events. Within the Ks peaks range, we identified 584 three-copy retention groups in *S. alba*

and 943 in *L. speciosa* (Supplementary Figs. 12–14), indicating that the polyploidy event in these species was a hexaploidy (whole-genome triplication, WGT) event. This finding was further supported by genome-wide syntenic regions between *S. alba* and *P. granatum*, as well as *L. speciosa*, and *P. granatum* (Fig. 2c and Supplementary Fig. 15). Furthermore, we presented an expected signature of the whole-genome triplication event through collinear genes in the modern genome (Supplementary Fig. 16). While the Ks peak appears slightly different between *S. alba* and *L. speciosa*, we performed gene tree reconstructions of the syntenic gene groups and confirmed that the WGT event occurred prior to the divergence between *Sonneratia* and *Lagerstroemia* (Fig. 2d). The placement of the WGT event was also validated using the multi-taxon paleopolyploidy search (MAPS) analysis and corresponding simulations (Supplementary Fig. 17). This multipronged approach allows us to overcome the challenges posed by divergent evolutionary rates in different plants, enabling the identification of more accurate features and positions of polyploidy events<sup>23,68–70</sup>.

Extrapolating from the divergence time in Lythraceae, we further estimated that the shared WGT event of *Sonneratia* and *Lagerstroemia* occurred around 64 Mya (Fig. 2a, see Methods for details), slightly after the divergence from *P. granatum*. This WGT event coincided with a brief period of dramatic global climate change resulting from a large asteroid collision with the Earth, known as the Cretaceous-Paleogene (K-Pg) boundary, which took place around 66 Mya<sup>48</sup>. Polyploidy events play a significant role in reshaping gene regulatory networks in response to environmental stresses<sup>9,71</sup>. A series of ancient WGD events occurred independently in numerous plant lineages around the K-Pg boundary<sup>43,45,49</sup>. These events served as a buffer for plants, enhancing their ability to survive and adapt to rapidly changing environments by increasing genomic plasticity and generating diverse genotypic combinations. We suggest that the WGT events may have contributed to the survival of plants during the extinction event. Not only that, at approximately 55 Mya, there was a significant global temperature increase (warming by ~6 °C within ~20,000 years) and a rise in eustatic sea levels, known as the Paleocene-Eocene Thermal Maximum (PETM)<sup>72</sup>. The combination of sea level rise, mass extinction, and the





**Fig. 2 | The whole-genome triplication (WGT) event is shared in *S. alba* and *L. speciosa*.** **a** Phylogenetic tree of seven eudicots, including *S. alba* and relatives. Blue node bars are 95% confidence intervals. Red nodes indicate two fossil calibration nodes. The star represents the WGT event. The occurrences of the K-Pg boundary and PETM are indicated by the arrows on the timeline. **b** Ks distribution between paralogous genes within the same species and orthologous genes from pairs of species. **c** Synteny between the homologous regions of *S. alba* (red), *L. speciosa* (pink), and *P. granatum* (green). It reflects the overall synteny relationship with a 3:1

ratio between *S. alba* vs. *P. granatum*, *L. speciosa* vs. *P. granatum*, respectively. The representation showcases partial regions of the genomes. **d** Numbers of homologous gene groups supporting different scenarios on the order of speciation and WGT events in *Sonneratia* and *Lagerstroemia*. **e** Chromosome evolution following WGT from the ALK (the ancestral Lythraceae karyotype). The star represents the WGT event. **f** Macrosynteny patterns among the three Lythraceae plants. “sal” represents *S. alba*, “pgr” represents *P. granatum*, and “lsp” represents *L. speciosa*.

WGT event potentially provided an opportunity in environmental and genetic aspects for offshore woody plants to develop a series of highly specialized traits (such as salt tolerance and aerial roots) to survive, leading to the emergence of the mangrove *Sonneratia*.

Genome evolution is a long-term and dynamic process. Early WGD events ( $\zeta$ ,  $\epsilon$ ,  $\gamma$ ) occurred hundreds of millions of years ago<sup>10,11</sup> and their corresponding collinearity has faded with time or been influenced by subsequent WGD events. Plants that have undergone recent WGD events within the past 20 million years still possess numerous redundant regions in the genomes. Therefore, the WGT event in *Sonneratia* (~64 Mya) provides a valuable opportunity to study the polyploidization–rediploidization process in angiosperms (Fig. 2a). Furthermore, by integrating appropriate genomic data, we positioned the WGT event within a narrow time window between two close speciation events, whose pattern is similar to the  $\gamma$ -WGT event associated with the early diversification of core eudicots.

### Chromosome evolution following the WGT event

Many ancient polyploidy events have been followed by striking reductions in duplicated redundancy and chromosome number<sup>59</sup>. For example, *Utricularia gibba*, despite having a small plant genome, has a haploid chromosome number ( $n$ ) of 14, yet it has undergone three WGD events since the well-known  $\gamma$  event shared by core eudicots<sup>16</sup>. If we exclusively consider polyploidy, the haploid chromosome number of *Utricularia gibba* would be  $7 \times 3 \times 2 \times 2 \times 2$  or  $n = 168$ , based on the ancestral chromosome number ( $n = 7$ ) before experiencing  $\gamma$ -WGT event<sup>73</sup>. Conducting a chromosome-scale comparative investigation among the Lythraceae plants allowed us to explore the paleo-history following the WGT event. Our analysis inferred that the chromosome number of the common ancestor of *Sonneratia* and *Lagerstroemia* is  $n = 24$  (post-WGT) and  $n = 8$  (pre-WGT) (Supplementary Fig. 18). Additionally, the chromosome number of the common ancestor of the three Lythraceae plants is  $n = 8$ , which is the same as the chromosome number in *P. granatum*.

To gain further insights into the evolutionary history of chromosomes, we reconstructed the ancestral Lythraceae karyotype (ALK) using WGD1 based on adjacent conserved collinear blocks. Our evolutionary scenario suggests that the ALK of *S. alba*, *L. speciosa*, and *P. granatum* genomes consisted of eight proto-chromosomes with 18,885 proto-genes. As shown in Fig. 2e, the ancestor underwent a WGT event and subsequently experienced chromosomal rearrangements to attain their modern genome structure. The chromosome origin of *S. alba* appears more intricate than that of *L. speciosa*. *S. alba*'s chromosomes underwent a greater number of fission and fusion events compared to *L. speciosa*, although intra-chromosomal inversions were common in the chromosome histories of both species (Fig. 2e and Supplementary Fig. 19). Due to the complexity of chromosome evolutionary history in *S. alba*, we illustrated it using reciprocally translocated chromosome arms (RTA), end-to-end joining (EEJ), nested chromosome fusion (NCF) events, fission events, and chromosome inversions to depict a probable karyotype evolution (Supplementary Fig. 20).

Although the reconstructed ancestral karyotype is highly likely to possess a structure very similar to the true ancestral genome, it may not be entirely identical<sup>60</sup>. Furthermore, we performed synteny analysis among the modern genomes of the three Lythraceae species and confirmed numerous chromosome rearrangements (Fig. 2f). In contrast to intra-chromosomal inversions observed in related inland species, *S. alba* exhibited significant fission and fusion events (Supplementary Fig. 21). These findings indicate that the mangrove species has a reduced number of chromosomes and underwent more chromosomes rearrangements compared to its closely related inland species *L. speciosa*.

### Adaptation through polyploidization–rediploidization cycles

During periods of dramatic environment and climate change, newly formed polyploids can possess fitness advantages over diploids. This is supported by evidence that the persistence of WGD correlates with times of environmental and climate change, suggesting potential benefit for the WGD in the face of challenges<sup>4,35,45,74–77</sup>. Nevertheless, polyploids may also face substantial disadvantages, including redundant components, gene dosage imbalance, increased replication and metabolic costs, cellular mismanagement, and a higher propensity for polyploid mitosis and meiosis to produce aneuploid cells<sup>35,58,77,78</sup>. Despite these immediate challenges, some polyploid lineages have persisted and even thrived<sup>79,80</sup>. As climatic conditions stabilize and environmental conditions improve, polyploids may experience reduced fitness compared to diploids due to the accumulation of genetic load, increased mutational load, slower positive selection, and reduced growth rates<sup>35,37,81,82</sup>. Therefore, the process of rediploidization following polyploidization may be inevitable for polyploids, ultimately leading to modern descendants as normal diploids cytogenetically, generating important genetic and taxonomic diversity. In fact, nearly all angiosperms have undergone successive rounds of polyploidization and rediploidization process (Supplementary Fig. 22)<sup>4,10,11,83</sup>. Considering the potential role of ploidy changes in genome evolution, we improve a model based on genomic evidence and the previous studies<sup>35,58,59,77,84,85</sup>. This model explains the polyploidization–rediploidization process, elucidating the adaptive evolution during global upheavals and restoration (Fig. 3 and Supplementary Data 2). Specifically, rediploidization through redundancy reduction, gene divergence and chromosome rearrangement confers advantages, such as shortening DNA replication and the cell cycle, and reducing recombination of locally adapted alleles, thereby facilitating the survival of the mangrove in barren intertidal zones.

