the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2024) **117**, 432–448 doi: 10.1111/tpj.16499

Evolution of sea-surfing plant propagule as revealed by the genomes of *Heritiera* mangroves

Jiayan Wang[1], Wei Xie[1,2], Fa Si[1], Ziwen He[1], Xinfeng Wang[1], Shao Shao[1], Suhua Shi[1,*] and Zixiao Guo[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, Guangzhou, Guangdong, China, and* [2]*School of Life Sciences, Guizhou Normal University, Guiyang, Guizhou, China*

Received 3 May 2023; revised 13 August 2023; accepted 4 October 2023; published online 18 October 2023.
*For correspondence (e-mail guozx8@mail.sysu.edu.cn, lssssh@mail.sysu.edu.cn).

SUMMARY

Coastal forests, such as mangroves, protect much of the tropical and subtropical coasts. Long-distance dispersal via sea-surfing propagules is essential for coastal plants, but the genomic and molecular basis of sea-surfing plant propagule evolution remains unclear. *Heritiera fomes* and *Heritiera littoralis* are two coastal plants with typical buoyant fruits. We *de novo* sequenced and assembled their high-quality genomes. Our phylogenomic analysis indicates *H. littoralis* and *H. fomes* originated (at ~6.08 Mya) just before the start of Quaternary sea-level fluctuations. Whole-genome duplication occurred earlier, permitting gene copy gains in the two species. Many of the expanded gene families are involved in lignin and flavonoid biosynthesis, likely contributing to buoyant fruit emergence. It is repeatedly revealed that one duplicated copy to be under positive selection while the other is not. By examining *H. littoralis* fruits at three different developmental stages, we found that gene expression levels remain stable from young to intermediate. However, ~1000 genes are up-regulated and ~ 3000 genes are down-regulated as moving to mature. Particularly in fruit epicarps, the upregulation of *WRKY12* and *E2Fc* likely constrains the production of p-Coumaroyl-CoA, the key internal substrate for lignin biosynthesis. Hence, to increase fruit impermeability, methylated lignin biosynthesis is shut down by down-regulating the genes *CCoAOMT*, *F5H*, *COMT*, and *CSE*, while unmethylated lignins are preferentially produced by upregulating *CAD* and *CCR*. Similarly, cutin polymers and cuticular waxes accumulate with high levels before maturation in epicarps. Overall, our genome assemblies and analyses uncovered the genomic evolution and temporal transcriptional regulation of sea-surfing propagule.

Keywords: coastal plant, comparative genomics, *Heritiera*, lignin biosynthesis, sea-surfing propagule, transcriptional regulation, genome structure and evolution.

INTRODUCTION

Over 2.6 billion people live in coastal regions, defined as 100 km strips adjacent to coasts (Sale et al., 2014). Coastal forests are of extraordinary high ecological values in protecting coastal regions from typhoons and tsunamis in the short term and the impacts of sea-level rise in the long term. As the threat of extreme climate events and sea-level rise intensifies, studies of the construction and maintenance of coastal shelter forests are increasing in relevance. Mangrove forests, well-known as "coast guards", are the most important coastal forests in tropical and subtropical regions, where typhoons and tsunamis are prevalent.

Many coastal plants, including the majority of mangrove species, produce propagules that disperse via seawater. The propagules, such as fruits, seeds, or viviparous hypocotyls, disperse across open seas by floating in seawater and moving with sea currents (sea-surfing). These traits are essential for the growth and maintenance of these plants. Long-distance dispersal has allowed gene flow among populations on distant coasts over evolutionary time, introducing evolutionary innovations that may have been useful for these species' adaptive success. At a shorter timescale, propagule dispersal is the key for a species to shift its distribution range in response to environmental changes (Guo et al., 2016, 2018; Van Der Stocken, Carroll, et al., 2019; Zhang et al., 2022).

To adapt to seawater hydrochory, sea-surfing coastal plants have evolved a set of morphological and physiological features in propagules (Tomlinson, 2016; Van Der Stocken, Wee, et al., 2019). *Heritiera* spp. have evolved a fibrous mesocarp and woody epicarp. *Sonneratia* spp., *Xylocarpus* spp., and *Nypa fruticans* have evolved corky testa. *Rhizophora* spp. have evolved spongy hypocotyls with air-filled walls. Due to the different morphological

adaptations, their long-distance dispersal abilities are remarkably variant. For example, *Aegiceras corniculatum* fruits are almost not buoyant, while *H. littoralis* fruit and *Rhizophora* spp. hypocotyls can be buoyant in seawater for months (Van Der Stocken, Wee, et al., 2019). For a successful long-distance dispersal, the most essential part is keeping the seeds viable throughout long-term seawater immersion until germination at the new location.

We set out to elucidate the genomic and molecular basis underlying the origin and evolution of seawater-dispersing propagules in coastal plants. We investigated the question by conducting a multi-omics study into the genomes of *H. fomes* and *H. littoralis* (family Malvaceae), as well as the transcriptome profiles of *H. littoralis* fruit at different development stages. *H. fomes* and *H. littoralis* are mangrove species with the most typical of seawater-dispersing propagules. Of the ~40 species in the *Heritiera* genus (POWO, 2022), only these two species and *H. globosa* have invaded the intertidal zones and become mangroves. Most *Heritiera* species have samara-style fruit, adapted for wind dispersal, while the three mangroves have reduced their "wing" to a vestigial "keel" (Tomlinson, 2016).

The strong long-distance dispersal ability of *H. littoralis* and *H. fomes* could mainly be attributed to the fruits with hard woody epicarp, fibrous mesocarp, and air spaces (Tomlinson, 2016; Van Der Stocken, Wee, et al., 2019). The fibrous mesocarp and air spaces decrease fruit density and allow it to float. In addition, highly lignified epicarp keeps the seeds viable throughout potentially long sea journeys. Lignins often accumulate in cell walls to prevent permeability to seawater and also contribute to inhibition of seed germination (Pierce et al., 2019). Lignins, Phenylpropanoid-based polymers, are also required for mechanical support of plant growth. The fruit is also coated with a developed cuticle, mainly composed of cutin and waxes, that protects them from water loss through transpiration, high light intensity, drought, pathogens, insect herbivores, and toxic chemicals. In addition, tannins, a collection of flavonoid polymers, have antioxidant activity and play a role in photoprotection, pathogen, and herbivory resistance, as well as salt tolerance.

We sequenced and analyzed the genomes of *H. fomes* and *H. littoralis*. We also captured and compared the transcriptomes of *H. littoralis* fruits at the young, intermediate, and mature developmental stages, to explore the molecular mechanisms of constructing the highly lignified fruit that are coated with cuticles. Particular attention was paid to the evolution of genes involved in lignin, tannin, and cutin and wax biosynthesis. Combining both genomic and transcriptomic analyses, we advance the understanding of the genetic changes underlying the evolution of plant seawater dispersal.

RESULTS

High-quality *H. fomes* and *H. littoralis* genome assemblies

We obtained 440.19 Gb of long reads for *H. fomes* and 160.23 Gb for *H. littoralis* using the PacBio SMRT sequencing technology (Table S1). The final assemblies are 817.96 and 818.66 Mb in length for *H. fomes* and *H. littoralis*, respectively. These numbers are close to their K-mer estimated genome sizes of 820.61 and 823.59 Mb (Figure 1a; Figure S1, Table 1) as well as the flow cytometry-estimated *H. littoralis* genome size of 832 Mb (He et al., 2022). The *H. fomes* assembly is in 617 scaffolds (scaffold N50 = 37.5 Mb) and the *H. littoralis* assembly is in 277 scaffolds (scaffold N50 = 5.09 Mb; Figures 1a,b; Table 1). The 25 long scaffolds of *H. fomes* (> 10 Mb and other short scaffolds are <0.8 Mb) cover 96.8% of its total sequences. Both of the *Heritiera* species contain 38 chromosomes ($2n = 38$) (Das et al., 2001), therefore, our *H. fomes* assembly approaches chromosome level. The *H. littoralis* assembly is more fragmented, with the 19 longest scaffolds (8.4–25.4 Mb) covering 28.0% of its total sequences. BUSCO (benchmarking universal single-copy orthologs) mapping indicates 98.1% and 96.6% of the complete eukaryotic conserved genes were found in the two assemblies (Table 1; Table S2). Moreover, 98.6% and 96.1% of the long reads (Table S1), 96.6% and 95.3% of the short reads (Table S3) were successfully mapped back to the assembled genomes of *H. fomes* and *H. littoralis*, respectively. These assessments reflect the good quality of the two assemblies.