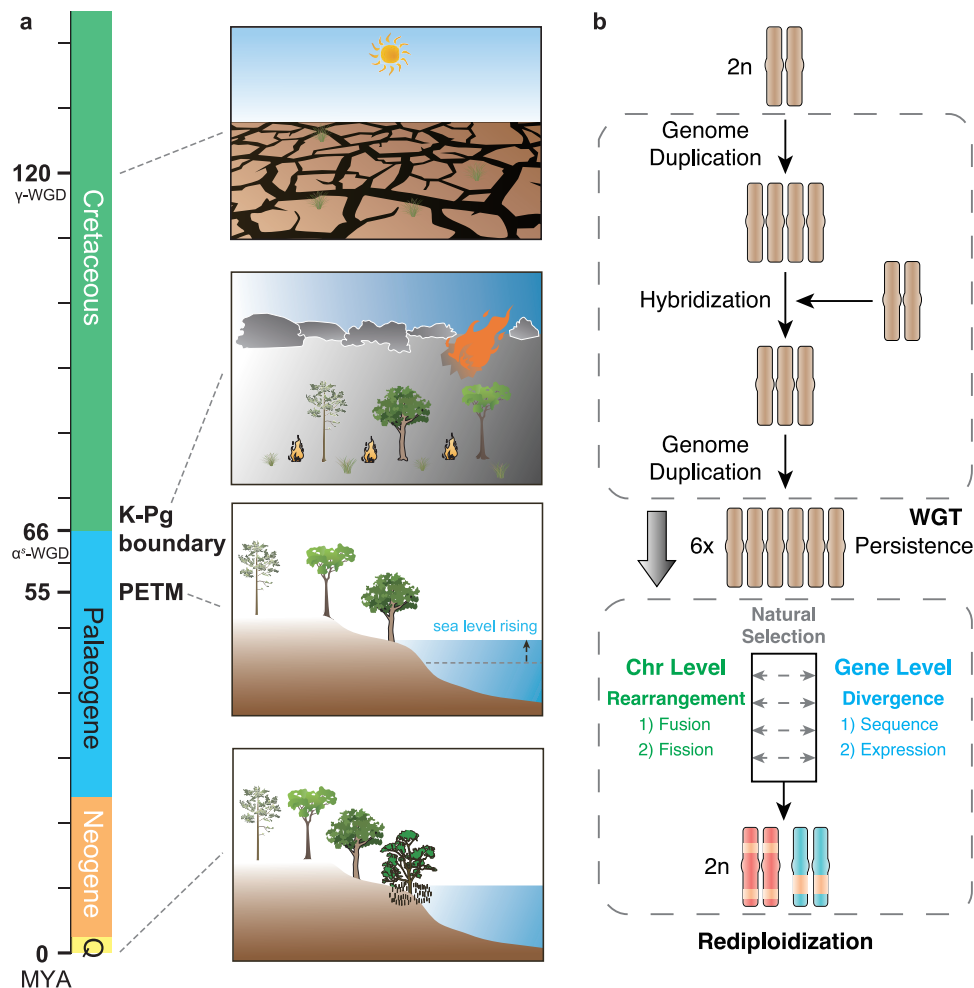
### Divergence of WGT retained genes in the mangrove genome

The differentiation of retained genes plays a crucial role in reducing gene redundancy and serves as a primary genetic basis for genome

evolution. We observed that paralogous gene pairs generated by the WGT event in *S. alba* exhibited higher genetic divergences (Supplementary Fig. 23) and Ks values (Fig. 2b), indicating sequence differentiation. Besides the sequence divergence, expression divergence is also important. Therefore, we conducted transcriptome sequencing of leaf, root, flower, and fruit tissues of *S. alba* (Supplementary Fig. 24 and Supplementary Table 5) and employed the exact conditional test to investigate the expression divergence of WGT retained genes. We identified that approximately 58.04% to 64.57% of the paralogous gene pairs generated by the WGT were differentially expressed across these four tissues in the mangrove species (Supplementary Table 6). The different tissues harbored a similar number of differentially expressed gene pairs (DEGPs), with slightly higher numbers in leaf tissue and lower numbers in fruit tissue. These differentially expressed pairs belonged to 481–516 three-copy retention groups and 1937–2136 two-copy retention groups in different tissues (Supplementary Table 6). Moreover, we identified 1,789 gene pairs that showed differential expression across all four tissues (Supplementary Fig. 25). To investigate the functional roles of these DEGPs, we performed gene ontology (GO) enrichment analysis. The DEGPs in different tissues were predominantly enriched in the metabolic process and catalytic activity GO categories, while the DEGPs shared across all four tissues were enriched in more specific GO categories related to metabolic process, gene expression, biosynthetic process, mitochondrial envelope, and catalytic activity (Supplementary Data 3), which are critical for plant growth and adaptation. Similarly, we explored the expression divergence of WGT retained genes in the closely related inland plant *L. speciosa*. We also identified that around 60% of the paralogous gene pairs resulting from the WGT exhibited differential expression across four tissues in the related species (Supplementary Tables 7, 8), mirroring findings in the mangrove species. These results suggest the potential neo- and sub-functionalization of the retained genes following the polyploidization–rediploidization process.

### Strong selection in WGT retained gene groups

Polyploidy is widely recognized as a major source of novel genetic material, which can undergo mutation and selection to give rise to new or specialized functions to aid adaptation<sup>86,87</sup>. To assess the impact of both negative and positive selection on sites located in different-copy (one-, two-, three-copy) retention groups in the mangrove species, we used the DFE-alpha approach to estimate the distribution of fitness effects (DFE) of new mutations and the proportion of adaptive divergence ( $\alpha$ )<sup>88–91</sup>, based on the folded site frequency spectrum (SFS) and divergence between *S. alba* and *S. apetala*. We also estimated constraint and selection effects using SnIPRE<sup>92</sup>, as well as fixation index (FI)<sup>93</sup> for genes belong to different-copy retention groups to demonstrate the strength of selection (Fig. 4 and Supplementary Fig. 26). We observed that the one-copy retention group exhibited a lower proportion of strongly deleterious mutations ( $N_e s < -100$ ) and a higher proportion of slightly deleterious mutations ( $-1 < N_e s < 0$ ) compared to other copy retention groups (Fig. 4a and Supplementary Fig. 26a). The proportion of adaptive divergence ( $\alpha$ ) at zero-fold nonsynonymous sites and neutral divergence ( $\omega_n$ ) were both higher in the three-copy retention group (Fig. 4a and Supplementary Fig. 26a). They suggested that the strength of negative selection and positive selection were increased in turn from the one-copy to the three-copy retention groups, consistent with estimates of FI (Fig. 4b and Supplementary Fig. 26b). Furthermore, the three-copy retention group demonstrated a lower constraint effect and a higher selection effect compared to other copy retentions (Fig. 4c, d and Supplementary Fig. 26c, d). Collectively, these results indicate preferential retentions of three-copy genes following the polyploidization–rediploidization process, driven by strong selection and possessing potential adaptive value in response to new environments.