We predicted 33 189 protein-coding genes in *H. fomes* and 38 961 in *H. littoralis* (Figure 1b and Table 1; Table S4). Repeat sequences occupy 72.1% and 70.8% of these genomes, with long terminal repeats retrotransposons (LTR-RTs) representing the largest fractions (45.3% and 45.2%; Table 1; Table S5). The Gypsy element is the most abundant LTR-RT in both genomes, followed by the Copia element (Figure 1c; Table S5). Interestingly, most LTR-RTs were inserted into the genomes of both species more recently than one million years ago (Figure 1d). The parallel insertion of LTR-RTs in the genomes was the most intensive at very young ages ~65 (*H. fomes*) and ~ 67 (*H. littoralis*) thousand years ago (Figure 1d), which coincides with the Last Glacial Maximum. The LTR-RT activities may have contributed to their survival through the drastic environmental changes in the Last Glacial Maximum.

Ancient genome polyploidization predated the origin of *Heritiera* mangroves

To investigate phylogenomic origin of these two *Heritiera* species, the transcriptomes of three closely related species (*Heritiera angustata*, *Heritiera parvifolia*, and *Sterculia monosperma*) were sequenced and assembled *de novo* (Figure S2 and Table S6), and genome sequences of *Durio*

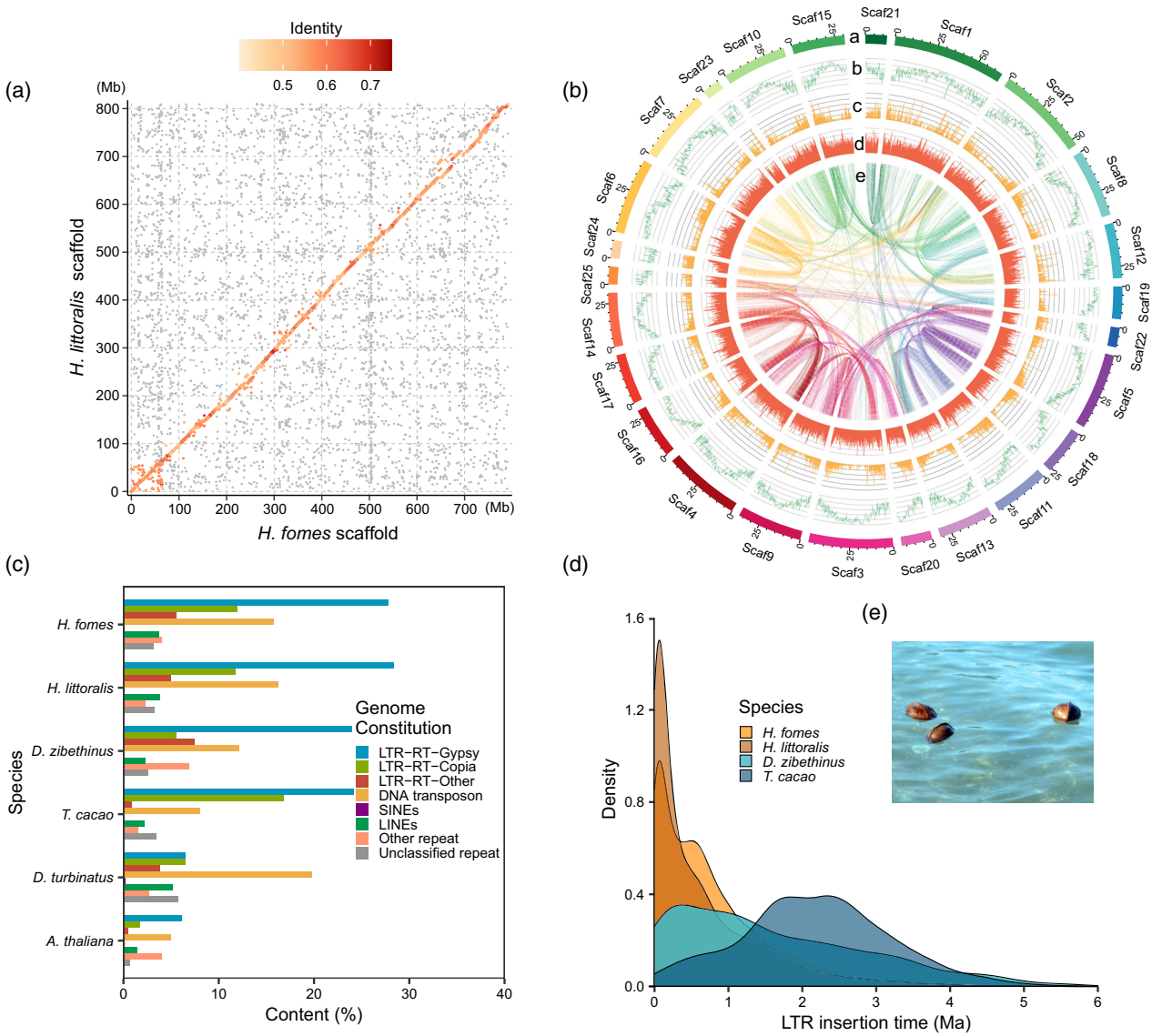

Figure 1. *Heritiera fomes* and *H. littoralis* genome assemblies and evolution of long terminal repeat retrotransposons (LTR-RTs).
(a) Dot plot showing scaffold synteny between *H. fomes* and *H. littoralis*.
(b) The genomic landscape of *H. fomes*. Tracks displayed are: a, assembled scaffolds; b, repeat density; c, gene density; d, GC content; e, intragenomic synteny. The 25 scaffolds longer than 10 M are presented, covering ~96% of the assembly.
(c) Repeat sequences in the *Heritiera* and four related species.
(d) Distributions of LTR-RT ages in the two *Heritiera* and two related species.
(e) Photo of *H. littoralis* fruits floating on seawater. The photo was taken by the first author in Shenzhen, China.

zibethinus, *Theobroma cacao*, *Dipterocarus turbinatus* and *Arabidopsis thaliana* were downloaded (Table S7). We clustered the genes from these nine species into homologous groups. A homologous group containing only one copy from each of the nine genomes is defined as a single-copy orthologous group. We identified 658 single-copy orthologous groups and used them to construct a phylogenetic tree of these species (Figure S3). With fossil calibrations (Table S8), molecular dating suggests that *H. fomes* diverged with *H. littoralis* ~ 5.02 Mya. Their most recent common ancestor diverged from the terrestrial relative *Heritiera angustata* in the late Miocene (at ~6.08 Mya, Figures 2a,b), just before the Quaternary cooling and sea-level fluctuation. The *Heritiera* genus diverged from its sister genus *Sterculia* ~ 18.59 Mya (Figures 2a,b).

Whole-genome duplication (WGD) has been suggested to be important for gene gains and functional innovation. We explored the newly sequenced genomes for collinear blocks that may be relics of ancient WGD events. We define collinear blocks as these blocks with arrays of five or more collinear pairs of homologous genes. We found 394 collinear blocks in *H. fomes* and 630 in *H.*

Table 1 Sequencing, assembly, and annotation of *Heritiera fomes* and *Heritiera littoralis* genomes

	H. fomes	*H. littoralis*
Sequencing technology	SMRT(HiFi)	SMRT
Expected genome size (Mb, K-mer)	820.61	823.59
Assembly size (Mb)	817.96	818.66
Total scaffold number	617	277
Longest Scaffold (Mb)	64.86	25.39
Scaffold N50 (Mb)	37.50	5.09
GC content (%)	34.43	33.86
Heterozygosity (%)	0.3077	0.3071
Repeat content (%)	72.05	70.85
BUSCOs completely mapped (%)	98.1	96.6
Gene number/total length (Mb)	33 189/127.49	38 961/151.59
Average gene length (bp)	3841	3891
Exon number/total length (Mb)	176 723/40.20	210 080/45.71
Average exon length (bp)	227	218
Intron number/total length (Mb)	143 534/87.44	171 119/106.04
Average intron length (bp)	609	620

littoralis. These blocks comprise ~18 000 genes in both species, occupying 54.0% of total *H. fomes* genes and 44.5% of total *H. littoralis* genes (Table S9). There are more such genes located in collinear blocks in *D. zibethinus* (67.71%), likely due to that it has undergone a whole-genome triplication (WGT) (Wang et al., 2019) instead of a WGD. In contrast, much smaller percentages of genes of *T. cacao* (16.32%) and *A. thaliana* (27.97%) are located in collinear blocks (Table S9).

If these collinear blocks were relics of WGD events, the genes they contain would have been duplicated at the same time and thus would have similar synonymous substitution rates (Ks values). Ks value distributions estimated in the four *Heritiera* species, *S. monosperma*, and *D. zibethinus* genomes are bimodal (Figure 3c). The more ancient low and flat peak with Ks between 1.4 and 2.2 (Figure 3c) is consistent with the gamma event, an ancient WGT event shared by all core eudicots (Murat et al., 2017). The younger and sharper peak (Ks between 0.19 and 0.27) corresponds to a more recent duplication event. *T. cacao* has undergone no post-gamma WGD and, as expected, does not have this peak in its Ks value distribution. *D. zibethinus* is known to have experienced a lineage-specific WGT (Wang et al., 2019), and we also observed 2:3 copy number ratios among *H. fomes* and *D. zibethinus* in 416 sets of syntenic genome blocks (Figure 3b,d). In contrast, we observed 241 *H. fomes* genome blocks in synteny with *T. cacao*, with a ratio of *H. fomes*: *T. cacao* = 2:1 (Figure 3a,d). Therefore, the duplication event corresponding to the younger Ks peak was inferred to have occurred after the divergence between the *D. zibethinus* and the common ancestor of *Heritiera* and *Sterculia* (Figure 2).

Gene copy number evolution and positive selection detection

We clustered genes annotated in the two *Heritiera* species, *D. zibethinus*, *T. cacao*, *D. turbinatus*, and *A. thaliana* into 20 302 homologous groups using OrthoFinder. We used CAFE to identify 1635 groups that significantly expanded on the *H. fomes* branch and 1328 that have contracted (*P*-value <0.05, Figure 2b). We also found 1985 expanded and 2596 contracted groups in *H. littoralis* (*P*-value <0.05, Figure 2b). In addition, we found 1517 expanded and 791 contracted groups in the most recent common ancestor of *H. fomes* and *H. littoralis*. The genes in the expanded groups of the common ancestor were disproportionally represented by GO terms such as proton export across plasma, cellular response to potassium ion starvation, cell death in response to hydrogen peroxide, and regulation of circadian rhythm (Table S10). Expansion in these gene groups may be beneficial for their adaptation to intertidal habitats with harsh conditions of highly salinity, anoxia, and tidal flooding.

By providing genetic material for adaptive evolution, copy number expansion of homologous gene groups might have played important roles in *Heritiera* species' adaptation to intertidal zones. We used DupGen_finder to investigate how the expansions occurred in these taxa. We identified 27 795 duplicated genes (genes with at least one homolog within the genome) in the *H. fomes* genome and 32 246 in *H. littoralis*. The excess of duplicated genes in *H. littoralis* has significantly contributed to the larger gene counts in its genome.

Duplicated genes were classified into five categories according to duplication mechanisms: whole genome (WGD), tandem (TD), transposed (TRD), proximal (PD), and dispersed duplicates (DSD). In both *H. fomes* and *H. littoralis*, the WGD contributed approximately half of the gene gain from duplications, while each of the other four mechanisms contributed to ~10% (Figure 4a; Table S11). We also observed higher Ka/Ks values between pairs of proximal and tandem duplicated genes, in comparison to the genes duplicated by the other three mechanisms (Figure S4).