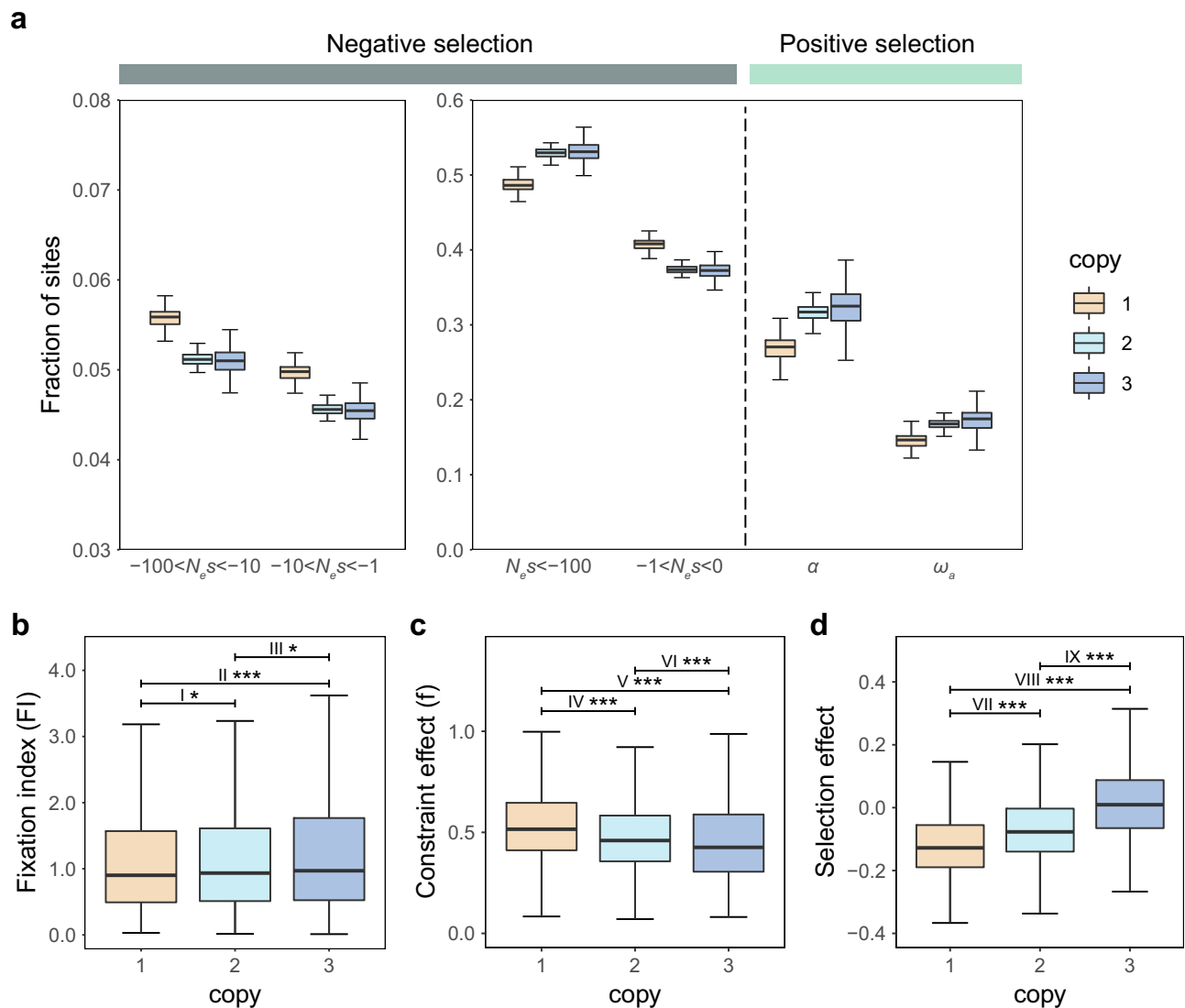


**Fig. 3 | A model of the polyploidization–rediploidization process in plants during global climate change.** **a** Whole-genome triplication ( $\gamma$ -WGT and  $\alpha^s$ -WGT) events coincide with dramatic global climate changes. The sea level rise, massive extinction, and WGT event might provide the opportunity for the offshore woody plants to survive, leading to the emergence of the mangrove. The images portray the prevalent environments during various periods. In the Early Cretaceous (around 120 Mya), an arid climate prevailed<sup>47</sup>. At the K-Pg boundary (around 66 Mya), the dramatic global climate change following a significant asteroid collision with Earth<sup>48</sup>. During the PETM (around 55 Mya), there was a notable global temperature increase and a rise in eustatic sea levels<sup>72</sup>. Finally, the image depicts the current environment. The cartoon elements have been sourced and modified from materials contributed by Christine Thurber, Dieter Tracey, Jane Hawkey, Jane Thomas, and Tracey Saxby (available in the IAN Image Library at <https://ian.umces.edu/media-library/>) under a CC BY-SA 4.0 License. Detailed credits for these cartoon materials can be found in Supplementary Data 2. **b** A hypothetical model of polyploidization–rediploidization process. The initial diploid genome experiences whole-genome triplication around the period of dramatic global environment and climate change. Polyploidy may persist during this period. Rediploidization post polyploidization is a major process for polyploids, driving the genome toward a diploid state through divergence of homologous genes in terms of sequence and expression, redundancy reductions, and large chromosome rearrangements such as fusion and fission events. The round of polyploidization and rediploidization process is widespread in angiosperms. The brown chromosomes represent homologous chromosomes, while the red and blue chromosomes represent significantly diverged chromosomes. The yellow bands indicate regions derived from other ancestral chromosomes through chromosomal rearrangements.

### WGT retained genes for root development and salt tolerance

Mangrove species live in environments characterized by high salinity and waterlogging, which pose challenges to plant growth and productivity<sup>94,95</sup>. The special root systems and high salt tolerance observed in mangroves are particularly noteworthy. *Sonneratia alba*, a prevalent and salt-tolerant mangrove species found in low intertidal zones, has evolved specialized structures like pneumatophores to enhance its waterlogging tolerance<sup>51</sup>. Following the WGT event, duplicated genes are often rapidly lost, while retained duplicates potentially changing expression or acquiring new functions serve as important sources of evolutionary innovation and aid in survival within the newly acquired habitat<sup>56,57,79,96</sup>. Therefore, we conducted functional analyses among the retained genes, which encompassed GO enrichment (Supplementary Fig. 27 and Supplementary Note 2) and gene function assessments based on annotations. Our focus was particularly directed toward the 584 three-copy retention groups generated by the

WGT event. Several gene groups involved in auxin distribution regulation, auxin signal transduction, reactive oxygen species (ROS) scavenging, ion transport, salt overly sensitive (SOS) signaling pathway, abscisic acid (ABA) signaling pathway, and transcriptional regulation retained the three duplicates ultimately (Supplementary Data 4). Interestingly, genetic and physiological experiments have demonstrated that salt modulates root growth direction by causing asymmetric auxin distribution and impairing the gravity response. In response to high salt levels, the SOS signaling pathway mediates the rapid degradation of amyloplasts in root columella cells, leading to the loss of root gravitropism<sup>97,98</sup>. *PIN* and *ABCB* genes, coding transporters polarly localized at the plasma membrane, promote auxin efflux activity<sup>99,100</sup>, while *PP2A* proteins also influence *PIN* localization and participate in the regulation of auxin distribution<sup>101</sup>. Peroxidase protein-coding genes (*POD*) play a role in reducing ROS level, thereby preventing ROS from catalyzing auxin oxidation. These mechanisms



**Fig. 4 | Natural selection patterns among different-copy retention groups.**

**a** The inferred distribution of fitness effects ( $N_e s$ ), proportion of adaptive divergence ( $\alpha$ ), and rate of adaptive substitution relative to neutral divergence ( $\omega_a$ ) for different-copy retention groups generated by the WGT event. Each distribution was estimated based on 200 bootstrap replicates. Traditional MK test with fixation index (**b**), constraint effect (**c**), and selection effect (**d**) for these different-copy retention groups. The two-tailed  $t$  test was applied to test for all pairwise differences.  $P$ -values are indicated by single asterisks ( $P$ -value  $< 0.05$ ) or triple asterisks

( $P$ -value  $< 0.001$ ). I\* ( $P$ -value =  $3.9 \times 10^{-2}$ ), II\*\*\* ( $P$ -value =  $4.5 \times 10^{-4}$ ), III\* ( $P$ -value =  $1.6 \times 10^{-2}$ ), VI\*\*\* ( $P$ -value =  $8.3 \times 10^{-4}$ ), IV\*\*\*, V\*\*\*, VII\*\*\*, VIII\*\*\*, and IX\*\*\* ( $P$ -value  $< 1 \times 10^{-15}$ ) represent the significantly different values for pairwise comparisons between groups. Box edges indicate upper and lower quartiles, center lines indicate median values, and whiskers extend to 1.5 times the interquartile range. The number of genes in each retention group (one-copy, two-copy, and three-copy) was 3439, 4832, and 1171, respectively. Source data are provided as a Source Data file.

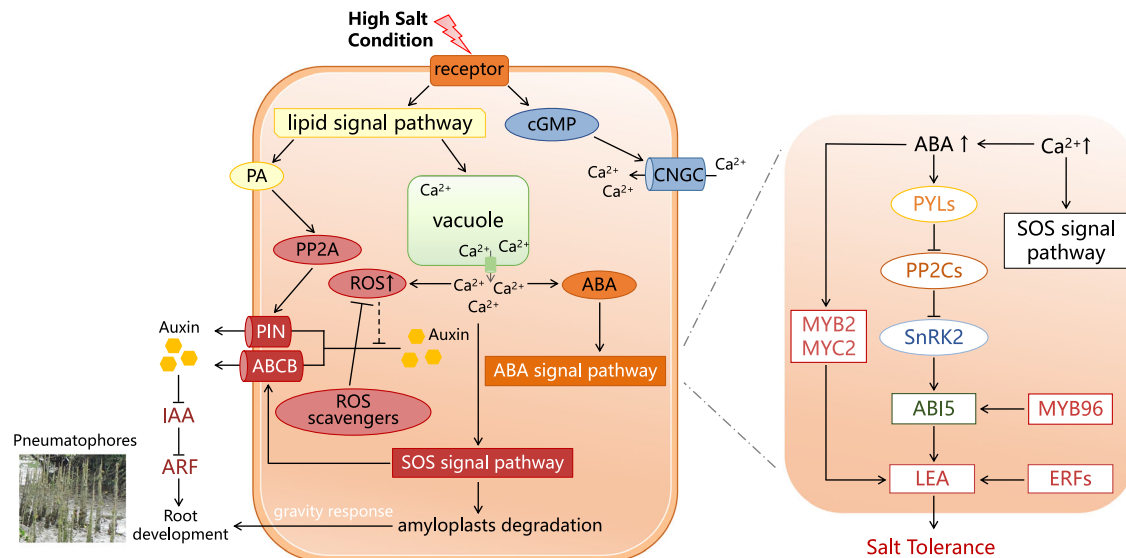
likely together facilitate the development of erect lateral branches in horizontal roots and shape the pneumatophores of *S. alba*, enhancing its waterlogging tolerance (Fig. 5).