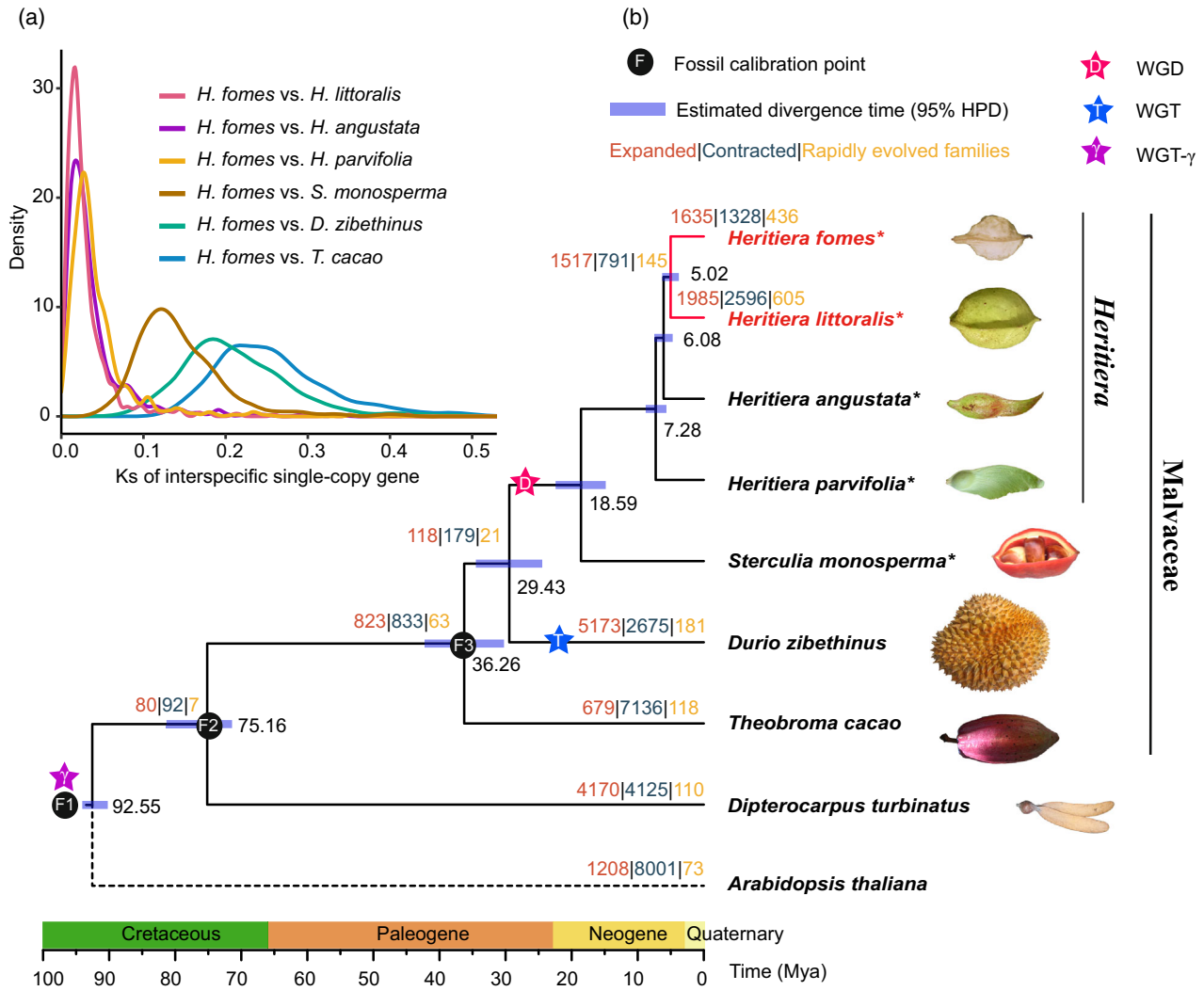

Figure 2. Phylogeny of the *Heritiera* and related species and evolution of homologous gene groups.
(a) Distributions of species-pairwise synonymous substitution rates (Ks) estimated from single-copy homologs of *Heritiera fomes*, *H. littoralis*, and five-related species.
(b) The dated phylogeny of *Heritiera* species and their relatives. Red branches indicate mangrove species. Blue bars at nodes represent 95% confidence intervals of node ages. Black nodes indicate the three nodes calibrated using fossil evidence. Stars on internal branches indicate the inferred polyploidization events (whole-genome duplication/whole-genome triplication, WGD/WGT). Red, blue, and yellow numbers on branches indicate numbers of expanded, contracted, and rapidly evolving gene groups. Asterisks mark the species we sequenced genome or transcriptome *de novo*. The fruit photos are provided by the authors or the contributors named in the ACKNOWLEDGEMENT section. To save figure space, the photos are not shown in real fruit sizes.

To determine the WGD contribution to gene family expansion we looked at the representation of genes generated by this event in the expanded gene families (Figure 4b; Table S12). We found 4681 such loci in *H. littoralis*. These genes disproportionally belong to phenylpropanoid, lignin, and flavonoid biosynthetic/metabolic process GO categories, as well as root cap development (Figure 4d; Table S13). Moreover, expansions of telomere organization and RNA-mediated transposition gene groups mainly resulted from DSD (Figure 4d; Table S13). In comparison, gene families that expanded in *H. fomes* contain 3247 genes that were duplicated during the WGD event (Figure 4b; Table S12). Seed dormancy, nodulation, and negative regulation of response to salt stress are the GO terms overrepresented among this set of loci (Figure 4d; Table S14). Notably, 788 proximately duplicated genes belong to expanded gene groups. They disproportionally affect the regulation of root development and sesquiterpenoid catabolic processes (Figure 4d; Table S14).

Gene duplication may have contributed to the origin of sea-surfing fruits of the two *Heritiera* species. According to previous literature (Dong & Lin, 2021; Fich et al., 2016; Kong et al., 2020; Lewandowska et al., 2020; Xie et al.,

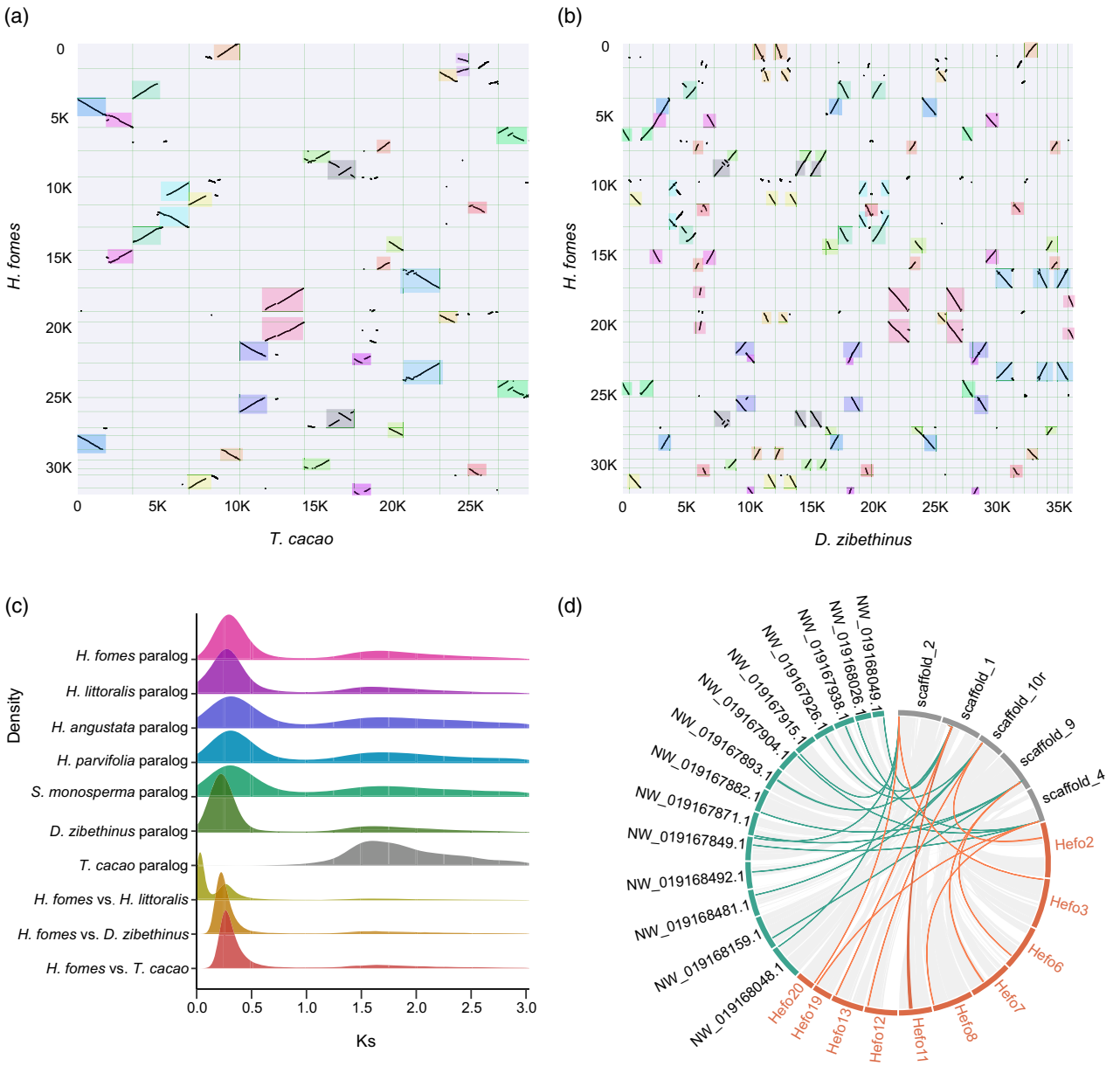

Figure 3. Evidence indicating whole-genome duplication in the *Heritiera* lineage.
(a) Dot plot indicates homologous (syntenic) scaffolds between *H. fomes* and *T. cacao* (2:1).
(b) Dot plot indicates homologous (syntenic) scaffolds between *H. fomes* and *D. zibethinus* (2:3).
(c) Ks distributions of intraspecific paralogs and interspecific homologs. (d) Diagram of syntenic genome blocks. Red links indicate collinear gene pairs with a 1:2 ratio between *T. cacao* and *H. fomes*. Green links indicate collinear gene pairs with a 1:3 ratio between *T. cacao* and *D. zibethinus*. Red, green, and gray block rectangles indicate genome scaffolds of *H. fomes*, *D. zibethinus*, and *T. cacao*, respectively.