Furthermore, we integrated transcriptomes by salt gradient experimental treatments to elucidate the mechanism underlying salt tolerance in *S. alba*. Using the HISAT2-HTseq-DESeq2 workflow, we examined expression profiles and identified differentially expressed genes (Supplementary Fig. 24). We observed 83 three-copy retention groups and 283 two-copy retention pairs with at least one copy showing up-regulation in leaf or root tissues under high salt conditions (Supplementary Table 9). We noticed that a subset of these genes, particularly the three-copy retentions, were associated with the phytohormone abscisic acid (ABA) (Fig. 5), including the ABA pathway, ABA transport, and other related processes (Supplementary Fig. 28). In detail, the release of  $\text{Ca}^{2+}$  in response to high salt triggers ABA biosynthesis<sup>102</sup>. Proteins such as PP2C and ABI5 function in the core

ABA signaling pathway and regulate downstream stress response genes, including *late embryogenesis abundant (LEA)*<sup>103</sup>. The expression of two *LEA* genes, *Sal009147* and *Sal011573*, was found to increase across salinity in leaf tissue. These hydrophilic and heat-stable proteins, with biased amino acid compositions, can sequester accumulated ions within cells and act as chaperones to prevent protein aggregation and inactivation<sup>104,105</sup>. Transcription factors have the potential to regulate multiple aspects of salt adaptation, with MYB and ERF positively influencing genes such as *ABIS* and *LEA*, and MYB2 and MYC2 acting as transcriptional activators in ABA-inducible gene expression<sup>106</sup>. Thus, the up-regulated retained genes can enhance plant desiccation and salt tolerance, contributing to adaptation in intertidal zones.

In summary, we successfully constructed chromosome-scale genomes for two Lythraceae plants, *S. alba* and *L. speciosa*, by combining PacBio SMRT sequencing, short reads sequencing, and Hi-C





**Fig. 5 | The pathway of root development and salt tolerance in *S. alba*.** Salt modulates root growth direction by causing asymmetric auxin distribution and reducing the gravity response. Several genes may together facilitate the erect lateral branches of the horizontal roots and shape the pneumatophores. The

duplicates with at least one copy up-regulated under high salt conditions, related to salt tolerance, are shown on the right. These genes are WGT retained duplicates and detailed three-copy retention groups are listed in Supplementary Data 4.

technologies. Based on genomic evidence and the previous studies<sup>35,58,59,77,84,85</sup>, we report an improved model of the polyploidization–rediploidization process in plants, shedding light on adaptive evolution during periods of global climate change. Our findings revealed that *S. alba* and *L. speciosa* underwent a WGT event at approximately 64 Mya, which coincided with the K-Pg boundary. Subsequently, the mangrove tree experienced extensive chromosomal rearrangements and fractionations, leading to its modern genome structure. We further discovered that the retained duplicates from the WGT event in *S. alba* exhibited not only sequence divergence but also significant expression divergence, which is a crucial mechanism for rediploidization. Overall, our study contributes valuable insights into the plant evolution.

## Methods

### Plant materials

We sampled mature specimens of *Sonneratia alba* (Supplementary Fig. 2) and *Lagerstroemia speciosa* (Supplementary Fig. 3) from the nursery of Dongzhai Harbor National Nature Reserve in Haikou and Sun Yat-sen University in Guangzhou with proper permission, respectively. Fresh and healthy tissues were carefully collected and rapidly frozen in liquid nitrogen. Subsequently, the samples were stored at  $-80^{\circ}\text{C}$  in the laboratory until DNA or RNA extraction was performed.

### Library construction and sequencing

High-molecular-weight (HMW) genomic DNA was isolated from *L. speciosa* leaf tissue using the CTAB (hexadecyltrimethylammonium bromide) method<sup>107</sup> for both PacBio Single-Molecule Real-Time (SMRT) long-read sequencing and Illumina short-read sequencing. A PacBio SMRT-bell library was prepared with 10 kb long inserts following the manufacturer's protocol and subsequently sequenced on a PacBio Sequel II platform (Pacific Biosciences, Menlo Park, CA, USA). The generated PacBio reads underwent data filtering and preprocessing, resulting in 9.57 million reads, corresponding to approximately 95.60 Gb of data and  $\sim 299\times$  coverage (assuming a genome size of 320 Mb, Supplementary Table 1). The same batch of genomic DNA was fragmented using sonication to construct a short-insert paired-end library with 500 bp inserts. This library was sequenced on an Illumina

HiSeq X Ten platform (San Diego, CA, USA), producing 41.16 Gb of data (Supplementary Table 1).

To facilitate gene prediction, total RNA was extracted from leaves of *L. speciosa* using the TRIzol universal reagent (Invitrogen) according to the manufacturer's instructions. The resulting RNA-seq library was sequenced on an Illumina HiSeq X Ten platform (San Diego, CA, USA). Furthermore, total RNA for the expression atlas of *S. alba* was extracted from leaf, root, flower, and fruit tissues of mature plants in the Dongzhai Harbor National Nature Reserve nursery. Each tissue contains three independent biological replicates. RNA-seq libraries were prepared for sequencing on an Illumina HiSeq 2500 platform (San Diego, CA, USA), generating 150 bp paired-end reads. The RNA-seq reads yielded a total of 76.76 Gb of data (Supplementary Table 5).

For Hi-C library construction, tender leaves of both *S. alba* and *L. speciosa* were subjected to formaldehyde fixation and subsequent lysis. The cross-linked DNA was digested with MboI, and the resulting restriction fragment ends were biotinylated and ligated. The purified DNA was then physically sheared to an approximate length of 400 bp. The Hi-C library of *L. speciosa* was sequenced on an Illumina NovaSeq 6000 platform (San Diego, CA, USA), while the Hi-C library of *S. alba* was sequenced on a BGISEQ-500 platform (Shenzhen, China).

### Genome assembly

We reported the genome assembly of *L. speciosa* and improved the previous assembly of *S. alba*<sup>54</sup>. The genome size of *S. alba* was estimated to be 211.67 Mb (Supplementary Fig. 29) through k-mer-based analysis<sup>108</sup>. The genome size of *L. speciosa* was initially estimated using flow cytometry and k-mer-based analysis. The flow cytometry measurement indicated a size of 361 Mb, consistent with the k-mer-based estimation of 340.46 Mb (Supplementary Fig. 30). Then we assembled the de novo genome of *L. speciosa* based on the PacBio long reads using wtdbg2<sup>109</sup> with optimized parameters. To improve the accuracy of the primary assembly, assemblies were further polished with Quiver (SMRT Analysis v2.3.0)<sup>110</sup> using long reads. We further removed residual errors using pilon (v1.22) based on Illumina paired-end reads<sup>111</sup>.

Based on Hi-C data and primary genome assemblies of *S. alba* and *L. speciosa* (Supplementary Table 10), we improved them to generate pseudo-chromosome scale genomes, respectively. The Hi-C data underwent quality evaluation and assessment using HiC-Pro<sup>112</sup>.



Subsequently, the Hi-C maps were generated using Juicer<sup>113</sup>, and the scaffolds were roughly separated using Juicebox<sup>114</sup>. Manual corrections were made to resolve any misassemblies based on the observed interactions. The validated assemblies were then utilized to construct pseudo-chromosomes using the 3D-DNA tool<sup>115</sup>. These pseudo-chromosomes provided a chromosome-scale representation of the genomes, enhancing their structural organization and contiguity.

### Genome annotations

We identified repetitive sequences in each of the whole genomes using a combination of homology-based and de novo approaches. Initially, known TEs within the genome were identified using RepeatMasker with the Repbase TE library, and RepeatProteinMask searches against the TE protein database were conducted. Subsequently, a de novo repeat library for each genome was constructed using RepeatModeler, allowing for comprehensive analysis, refinement, and classification of consensus models for potential interspersed repeats<sup>116</sup>. Additionally, a de novo search for long terminal repeat (LTR) retrotransposons in each genome sequence was performed using LTR\_FINDER (v1.0.7)<sup>117</sup>. Tandem repeats were identified using the Tandem Repeat Finder program, while non-interspersed repeat sequences were detected using RepeatMasker. The results from both approaches were integrated, and RepeatMasker was employed to identify the repeat sequences. We also estimated the age structures of long terminal repeat-retrotransposons (LTR-RTs). After de novo prediction of LTR-RTs, we imposed the criterion that an intact LTR-RT must be separated by 1 to 15 kb from other candidates, flanked by a pair of putative LTRs ranging from 100 bp to 3000 bp, with a similarity of over 80%, and possessed a complete *Gag-Pol* sequence. The timing of LTR-RT insertion was estimated based on the divergence between the 5'-LTR and 3'-LTR of the same transposon, using a mutation rate of  $1.3 \times 10^{-8}$  substitutions per year per site<sup>63</sup>.

We conducted gene model prediction of each genome using a combination of ab initio, homology-based, and RNA-seq-assisted prediction. We used Augustus (v3.3.1)<sup>118</sup> and GeneMark<sup>119</sup> to perform ab initio gene prediction based on the masked genome except for the low complexity or simple repeats, because some of these repeats could be found in the genes. The protein sets were collected for homology-based prediction and chosen as homology-based evidence from sequenced relative plant species and model plant species. Then exonerate (v2.2.0) was used to generate the gene structures based on the homology alignments<sup>120</sup>. Clean RNA-seq reads were mapped against the assembly using Tophat2 (v2.1.1), and transcripts were identified using Cufflinks (v2.2.1)<sup>121,122</sup>. Finally, EvidenceModeler was used to integrate all predictions to generate non-redundant and consensus gene models<sup>123</sup>. Gene functions were annotated based on the best alignment matches to the NCBI (NR), Swissprot, TrEMBL, InterPro, the Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), and Pfam non-redundant protein databases. The transcription factor identification was performed using iTAK (v1.7)<sup>124</sup>. The quality of genome assembly and annotation was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCOv5) with the plant-specific dataset (eudicotyledons\_odb10)<sup>125</sup>.

### Phylogeny reconstruction and molecular dating

We downloaded genome and annotation data for *Vitis vinifera* (Genoscope.12X)<sup>73</sup>, *Arabidopsis thaliana* (TAIR10)<sup>126</sup>, *Eucalyptus grandis* (v2.0) from Phytozome v12.1 database<sup>83</sup>, *Nelumbo nucifera* from Nelumbo Genome Database<sup>127</sup>, *Punica granatum* (GCF\_007655135.1) from NCBI database<sup>128</sup>. We used OrthoFinder to identify orthologous genes among *S. alba*, *L. speciosa*, and these five eudicot species, resulting in the identification of single-copy gene groups<sup>129</sup>. For each group, we aligned the corresponding single-copy orthologous proteins and generated codon alignments using MAFFT<sup>130</sup> and PAL2NAL<sup>131</sup>. To ensure data quality, we further applied Gblocks

0.91b<sup>132</sup> to trim the alignments and discarded ambiguity alignments shorter than 150 bp. Using the concatenated alignment of these groups, we inferred a phylogenetic tree using RAxML-NG<sup>133</sup> with the GTR + GAMMA + I model and performed 1,000 bootstrap replicates. Following its reconstruction, we estimated the divergence time among the seven species using MCMCTREE from the PAML (v4.9j) package with approximate likelihood calculation<sup>134,135</sup>. The HKY85 + G nucleotide substitution model and independent-rates clock model were employed in the molecular dating. To provide calibration points, we incorporated two reliable fossil calibrations. Firstly, the root node of eudicots was placed at 119.6–128.63 Mya<sup>136</sup>. Secondly, the common ancestor of *Sonneratia* and *Lagerstroemia* was set to a time earlier than 55.8 Mya, since the earliest convincing fossils of *Sonneratia*-like pollen<sup>137</sup>. In order to delineate the positions of these plants within Lythraceae, we expanded our analysis by constructing a more extensive phylogenetic tree using these seven plants, other 35 genome-sequenced angiosperms, and the gymnosperm *Gnetum montanum* as an outgroup (Supplementary Data 5). Utilizing the embryophyta\_odb10 lineage ancestral variant dataset (comprising a consensus sequence and variants of extant sequences) in BUSCOv5<sup>129</sup>, we identified 868 low-copy nuclear genes. We then performed sequence alignment and phylogenetic inference as described earlier. The early divergence times in angiosperms were set to 125–247.2 Mya<sup>138,139</sup>. All MCMC analyses were independently run twice to ensure convergence, with 10 million generations and sampling every 500 generations after a burn-in of 1,000,000 iterations. The phylogenetic trees were visualized using the R package GGTREE<sup>140</sup>.

### Whole-genome triplication analyses

In order to identify and locate putative WGDs in Lythraceae species, we used a multipronged approach, including the intra- and interspecies synteny analysis, Ks-base estimation, and phylogenetic reconciliation. Initially, we utilized the BLASTP program to align protein sequences between species (*P. granatum* vs. *S. alba*, *P. granatum* vs. *L. speciosa*) and within species, applying the parameters (identity  $\geq 30\%$ , e-value  $< 1e-10$ , alignment length  $\geq 30\%$  of both query and reference sequences). We identified syntenic blocks containing a minimum of five shared genes using MCScanX<sup>141</sup>, and the resulting syntenic blocks between species were visualized by Circos<sup>142</sup>. Subsequently, we applied KaKs\_Calculator to calculate synonymous substitution rates (Ks) with the YN substitution model<sup>143</sup> based on alignments of all syntenic gene pairs and constructed Ks distribution. To identify paralogous genes generated from the WGD event in the *S. alba* genome, we selected blocks with median Ks values in the range of 0.2–1.0, excluding gene pairs with Ks values larger than 1.26. Using the R package igraph (<https://igraph.org>), we further classified different-copy retention groups after the WGD event (Supplementary Fig. 12). We also identified different-copy retention groups after the WGD event in *L. speciosa* using the same workflow (Supplementary Fig. 13). The analyses of synteny and Ks-base indicated that both *S. alba* and *L. speciosa* had undergone a whole-genome triplication (WGT) event. Simultaneously, we illustrated the distribution of gene densities for different-copy retention groups in both species (Supplementary Fig. 14). To investigate whether the WGT event was shared between *S. alba* and *L. speciosa*, we identified 306 *P. granatum* genes that possessed three orthologs generated by the WGT in both *S. alba* and *L. speciosa*. Then we performed gene tree reconstruction using RAxML-NG and classified the phylogenetic trees based on their topologies.

We also inferred and located the putative WGT placement using the multi-taxon paleopolyploidy search (MAPS) tool<sup>144</sup>. Clustering gene families among five species, including *S. alba*, *L. speciosa*, *P. granatum*, *E. grandis*, and *A. thaliana*, by OrthoFinder, we retained the gene families with at least one gene present in each species. We constructed gene trees based on multiple sequence alignments of each

gene family as described above and rooted each tree using Notung (v2.9.1.5)<sup>145</sup>. By mapping these gene trees to the given species tree, we calculated the percentage of subtrees with gene duplications shared by all species descended from each node using the MAPS tool. To validate the WGT placement, we compared the subtree percentages at each node among observed, null simulated, and positive simulated data and recognized a significant gene duplication burst indicative of a polyploidy event. Background gene birth and death rates were estimated using the R package WGDgc<sup>146</sup> and the mean of a geometric distribution of the root was calculated through CAFE analysis<sup>147</sup>. We performed 2,000 simulated gene tree simulations with 200 bootstrap replicates for both null and positive simulations. In the positive simulation, we designated a polyploidy event in the common ancestor of *S. alba* and *L. speciosa*, setting the `wgd_retention_rate` to 0.2. The observed and simulated data were compared to evaluate the location of the WGT events. The results of the synteny, Ks-based, and phylogenetic analyses consistently indicated that *S. alba* and *L. speciosa* had undergone a whole-genome triplication (WGT) event prior to their diverging from a common ancestor.

To determine the absolute timing of the whole-genome duplication (WGT) event, we conducted a molecular clock analysis on concatenated gene families, calibrated using species divergence times<sup>148</sup>. Specifically, we first identified 208 homologous gene groups among *S. alba*, *L. speciosa*, and *P. granatum* supporting the WGT event before the speciation event (Fig. 2e). For each group, we used the reciprocal BLASTP best-hit method between *P. granatum* and *E. grandis* to obtain the corresponding ortholog from *E. grandis*. Ultimately, we identified 170 gene families that exhibited a clear signal of the WGT event on the common ancestor of *S. alba* and *L. speciosa*. In each gene tree, the genes from *S. alba* and *L. speciosa* were divided into three clades. To improve the robustness and precision of estimation, we selected the outgroup clade and one of the two ingroup clades randomly to concatenate multiple sequence alignments. The phylogenetic tree was then constructed using RAXML-NG, and the molecular clock analysis was performed using the approximate likelihood calculation method in MCMCTREE under the appropriate model<sup>134,135</sup>. The nodes were constrained using species divergence times obtained from the phylogenetic tree as described above. Each analysis was independently run twice to ensure convergence.

### Chromosome evolution

In order to infer their evolutionary history, we selected representative species in the Lythraceae with chromosome-scale genome assemblies. We first inferred the ancestral chromosome numbers across the phylogenetic tree (Supplementary Fig. 18) using ChromEvol (v2.1)<sup>149</sup>. The haploid chromosome number ( $n$ ) of *Sonneratia alba* was reported by S. Graham<sup>150</sup>, while the  $n$  of other Lythraceae species (*Lagerstroemia speciosa*, *Punica granatum*, *Pemphis acidula*) and outgroup (*Eucalyptus grandis*) were obtained from the Chromosome Counts Database (CCDB)<sup>151</sup>. The chromosome number of the most recent common ancestor (MRCA) among *S. alba*, *L. speciosa*, and *P. granatum* was the same as *P. granatum* ( $n = 8$ ). We utilized WGDl (v0.6.5) to identify adjacent conserved collinear genes and blocks among all chromosome pairs within the three Lythraceae species, and then reconstructed the Ancestral Lythraceae Karyotype (ALK), excluding interference from fragmented collinear regions, following the tutorial<sup>152,153</sup>. Subsequently, we visualized the global pattern of chromosomal changes in extant species. Furthermore, we depicted the evolutionary history of *S. alba* chromosomes to provide a clearer representation of the karyotype evolution<sup>152</sup>. While the reconstructed ancestral karyotype almost certainly has a very similar structure to the true ancestral genome, it may not be absolutely identical<sup>60</sup>. We also conducted synteny analysis among the modern genomes of the three Lythraceae species using MCScanX and JCVI to discover chromosome rearrangements<sup>154</sup>.