2018; Yeats & Rose, 2013), we summarized the enzyme proteins and encoding genes involved in lignin (Table S15), tannin (Table S16), as well as cutin and wax (Table S17) biosynthesis. In the *H. fomes* and *H. littoralis* genomes, we identified homologous genes encoding 76 of these enzyme proteins. Twenty-three of the proteins are encoded by genes with duplicated paralogs. Of the 99 orthologous single-copy gene clusters encoding these 76 proteins, 18 appear to be under positive selection in *H. fomes* (nine genes originating from duplication events) and eight in *H. littoralis* (four genes of them duplicated, Table S18). Twenty genes show signs of adaptive evolution in the most recent common ancestor of the two species (eight of them duplicated, Table S18). All told, the genes of 23 proteins (10 from duplicated genes) appear to have been under positive selection on at least one branch. Interestingly, it is repeatedly observed that one copy

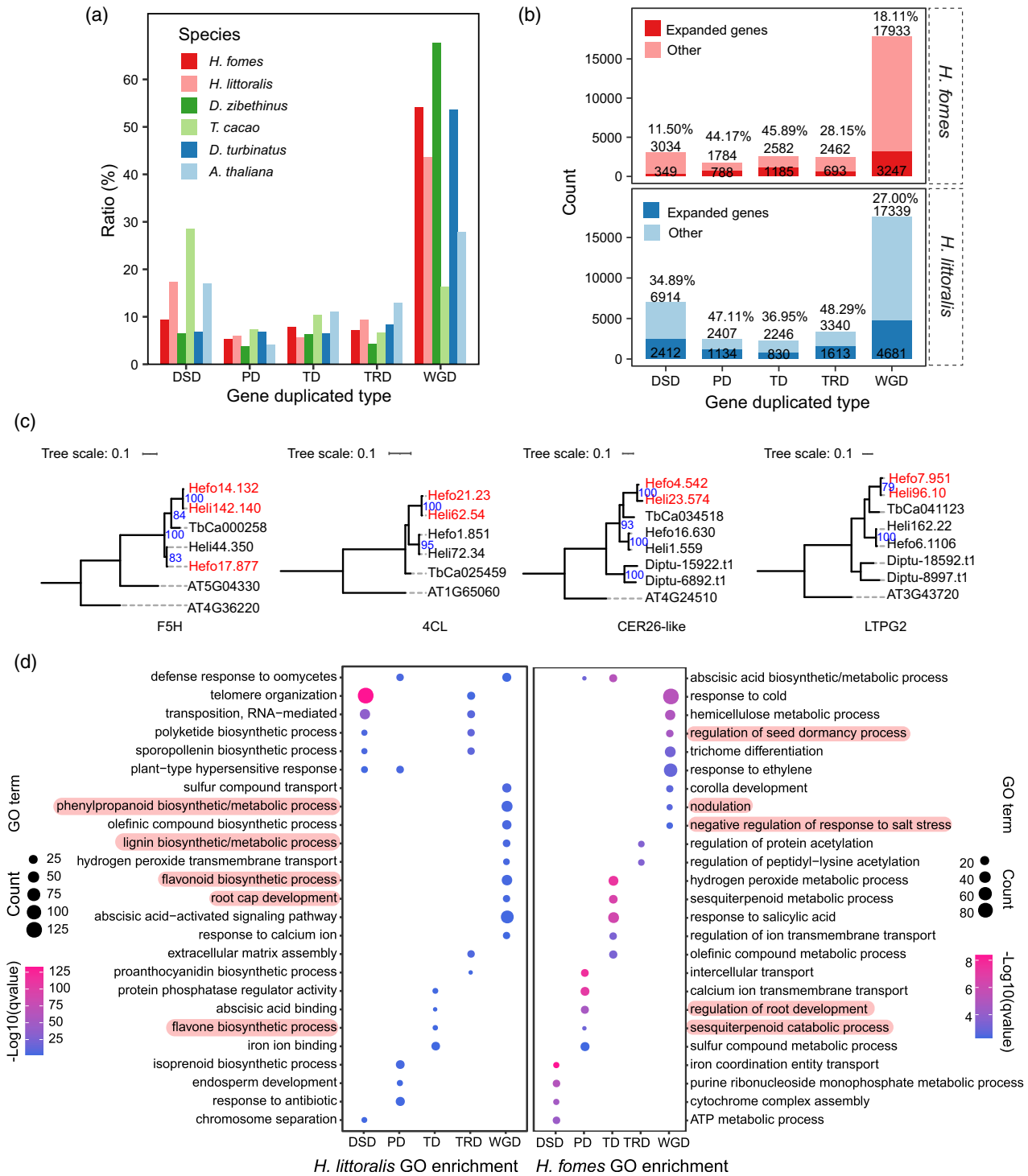

Figure 4. Gene duplication and positive selection.
(a) Bar plots showing ratios of genes duplicated using different mechanisms.
(b) Expanded gene groups (identified by CAFE) that have duplicated using various mechanisms. On each bar, the numbers from bottom to up are the number of genes included in expanded gene groups (Expanded), the total number of duplicates from the corresponding mechanism (Duplicated), and the percentage of Expanded/Duplicated.
(c) Cases of duplicated genes under differential natural selection. Gene name in red indicates the gene copy under positive selection. Blue numbers on branches indicate bootstrap support rates of the gene tree.
(d) Functional enrichment of expanded gene groups originated from various gene duplication mechanisms. DSD: dispersed duplicates; PD: proximal duplicates; TD: tandem duplicates; TRD: transposed duplicates; WGD: whole-genome duplicates.

among each of the 10 duplicates shows signs of adaptive evolution while the other copy not, suggesting selection for novel functions (Figure 4c; Figure S5).

These results suggest that the recent whole-genome duplication has significantly contributed to gene gain and gene family expansion in the two *Heritiera* species. Particularly, the expansion of gene families involved in lignin and flavonoid biosynthetic and metabolism may have been important for the origin of sea-surfing fruits. Interestingly, the duplicated copies of genes involved in lignin, tannin, and cutin and wax biosynthesis are under differential natural selection, implying the potential selective usage of duplicated genes in evolving sea-surfing fruit.

Identification of differentially expressed genes in the *Heritiera* fruit development process

We used RNA-seq to measure gene expressions in *H. littoralis* epicarps and seeds at young, intermediate, and