### Transcriptome sequencing and analysis

RNA-seq reads from the leaf, root, flower, and fruit tissues of *S. alba* were first filtered using SolexaQA++ (v3.1.7.1)<sup>155</sup>. Clean reads were aligned to the *S. alba* genome using HISAT2 (v2.2.0)<sup>156</sup>. The HTSeq (v0.13.5) was utilized to determine the number of reads uniquely mapped to each gene in the tissue samples<sup>157</sup>. To detect the differential expression of duplicated genes, we employed the exact conditional test<sup>158</sup>, which has been successfully applied in soybean, *Brassica*, and *Avicennia*<sup>159–161</sup>. For each pair of duplicated genes, we computed the  $P$ -value using the R function `binom.test`. Multiple testing was corrected by applying the Bonferroni correction method. Differential expression was considered significant for gene pairs with a corrected  $P$ -value below 5%. Only gene pairs whose at least one gene read number more than 0 were included in the analysis. We also applied the method for three-copy duplicated gene groups through pairwise comparisons. We finally identified duplicated gene pairs with differential expression for each tissue based on the consistency in three replicates. To investigate the functional roles of these differentially expressed gene pairs, we performed GO enrichment analysis using BiNGO in Cytoscape (v3.7.2)<sup>162</sup>. Additionally, we performed RNA-seq on leaf, stem, flower, and fruit tissues of *L. speciosa* (Supplementary Table 7 and Supplementary Fig. 31). We employed the HISAT2–HTSeq–exact conditional test workflow, as described earlier, to identify differentially expressed duplicated gene pairs. Subsequently, we conducted GO enrichment analysis on these gene pairs in *L. speciosa* (Supplementary Fig. 32).

We also collected the transcriptomes of leaf and root tissues of *S. alba* under different salinity conditions<sup>163</sup>. Specifically, *S. alba* plants were divided into three groups and subjected to irrigation with solutions containing 0, 250, and 500 mM NaCl to simulate low, medium, and high salinity conditions, respectively. After reads mapping and counting, we identified differentially expressed genes (DEGs) from two comparisons (medium *vs.* low salinity condition, high *vs.* medium salinity condition) using the R package DESeq2<sup>164</sup>. The  $P$ -value below 5% and fold change greater than two was set as the significantly differential expression threshold.

### Distribution of fitness effects of variants

To quantify the impact of selection on different-copy retention groups after the WGT event, we estimated the distribution of fitness effects (DFE) of new mutations and the proportion of adaptive divergence ( $\alpha$ ) at zero-fold nonsynonymous sites using DFE-alpha<sup>88–90</sup>. The whole-genome resequencing data from two populations (Cebu, Philippines; Davao, Philippines) of *Sonneratia alba* (12 individuals per population), and one individual of congeneric species *S. apetala* were used in the analysis. We first binned protein-coding genes into three subsets according to retained copy numbers (one-, two-, three-copy). We used DFE-alpha to compare the folded site frequency spectrum (SFS) and divergence of zero-fold nonsynonymous sites with those for four-fold synonymous sites. The four-fold synonymous sites were assumed to be neutral. We also estimated DFE and  $\alpha$  with 200 bootstrap replicates generated by randomly sampling genes of each subset.

### Reporting summary

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

### Data availability

The raw genomic Illumina reads, PacBio reads, Hi-C reads, and RNA-seq reads reported in this paper have been deposited in the Genome Sequence Archive (GSA, <https://ngdc.cncb.ac.cn/gsa>) in National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation, under accession number CRA004284 with BioProject ID PRJCA005319. The genome assembly sequences have been deposited in the Genome Warehouse (GWH, <https://ngdc.cncb.ac.cn/gwh>) in National Genomics Data Center

under accession number GWHBCIQ00000000 [<https://ngdc.cnbc.ac.cn/gwh/Assembly/20653/show>], GWHBCKL00000000 [<https://ngdc.cnbc.ac.cn/gwh/Assembly/20692/show>] with BioProject ID PRJCA004930 and BioSample ID SAMC353197, SAMC353201. The genome assemblies and annotations are also available at Figshare: *Sonneratia alba* [<https://doi.org/10.6084/m9.figshare.25118819>], *Lagerstroemia speciosa* [<https://doi.org/10.6084/m9.figshare.25118831>]. Source data are provided with this paper.

## References

- Darwin, F. & Seward, A. C. *More Letters of Charles Darwin* (John Murray, 1903).
- Friedman, W. E. The meaning of Darwin's 'abominable mystery'. *Am. J. Bot.* **96**, 5–21 (2009).
- Wood, T. E. et al. The frequency of polyploid speciation in vascular plants. *Proc. Natl Acad. Sci.* **106**, 13875–13879 (2009).
- Van de Peer, Y., Mizrahi, E. & Marchal, K. The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424 (2017).
- Ren, R. et al. Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Mol. Plant* **11**, 414–428 (2018).
- Clark, J. W. & Donoghue, P. C. J. Whole-genome duplication and plant macroevolution. *Trends Plant Sci.* **23**, 933–945 (2018).
- Rice, A. et al. The global biogeography of polyploid plants. *Nat. Ecol. Evol.* **3**, 265–273 (2019).
- Qiao, X. et al. Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biol.* **20**, 38 (2019).
- Wu, S., Han, B. & Jiao, Y. Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. *Mol. Plant* **13**, 59–71 (2020).
- Jiao, Y. et al. Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**, 97–100 (2011).
- Jiao, Y. et al. A genome triplication associated with early diversification of the core eudicots. *Genome Biol.* **13**, R3 (2012).
- Bowers, J. E., Chapman, B. A., Rong, J. & Paterson, A. H. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**, 433–438 (2003).
- Monnahan, P. et al. Pervasive population genomic consequences of genome duplication in *Arabidopsis arenosa*. *Nat. Ecol. Evol.* **3**, 457–468 (2019).
- Schmutz, J. et al. Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 178–183 (2010).
- Iorizzo, M. et al. A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nat. Genet.* **48**, 657–666 (2016).
- Ibarra-Laclette, E. et al. Architecture and evolution of a minute plant genome. *Nature* **498**, 94–98 (2013).
- Bock, D. G., Kane, N. C., Ebert, D. P. & Rieseberg, L. H. Genome skimming reveals the origin of the Jerusalem Artichoke tuber crop species: neither from Jerusalem nor an artichoke. *N. Phytologist* **201**, 1021–1030 (2014).
- Mandáková, T., Pouch, M., Brock, J. R., Al-Shehbaz, I. A. & Lysak, M. A. Origin and evolution of diploid and allopolyploid *Camelina* genomes was accompanied by chromosome shattering. *Plant Cell* **31**, 2596–2612 (2019).
- Aköz, G. & Nordborg, M. The *Aquilegia* genome reveals a hybrid origin of core eudicots. *Genome Biol.* **20**, 256 (2019).
- Moghe, G. D. et al. Consequences of whole-genome triplication as revealed by comparative genomic analyses of the wild radish *Raphanus raphanistrum* and three other Brassicaceae species. *Plant Cell* **26**, 1925–1937 (2014).
- Kagale, S. et al. The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. *Nat. Commun.* **5**, 3706 (2014).
- Cheng, S. et al. The *Tarenaya hassleriana* genome provides insight into reproductive trait and genome evolution of Crucifers. *Plant Cell* **25**, 2813–2830 (2013).
- Wang, J. et al. Recursive paleohexaploidization shaped the Durian genome. *Plant Physiol.* **179**, 209–219 (2019).
- Wang, Z. et al. The genome of *Hibiscus hamabo* reveals its adaptation to saline and waterlogged habitat. *Hortic. Res.* **9**, uhac067 (2022).
- Tu, L. et al. Genome of *Tripterygium wilfordii* and identification of cytochrome P450 involved in triptolide biosynthesis. *Nat. Commun.* **11**, 971 (2020).
- Hane, J. K. et al. A comprehensive draft genome sequence for lupin (*Lupinus angustifolius*), an emerging health food: insights into plant-microbe interactions and legume evolution. *Plant Biotechnol. J.* **15**, 318–330 (2017).
- Lee, E. S. et al. Engineering homoeologs provide a fine scale for quantitative traits in polyploid. *Plant Biotechnol. J.* **21**, 2458–2472 (2023).
- Bombarely, A. et al. Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nat. Plants* **2**, 16074 (2016).
- Sun, G. et al. Large-scale gene losses underlie the genome evolution of parasitic plant *Cuscuta australis*. *Nat. Commun.* **9**, 2683 (2018).
- Liang, Y. et al. The genome of *Eustoma grandiflorum* reveals the whole-genome triplication event contributing to ornamental traits in cultivated lisianthus. *Plant Biotechnol. J.* **20**, 1856–1858 (2022).
- Liu, B. et al. *Mikania micrantha* genome provides insights into the molecular mechanism of rapid growth. *Nat. Commun.* **11**, 340 (2020).
- Palfalvi, G. et al. Genomes of the venus flytrap and close relatives unveil the roots of plant carnivory. *Curr. Biol.* **30**, 2312–2320.e5 (2020).
- Song, A. et al. Analyses of a chromosome-scale genome assembly reveal the origin and evolution of cultivated chrysanthemum. *Nat. Commun.* **14**, 2021 (2023).
- Nakano, M. et al. A chromosome-level genome sequence of *Chrysanthemum seticuspe*, a model species for hexaploid cultivated chrysanthemum. *Commun. Biol.* **4**, 1167 (2021).
- Baduel, P., Bray, S., Vallejo-Marin, M., Kolář, F. & Yant, L. The “Polyploid Hop”: shifting challenges and opportunities over the evolutionary lifespan of genome duplications. *Front Ecol. Evol.* **6**, 117 (2018).
- Comai, L. The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* **6**, 836–846 (2005).
- Meeus, S., Šemberová, K., De Storme, N., Geelen, D. & Vallejo-Marin, M. Effect of whole-genome duplication on the evolutionary rescue of sterile hybrid monkeyflowers. *Plant Commun.* **1**, 100093 (2020).
- Selmecki, A. M. et al. Polyploidy can drive rapid adaptation in yeast. *Nature* **519**, 349–352 (2015).
- Ebadi, M. et al. The duplication of genomes and genetic networks and its potential for evolutionary adaptation and survival during environmental turmoil. *Proc. Natl Acad. Sci.* **120**, e2307289120 (2023).
- Rudman, S. M. et al. What genomic data can reveal about eco-evolutionary dynamics. *Nat. Ecol. Evol.* **2**, 9–15 (2018).
- Aguirre-Liguori, J. A., Ramirez-Barahona, S. & Gaut, B. S. The evolutionary genomics of species' responses to climate change. *Nat. Ecol. Evol.* **5**, 1350–1360 (2021).
- Borevitz, J. Utilizing genomics to understand and respond to global climate change. *Genome Biol.* **22**, 91 (2021).
- Vanneste, K., Baele, G., Maere, S. & Van de Peer, Y. Analysis of 41 plant genomes supports a wave of successful genome



- duplications in association with the Cretaceous–Paleogene boundary. *Genome Res.* **24**, 1334–1347 (2014).
44. Soltis, P. S. & Soltis, D. E. Ancient WGD events as drivers of key innovations in angiosperms. *Curr. Opin. Plant Biol.* **30**, 159–165 (2016).
45. Cai, L. et al. Widespread ancient whole-genome duplications in Malpighiales coincide with Eocene global climatic upheaval. *N. Phytologist* **221**, 565–576 (2019).
46. Benton, M. J., Wilf, P. & Sauquet, H. The Angiosperm Terrestrial Revolution and the origins of modern biodiversity. *N. Phytologist* **233**, 2017–2035 (2022).
47. Heimhofer, U., Hochuli, P. A., Burla, S., Dinis, J. M. L. & Weissert, H. Timing of Early Cretaceous angiosperm diversification and possible links to major paleoenvironmental change. *Geology* **33**, 141 (2005).
48. Schulte, P. et al. The Chicxulub asteroid impact and mass extinction at the Cretaceous–Paleogene boundary. *Science* **327**, 1214–1218 (2010).
49. Zhang, L. et al. The ancient wave of polyploidization events in flowering plants and their facilitated adaptation to environmental stress. *Plant Cell Environ.* **43**, 2847–2856 (2020).
50. Duke, N. C. et al. A world without mangroves? *Science* **317**, 41–42 (2007).
51. Tomlinson, P. B. *The Botany Of Mangroves* 2nd edn. (Cambridge Univ. Press, 2016).
52. 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).
53. He, Z. et al. De novo assembly of coding sequences of the mangrove palm (*Nypa fruticans*) using RNA-seq and discovery of whole-genome duplications in the ancestor of palms. *PLoS ONE* **10**, e0145385 (2015).
54. He, Z. et al. Convergent adaptation of the genomes of woody plants at the land–sea interface. *Natl Sci. Rev.* **7**, 978–993 (2020).
55. 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).
56. Xu, S. et al. Where whole-genome duplication is most beneficial: adaptation of mangroves to a wide salinity range between land and sea. *Mol. Ecol.* **32**, 460–475 (2023).
57. Feng, X. et al. Genomic insights into molecular adaptation to intertidal environments in the mangrove *Aegiceras corniculatum*. *N. Phytologist* **231**, 2346–2358 (2021).
58. Mandáková, T. & Lysak, M. A. Post-polyploid diploidization and diversification through dysploid changes. *Curr. Opin. Plant Biol.* **42**, 55–65 (2018).
59. Wendel, J. F. The wondrous cycles of polyploidy in plants. *Am. J. Bot.* **102**, 1753–1756 (2015).
60. Murat, F., Armero, A., Pont, C., Klopp, C. & Salse, J. Reconstructing the genome of the most recent common ancestor of flowering plants. *Nat. Genet.* **49**, 490–496 (2017).
61. Pont, C. et al. Paleogenomics: reconstruction of plant evolutionary trajectories from modern and ancient DNA. *Genome Biol.* **20**, 29 (2019).
62. He, Z. et al. Evolution of coastal forests based on a full set of mangrove genomes. *Nat. Ecol. Evol.* **6**, 738–749 (2022).
63. 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. *N. Phytologist* **217**, 428–438 (2018).
64. Petrov, D. A. Evolution of genome size: new approaches to an old problem. *Trends Genet.* **17**, 23–28 (2001).
65. Bennetzen, J. L. & Wang, H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu. Rev. Plant Biol.* **65**, 505–530 (2014).
66. Ou, S. & Jiang, N. LTR\_retriever: a highly accurate and sensitive program for identification of long terminal repeat retro-transposons. *Plant Physiol.* **176**, 1410–1422 (2018).
67. Kumar, S. et al. TimeTree 5: an expanded resource for species divergence times. *Mol. Biol. Evol.* **39**, msac174 (2022).
68. Shi, T. & Chen, J. A reappraisal of the phylogenetic placement of the *Aquilegia* whole-genome duplication. *Genome Biol.* **21**, 295 (2020).
69. Roelofs, D. et al. Multi-faceted analysis provides little evidence for recurrent whole-genome duplications during hexapod evolution. *BMC Biol.* **18**, 57 (2020).
70. Wang, Y. et al. An ancient whole-genome duplication event and its contribution to flavor compounds in the tea plant (*Camellia sinensis*). *Hortic. Res.* **8**, 176 (2021).
71. De Smet, R. & Van de Peer, Y. Redundancy and rewiring of genetic networks following genome-wide duplication events. *Curr. Opin. Plant Biol.* **15**, 168–176 (2012).
72. 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).
73. The French–Italian Public Consortium for Grapevine Genome Characterization. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**, 463–467 (2007).
74. Armarego-Marriott, T. Doubled genome an asset. *Nat. Clim. Chang* **10**, 184 (2020).
75. Stevens, A. V., Nicotra, A. B., Godfree, R. C. & Guja, L. K. Polyploidy affects the seed, dormancy and seedling characteristics of a perennial grass, conferring an advantage in stressful climates. *Plant Biol.* **22**, 500–513 (2020).
76. Bowers, J. E. & Paterson, A. H. Chromosome number is key to longevity of polyploid lineages. *N. Phytologist* **231**, 19–28 (2021).
77. Doyle, J. J. & Coate, J. E. Polyploidy, the nucleotype, and novelty: the impact of genome doubling on the biology of the cell. *Int. J. Plant Sci.* **180**, 1–52 (2019).
78. Yant, L. & Bombliks, K. Genome management and mismanagement—cell-level opportunities and challenges of whole-genome duplication. *Genes Dev.* **29**, 2405–2419 (2015).
79. Van de Peer, Y., Ashman, T.-L., Soltis, P. S. & Soltis, D. E. Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell* **33**, 11–26 (2021).
80. Yu, H. et al. A route to de novo domestication of wild allotetraploid rice. *Cell* **184**, 1156–1170.e14 (2021).
81. Levin, D. A. Minority cytotype exclusion in local plant populations. *Taxon* **24**, 35–43 (1975).
82. Burton, T. L. & Husband, B. C. Fitness differences among diploids, tetraploids, and their triploid progeny in *Chamerion angustifolium*: mechanisms of inviability and implications for polyploid evolution. *Evolution* **54**, 1182–1191 (2000).
83. Myburg, A. A. et al. The genome of *Eucalyptus grandis*. *Nature* **510**, 356–362 (2014).
84. Van de Peer, Y., Maere, S. & Meyer, A. The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **10**, 725–732 (2009).
85. Li, Z. et al. Patterns and processes of diploidization in land plants. *Annu. Rev. Plant Biol.* **72**, 387–410 (2021).
86. Sessa, E. B. Polyploidy as a mechanism for surviving global change. *N. Phytologist* **221**, 5–6 (2019).
87. Carretero-Paulet, L. & Van de Peer, Y. The evolutionary conundrum of whole-genome duplication. *Am. J. Bot.* **107**, 1101–1105 (2020).
88. Keightley, P. D. & Eyre-Walker, A. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics* **177**, 2251–2261 (2007).