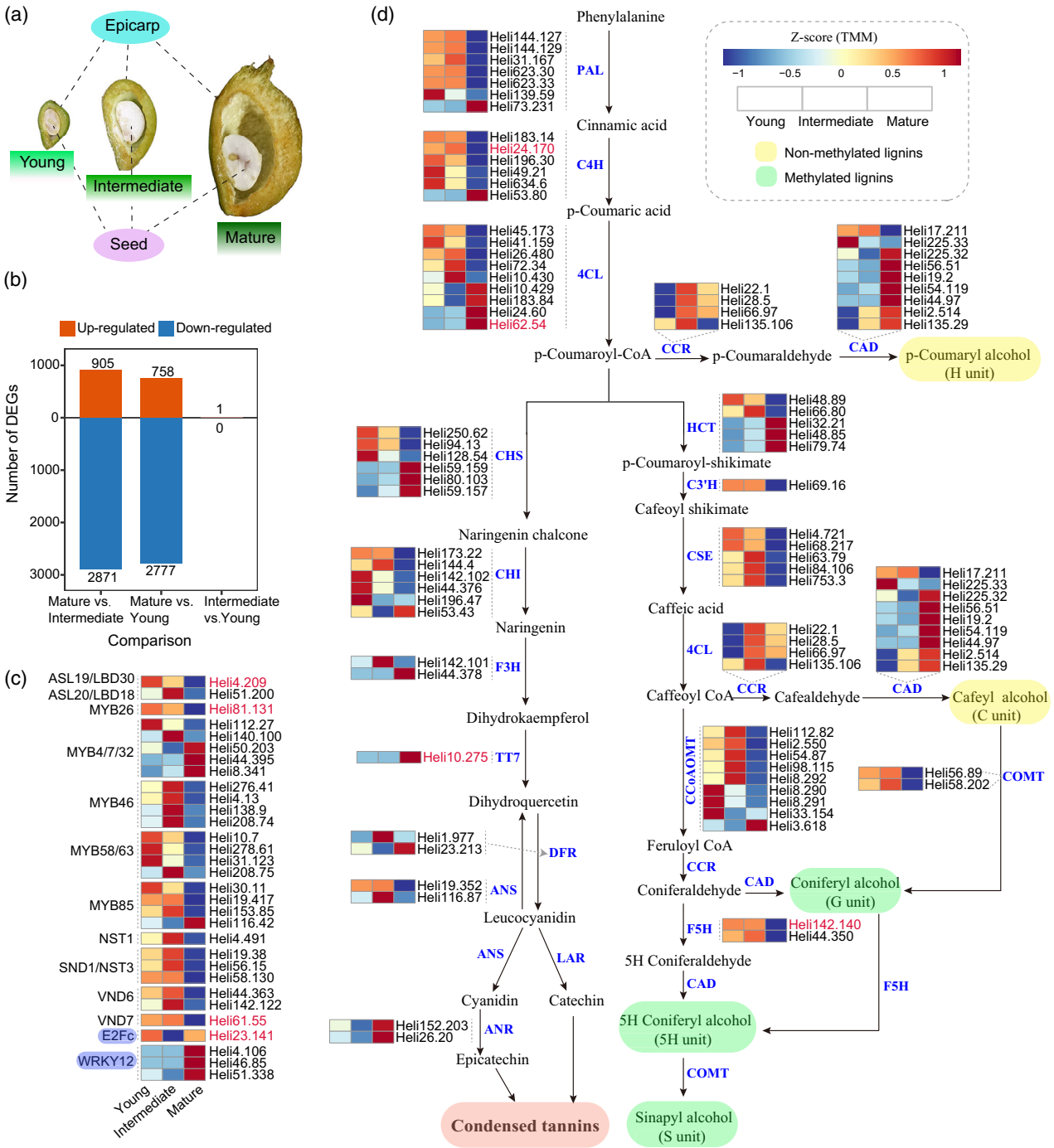

Figure 5. Expression profiles of genes in the lignins and condensed tannins biosynthesis pathways in *Heritiera littoralis*.
(a) *H. littoralis* fruit phenotypes at different developmental stages.
(b) Number of genes differentially expressed across developmental stages in *H. littoralis* fruit epicarps. Red and blue bars represent significantly differentially expressed genes (with adjusted *P*-value <0.05 and fold change >2). For example, up-regulated genes of Mature vs. Intermediate have higher expression at the Mature stage than at the Intermediate stage.
(c) Expression profile of the transcription factors regulating lignin biosynthesis in *H. littoralis*.
(d) Expression profiles of genes in lignin and condensed tannin biosynthesis pathways. Gene expression was scaled using the Z-scores of TMM (mean value of three biological replicates). In (c) and (d), gene names in red indicate the genes under positive selection. *PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate-CoA ligase; CCR, cinnamoyl-CoA reductase; CAD, cinnamyl alcohol dehydrogenase; C3H, coumarate 3-hydroxylase; HCT, hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase; CSE, caffeoyl shikimate esterase; CCoAOMT, caffeoyl-CoA 3-O-methyltransferase; COMT, caffeate/5-hydroxyferulate 3-O-methyltransferase; F5H, ferulate-5-hydroxylase; CHS, Chalcone synthase; CHI, Chalcone isomerase; F3H, Flavanone 3-hydroxylase; TT7, cytochrome P450 superfamily protein; DFR, Dihydroflavonol-4-reductase; ANS, Leucoanthocyanidin dioxygenase; LAR, Leucoanthocyanidin reductase; ANR, Anthocyanidin reductase.*

mature fruit stages (Figure 5a; Table S19). We estimated the reproducibility of our measurements by using three biological replicates for each stage and tissue. High among-sample Pearson's correlations suggest that our experiments were well controlled (Figure S6). Expressions of genes involved in lignin (Table S15), tannin (Table S16), cutin polymer, and cuticular wax (Table S17) biosynthesis pathways were particularly quantified and analyzed in the later sections.

We found few expression changes between the young and intermediate stage epicarps, suggesting that transcriptome changes do not drive this development interval (Figure 5b; Figure S7 and Table S20). In contrast, we found 905 up-regulated and 2871 down-regulated genes between the intermediate and mature stages (with adjusted *P*-value <0.05 and fold change >2, Figure 5b; Figure S7 and Table S20). As expected, similar numbers of genes changed their expression between the young and mature stages, given similar transcriptome profiles between young and intermediate epicarps (with adjusted *P*-value <0.05 and fold change >2, Figure 5b; Figure S7 and Table S20). These differentially expressed genes are disproportionally involved in phenylpropanoid biosynthesis, cutin, suberin, and wax biosynthesis according to the KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analysis (Figure S8 and Table S21). Hence, gene expression changes drastically in epicarps as fruit transit to maturation, affecting particularly the biosynthesis of lignin, cutin, and wax.

The pattern is very similar in seeds, where only three genes are down-regulated moving from the young to intermediate stage, while 1239 were up-regulated and 1531 down-regulated between the intermediate and mature stages (Figure S7 and Table S20). The young-to-mature stage comparison again confirmed this result. These differentially expressed genes are primarily involved in plant hormone signal transduction, as well as starch and sucrose metabolism (Figure S8). It thus appears that while both seed and epicarp transcriptome are extensively remodeled during fruit maturation, the processes involved are tissue-specific. Gene expression changes in seeds are less related to fruit impermeability.

Preferred unmethylated lignin biosynthesis enhances fruit impermeability

Lignins are the basic materials in the structures protecting seeds from seawater. Expression of several genes regulating key steps in the lignin biosynthetic pathway changes during epicarp development (Figure 5d; Figures S9 and S10, Tables S20 and S21). *CCoAOMT* (*caffeoyl-CoA 3-O-methyltransferase*), *F5H* (*ferulate-5-hydroxylase*), *COMT* (*caffeate/5-hydroxyferulate 3-O-methyltransferase*), and *CSE* (*caffeoyl shikimate esterase*) are highly expressed at the young and intermediate stages but are sharply down-regulated at the mature stage, indicating that G, 5H, and S lignins are dominantly synthesized before fruit maturation (Figure 5d; Figures S9 and S10). In contrast, the expressions of *CAD* (*cinnamyl alcohol dehydrogenase*) and *CCR* (*cinnamoyl-CoA reductase*) are low in young epicarps but increase in the intermediate and mature stages, indicating that H- and C-type lignins are mainly produced during maturation (Figure 5d; Figures S9 and S10). Interestingly, G, 5H, and S lignins are methylated, while H and C lignins are not. The unmethylated monolignols are incorporated into lignin polymers via benzodioxane bonds, forming a linear structure without side chains (Hao & Mohnen, 2014; Xie et al., 2018). Most importantly, linear lignins enhance the hydrophobicity and stability of plant tissues because they have fewer crosslinking bonds with other cell wall components (Stone et al., 2018). Hence, *H. littoralis* fruits may have reduced permeability due to the prevalence of unmethylated lignins, thus increasing seed viability after a long seawater immersion.

The *WRKY12* and *E2Fc* transcription factors are up-regulated during epicarp maturation (Figure 5c; Figure S11 and Table S22). These TFs inhibit phenylpropanoid and lignin biosynthesis by repressing genes involved in pathway internal steps, including *NST1*, *NST2*, *VND6*, *VND7*, *SND1/NST3*, *MYB46*, *MYB58*, and *MYB85* (Figure S12). This effect is reinforced by the contemporaneous shutting down of

the expression of *ASL19* and *ASL20*, transcription factors that up-regulate internal steps of the same pathway (Figure 5c; Figure S11). These concerted changes likely result in a sharp decrease in lignin biosynthesis during maturation (Figure 5c; Figure S11), reflected by the drastic reduction in expression of *PAL* (*phenylalanine ammonia lyase*), *C4H* (*cinnamate 4-hydroxylase*), and *4CL* (*4-coumarate-CoA ligase*) during epicarp maturation. These latter genes regulate the upstream steps of phenylpropanoid and lignin biosynthesis (Figures S9 and S11), thus reducing the production of the key internal substrate p-Coumaroyl-CoA. Therefore, the downstream unmethylated and methylated lignin biosynthesis pathways would compete for the substrate. This is in consistent with our observation that *H. littoralis* primarily synthesizes unmethylated lignins, as indicated by the expression increase of *CCR* and *CAD* and decrease of *CSE*, *CCoAOM*, *F5H*, and *COMT*. This further demonstrates that *H. littoralis* have evolved to accumulate unmethylated lignins in epicarps, thereby increasing fruit impermeability.

Many key lignin biosynthetic pathway genes are either not expressed or at very low levels in seeds (Figures S9 and S10). Examples include *HCT* (*hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase*), *CSE*, *CCoAOMT*, *COMT*, and *F5H*, indicating that almost no lignin biosynthesis occurs in seeds. In contrast, genes involved in condensed tannin biosynthesis are overexpressed in seeds compared to epicarps (Figures S9 and S13), implying that condensed tannins mainly accumulate in seeds. However, fruit epicarps do increase condensed tannin biosynthesis at the mature stage, as indicated by upregulation of *TT7* (*cytochrome P450 superfamily protein*) and *ANR* (*anthocyanidin reductase*) (Figure 5d; Figures S9 and S13). Tannins can enhance seed resistance to salt, pathogen, and insect herbivores.

High level of cutin and wax biosynthesis in fruit epicarps

Plant cuticles mainly consist of two layers: the lipophilic cutin polymer that cannot be dissolved by organic solvents, and the lipophilic complexes, cuticular waxes, that can be dissolved by such solvents. By synthesizing the findings of previous studies (Fich et al., 2016; Kong et al., 2020; Lewandowska et al., 2020; Yeats & Rose, 2013), we built a working model of cutin and wax biosynthesis (Figure 6, Supplemental Note in the Supplemental Information File S1).

We quantified the gene expression profiles of all genes known to be involved in cutin and wax biosynthesis. Many of the key genes in this pathway are not expressed or are at very low levels in seeds, including *CYP86A* (*CYP86 subfamily of cytochrome P450s*), *GPAT* (*Glycerol-3-phosphate acyltransferase*), *HTH* (*HOTHEAD*), *KCS* (*β-ketoacyl-CoA synthase*), *CER2/26* (*eceriferum 2/26*), *CUS* (*cutin synthase*), and *SHN* (*shine*) (Figure 6; Figure S14). Hence, cutin and wax biosynthesis appear to be down-regulated in seeds.

In contrast, most genes in this pathway are highly expressed in young and intermediate epicarps but down-regulated to very low levels at the mature stage (Figure 6; Figure S14). This indicates that cutin and wax were accumulated before maturation. Interestingly, *MAH1* (midchain alkane hydroxylase), an enzyme that catalyzes the biosynthesis of secondary alcohols and ketones from alkanes, is expressed at a very low level at all three development stages (Figure 6; Figure S14). This implies that secondary alcohols and ketones are not necessary for cuticle wax biosynthesis in *H. littoralis*. Similarly, the genes from the CYP77A subfamily (the homologs of CYP77A6), encoding midchain hydroxylases, are also expressed at extremely low levels at all three developmental stages. Therefore, monoacylglycerols, the keystone precursors of cutin polymer biosynthesis, are not synthesized via the two CYP77A-requiring pathways in *H. littoralis*. The *CYP86A* and *HTH* pathways are likely used instead.

DISCUSSION

WGD facilitated evolution of sea-surfing *Heritiera* fruit

Aerial roots, viviparity, and salt tolerance are the phenotypes that predominantly facilitated plant adaptation to coastal intertidal zones (Duke, 2014; Tomlinson, 2016). Genetic changes underlying intertidal adaptation of mangrove species were enumerated in several previous studies (Guo et al., 2022; He et al., 2022; Ma et al., 2021; Pootakham et al., 2022; Xu et al., 2017). Our phylogenomic dating indicates that the two *Heritiera* species originated in the late Miocene (Figure 2b) when sea levels rose and fell frequently (Miller et al., 2005). Their ancestors have probably been inundated by rising sea levels and gradually become mangroves (Guo et al., 2022; Xu et al., 2017). Although the opportunities for terrestrial plants to invade intertidal habitats have been continually available through the Cenozoic period (He et al., 2022), *Heritiera* species invaded the intertidal zones relatively late, much later than these species-rich mangrove taxa *Rhizophoreae*, *Sonneratia* and *Avicennia* (about 40–50 million years ago).

Like in many other mangrove taxa (but not all, Xie et al., 2023), the two *Heritiera* species invaded intertidal zones after a whole-genome duplication event (Guo et al., 2022; Hu et al., 2020; Ma et al., 2021; Pootakham et al., 2022; Xu et al., 2017). We identified a WGD with an important contribution to the expansion of homologous gene groups in both *H. littoralis* and *H. fomes* (Figure 4b). These gene gains have facilitated the evolution of root cap development in the *Heritiera* species (Figure 4d). Both *Heritiera* species have developed huge plank roots and *H. fomes* have evolved conical pneumatophores. Plank roots are essential for the trees to fix on the thin soil of intertidal