89. Eyre-Walker, A. & Keightley, P. D. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol. Biol. Evol.* **26**, 2097–2108 (2009).
90. Schneider, A., Charlesworth, B., Eyre-Walker, A. & Keightley, P. D. A method for inferring the rate of occurrence and fitness effects of advantageous mutations. *Genetics* **189**, 1427–1437 (2011).
91. Chen, Q. et al. Two decades of suspect evidence for adaptive molecular evolution—negative selection confounding positive-selection signals. *Natl Sci. Rev.* **9**, nwab217 (2022).
92. Eilertson, K. E., Booth, J. G. & Bustamante, C. D. SnpPRE: selection inference using a Poisson random effects model. *PLoS Comput. Biol.* **8**, e1002806 (2012).
93. Shapiro, J. A. et al. Adaptive genic evolution in the *Drosophila* genomes. *Proc. Natl Acad. Sci.* **104**, 2271–2276 (2007).
94. Ball, M. C. Mangrove species richness in relation to salinity and waterlogging: a case study along the Adelaide River floodplain, northern Australia. *Glob. Ecol. Biogeogr. Lett.* **7**, 73 (1998).
95. Singh, A. Soil salinization and waterlogging: a threat to environment and agricultural sustainability. *Ecol. Indic.* **57**, 128–130 (2015).
96. Freeling, M. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* **60**, 433–453 (2009).
97. Sun, F. et al. Salt modulates gravity signaling pathway to regulate growth direction of primary roots in *Arabidopsis*. *Plant Physiol.* **146**, 178–188 (2008).
98. Zhang, Y., Xiao, G., Wang, X., Zhang, X. & Friml, J. Evolution of fast root gravitropism in seed plants. *Nat. Commun.* **10**, 3480 (2019).
99. Friml, J., Wiśniewska, J., Benková, E., Mendgen, K. & Palme, K. Lateral relocation of auxin efflux regulator PIN3 mediates tropism in *Arabidopsis*. *Nat.* **415**, 806–809 (2002).
100. Han, E. H., Petrella, D. P. & Blakeslee, J. J. ‘Bending’ models of halotropism: incorporating protein phosphatase 2A, ABCB transporters, and auxin metabolism. *J. Exp. Bot.* **68**, 3071–3089 (2017).
101. Gao, H.-B., Chu, Y.-J. & Xue, H.-W. Phosphatidic acid (PA) binds PP2AA1 to regulate PP2A activity and PIN1 polar localization. *Mol. Plant* **6**, 1692–1702 (2013).
102. Julkowska, M. M. & Testerink, C. Tuning plant signaling and growth to survive salt. *Trends Plant Sci.* **20**, 586–594 (2015).
103. Skubacz, A., Daszkowska-Golec, A. & Szarejko, I. The role and regulation of ABI5 (ABA-Insensitive 5) in plant development, abiotic stress responses and phytohormone crosstalk. *Front. Plant Sci.* **7**, 1884 (2016).
104. Hundertmark, M. & Hinch, D. K. LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics* **9**, 118 (2008).
105. Battaglia, M., Olvera-Carrillo, Y., Garciarrubio, A., Campos, F. & Covarrubias, A. A. The enigmatic LEA proteins and other hydrophilins. *Plant Physiol.* **148**, 6–24 (2008).
106. Abe, H. et al. *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell* **15**, 63–78 (2003).
107. Doyle, J. J. & Doyle, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* **19**, 11–15 (1987).
108. Vurture, G. W. et al. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
109. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* **17**, 155–158 (2020).
110. Chin, C.-S. et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569 (2013).
111. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
112. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
113. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
114. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
115. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
116. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* **25**, 1–14 (2009).
117. Xu, Z. & Wang, H. LTR\_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
118. Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
119. Lomsadze, A. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res.* **33**, 6494–6506 (2005).
120. Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinform.* **6**, 31 (2005).
121. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
122. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
123. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EvidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, R7 (2008).
124. Zheng, Y. et al. iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol. Plant* **9**, 1667–1670 (2016).
125. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
126. Lamesch, P. et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40**, D1202–D1210 (2012).
127. Shi, T. et al. Distinct expression and methylation patterns for genes with different fates following a single whole-genome duplication in flowering plants. *Mol. Biol. Evol.* **37**, 2394–2413 (2020).
128. Luo, X. et al. The pomegranate (*Punica granatum* L.) draft genome dissects genetic divergence between soft- and hard-seeded cultivars. *Plant Biotechnol. J.* **18**, 955–968 (2020).
129. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
130. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
131. 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).
132. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
133. 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).
134. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
135. Dos Reis, M. & Yang, Z. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Mol. Biol. Evol.* **28**, 2161–2172 (2011).
136. Morris, J. L. et al. The timescale of early land plant evolution. *Proc. Natl Acad. Sci.* **115**, E2274–E2283 (2018).
137. Graham, S. A. Fossil records in the Lythraceae. *Botanical Rev.* **79**, 48–145 (2013).
138. Coiro, M., Doyle, J. A. & Hilton, J. How deep is the conflict between molecular and fossil evidence on the age of angiosperms? *N. Phytologist* **223**, 83–99 (2019).
139. Zhang, L. et al. The water lily genome and the early evolution of flowering plants. *Nature* **577**, 79–84 (2020).
140. 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).
141. Wang, Y. et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
142. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
143. Wang, D., Zhang, Y., Zhang, Z., Zhu, J. & Yu, J. KaKs\_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteom. Bioinform.* **8**, 77–80 (2010).
144. Li, Z. et al. Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proc. Natl Acad. Sci.* **115**, 4713–4718 (2018).
145. Stolzer, M. et al. Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics* **28**, i409–i415 (2012).
146. Rabier, C.-E., Ta, T. & Ané, C. Detecting and locating whole genome duplications on a phylogeny: a probabilistic approach. *Mol. Biol. Evol.* **31**, 750–762 (2014).
147. Mendes, F. K., Vanderpool, D., Fulton, B. & Hahn, M. W. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **36**, 5516–5518 (2021).
148. Clark, J. W. & Donoghue, P. C. J. Constraining the timing of whole genome duplication in plant evolutionary history. *Proc. R. Soc. B: Biol. Sci.* **284**, 20170912 (2017).
149. Glick, L. & Mayrose, I. ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. *Mol. Biol. Evol.* **31**, 1914–1922 (2014).
150. Graham, S. A., Oginuma, K., Raven, P. H. & Tobe, H. Chromosome numbers in *Sonneratia* and *Duabanga* (Lythraceae s.l.) and their systematic significance. *Taxon* **42**, 35–41 (1993).
151. Rice, A. et al. The Chromosome Counts Database (CCDB)—a community resource of plant chromosome numbers. *N. Phytologist* **206**, 19–26 (2015).
152. Sun, P. et al. WGDl: a user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Mol. Plant* **15**, 1841–1851 (2022).
153. Salse, J. Ancestors of modern plant crops. *Curr. Opin. Plant Biol.* **30**, 134–142 (2016).
154. Tang, H. et al. Synteny and collinearity in plant genomes. *Science* **320**, 486–488 (2008).
155. Cox, M. P., Peterson, D. A. & Biggs, P. J. SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinform.* **11**, 485 (2010).
156. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
157. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
158. Gu, K., Ng, H. K. T., Tang, M. L. & Schucany, W. R. Testing the ratio of two Poisson rates. *Biometrical J.* **50**, 283–298 (2008).
159. Roulin, A. et al. The fate of duplicated genes in a polyploid plant genome. *Plant J.* **73**, 143–153 (2013).
160. Liu, S. et al. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat. Commun.* **5**, 3930 (2014).
161. Feng, X. et al. Expansion and adaptive evolution of the *WRKY* transcription factor family in *Avicennia* mangrove trees. *Mar. Life Sci. Technol.* **5**, 155–168 (2023).
162. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
163. 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).
164. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

## Acknowledgements

The authors thank Zuyao Liu and Nan Wang for their technical support. The project was supported by the National Natural Science Foundation of China (32170230 to Z.H., 31971540 to Z.H. and 32330005 to S. Shi), the Guangdong Basic and Applied Basic Research Foundation (2023B1515020083 to Z.H.), the Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (311021006 to S. Shi), the Shenzhen Science and Technology Innovation Program (RCBS20221008093316043 to Q.C.), and the China Postdoctoral Science Foundation (2023M740690 to X.F.).

## Author contributions

Z.H. and S. Shi conceived the study. X.F., S. Shi, and Z.H. designed and conceptualized the study. X.F., Q.C., W.W., J.W., G.L., and Z.H. performed the data analysis. X.F., Q.C., G.L., S.X., C.Z., S. Shi, and Z.H. collected materials. X.F., Q.C., S.X., S. Shao, and M.L. performed the experiments. X.F., Q.C., S. Shi, and Z.H. wrote the manuscript. C.-I.W. revised the manuscript. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-46080-7>.

**Correspondence** and requests for materials should be addressed to Suhua Shi or Ziwen He.

**Peer review information** *Nature Communications* thanks Huan Liu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024