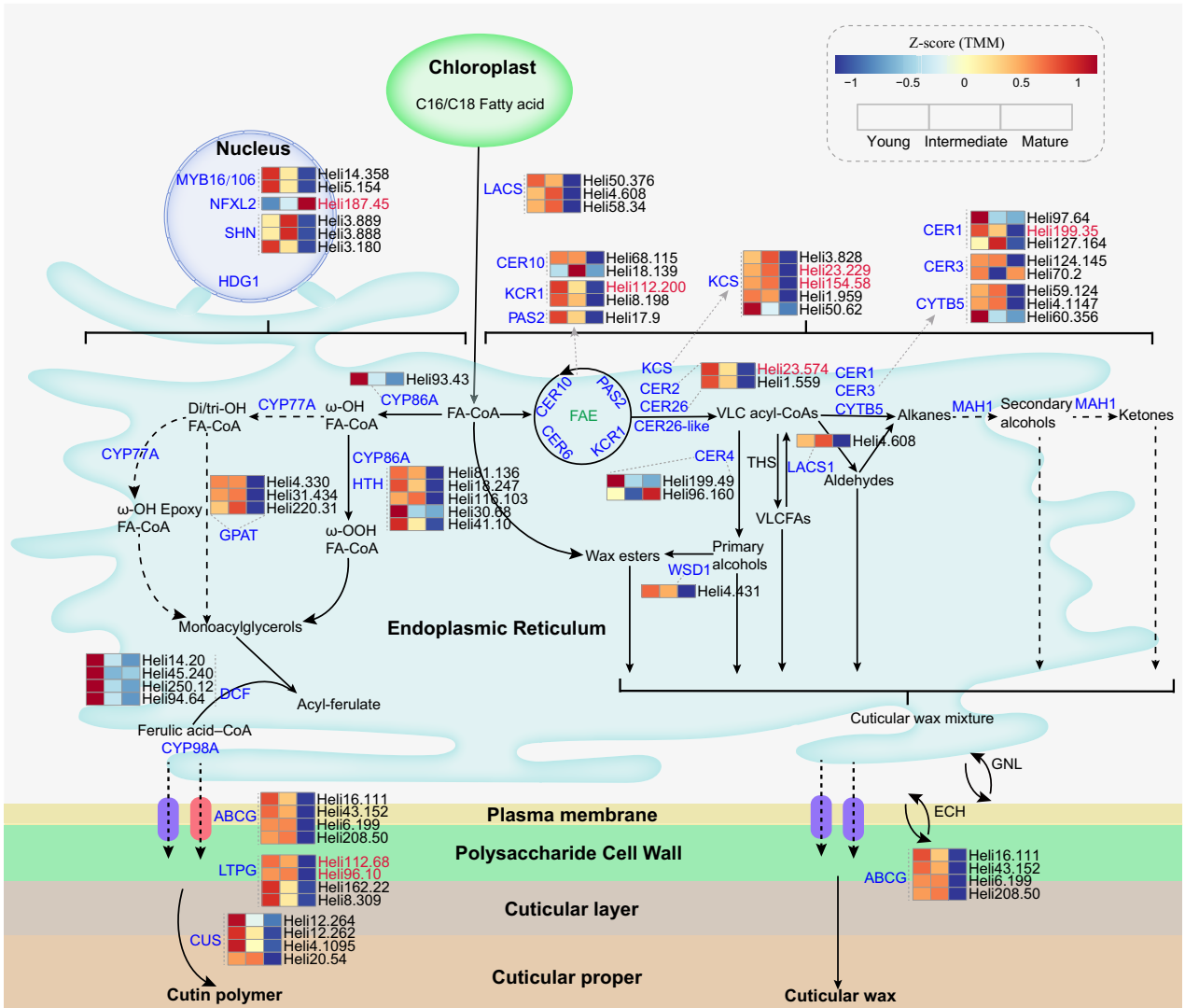

Figure 6. Expression profiles of genes in the cutin and wax biosynthetic pathways in *Heritiera littoralis* fruit epicarps.
Blue names indicate proteins in the pathways while black and red names indicate corresponding homologous *H. littoralis* genes. Gene names in red indicate the genes under positive selection. The heatmaps show the expression of genes involved in the pathway at young, intermediate, and mature stages of *H. littoralis* fruit development. The working model was built by synthesizing previous findings (Fich et al., 2016; Kong et al., 2020; Lewandowska et al., 2020; Yeats & Rose, 2013). *MYB16/106, MYB transcription factor; NFXL2, nfx1-like 2; SHN, shine; HDG1, homeodomain glabrous 1; LACS, long-chain acyl-coenzyme A synthase; CER, eceriferum (CER1, CER2, CER3, CER4, CER6, CER10, CER26, CER26-like); KCR1, β-ketoacyl-CoA reductase; PAS2, 3-hydroxyacyl-CoA dehydratase; KCS, β-ketoacyl-CoA synthase; CYTB5, cytochrome b; CYP86A, CYP86 subfamily of cytochrome P450s; CYP77A, CYP77 subfamily-encoded midchain hydroxylase; HTH, HOTHEAD; GPAT, Glycerol-3-phosphate acyltransferase; DCF, deficient in cutin ferulate; ABCG, ATP binding cassette G; LTPG, lipid transfer protein G; CUS, cutin synthase; MAH1, midchain alkane hydroxylase; WSD1, wax synthase/acyl-CoA: diacylglycerol acyltransferase; ECH, echidna; GNL, gnom-like.*

zones and thus resist strong winds, while conical pneumatophores help the plants breathe during tidal inundations (Tomlinson, 2016).

The current coastal forests are threatened by global changes, such as the rapid rise of sea levels, higher frequency and intensity of storms, and changes in regional precipitation and temperature regimes (Chen et al., 2011; Parmesan & Yohe, 2003; Poloczanska et al., 2013). Coastal plants may survive such changes by shifting their geographical range to track the conditions they have adapted to (Berg et al., 2010; Moritz & Agudo, 2013; Valladares et al., 2014). Geographical distribution and population structure of coastal plant species confined within the narrow intertidal zones along tropical and subtropical coastlines have been profoundly impacted by historic sea-level changes (Banerjee et al., 2022; Guo et al., 2016; Wang et al., 2022; Zhang et al., 2022). Long-distance dispersal is a key factor shaping the current distribution and population genetic patterns (Binks et al., 2019; Guo et al., 2018), and it will also be important for response to sea-level rise.

Our results suggest that gene families involved in lignin and flavonoid metabolic pathways have disproportionally expanded in *H. littoralis*. Furthermore, frequently only one of the duplicated copies of genes involved in lignin, tannin, and cutin and wax biosynthesis evolved under positive selection. This suggests that some neofunctionalization has occurred, with WGD duplicates perhaps functioning collaboratively in coping with different environmental conditions. Acquisition of derived differential expression has been previously found in other plant species (Xu et al., 2023). The differential natural selection pressure on duplicated genes found in the two *Heritiera* species is likely another case of adaptive evolution attributed to WGD.

Evolution of gene expression underlying sea-surfing fruits

By measuring gene expression genome-wide using RNA-seq, we inferred that methylated and unmethylated lignins are synthesized in young and intermediate stage *H. littoralis* fruit epicarps. The biosynthesis of both types of lignins may be necessary for fruit growth at the early development stages. Moreover, despite the reduction of internal substrate supply at the mature stage, *H. littoralis* fruit preferentially produces unmethylated lignins. Results from *Paphiopedilum armeniacum* demonstrate that down-regulation of the first lignin methylation gene *CCoAOMT* is crucial for unmethylated lignin production (Fang et al., 2020). Almost all homologous copies of *CCoAOMT* are down-regulated upon fruit maturation in *H. littoralis* (Figure 5d). This compensation mechanism likely plays an important role in increasing the hydrophobicity and stability of mature fruit in seawater (Stone et al., 2018). The heavily lignified secondary cell walls developed by *P. armeniacum* seeds also enhance seed survival in harsh conditions (Fang et al., 2020). However, the large amounts of unmethylated lignin in *P. armeniacum* seeds inhibit seed germination, making germplasm conservation and commercial production difficult (Fang et al., 2020). In contrast, unmethylated lignin accumulating in *H. littoralis* fruit epicarps is beneficial because impermeability is essential for seed viability maintenance during long-distance dispersal. In addition, *H. littoralis* seeds increase tannin biosynthesis upon maturation. This is likely also beneficial because tannins play a role in photoprotection, pathogen, and herbivory resistance, and salt tolerance (Kandil et al., 2004; Wang et al., 2014).

The cuticle covering the plant aerial surface is the first line of defense against environmental threats by controlling the exchange of water, solutes, and gases between plants and the surrounding environment (Lewandowska et al., 2020). Cutin polymers and cuticular waxes are also accumulated at the young and intermediate stages of *H. littoralis* fruit development and are likely important for the protection of the fruit from salty seawater and strong UV radiation. Gene expression analyses also suggest that secondary alcohols and ketones are not necessary for cuticle wax biosynthesis in *H. littoralis*. VLC ketones are likewise not necessary for seed cuticle synthesis in *Amphicarpaea edgeworthii* because *MAH1* is expressed at low levels in its aerial and subterranean seeds (Liu et al., 2021).

Genomic resources for coastal forest conservation and restoration

Our identification of the genes used by *H. littoralis* for lignin, tannin, and cutin and wax biosynthesis can be useful in molecular breeding to obtain varieties for coastal forest restoration. Higher expression of *WRKY12* and *E2Fc* inhibits p-Coumaroyl-CoA biosynthesis in mature *H. littoralis* fruit epicarps, leading to competition between methylated and unmethylated lignin biosynthesis pathways for the precursor pool. Meanwhile, the expression regulation of genes *CCoAOMT*, *F5H*, *COMT*, *CSE*, *CAD*, and *CCR* is the key for balancing the biosynthesis of methylated and unmethylated lignins. Using advanced gene editing tools, it may be possible to design a new, better dispersing *H. littoralis* variety. These findings are also likely applicable to other coastal plant species. In the past decades, coastal forests afforestation has been widely conducted over the world to increase forest area (Friess et al., 2019; Thomas et al., 2017). Additional coastal forests are surely needed to ameliorate the impact of rising sea levels in coastal regions. Furthermore, the genome sequences provided here would also be valuable for future evolutionary and genomic studies, such as large-scale comparative genomic studies on more species of the Malvaceae family.

EXPERIMENTAL PROCEDURES

Plant materials, DNA/RNA extraction

The *H. fomes* genome was newly sequenced in this study. The *H. littoralis* genome was sequenced as part of a mangrove genomics research project conducted by our lab (He et al., 2022), but genome description is newly provided in this paper. *H. littoralis* leaves were collected from an adult individual in Dongzhai Harbor of Hainan Island and those from *H. fomes* were sampled from an adult individual in Bangladesh (Table S23). Fresh *H. fomes* and *H. littoralis* leaves were harvested and immediately frozen in liquid nitrogen, followed by preservation at −80°C for subsequent nucleic acid extractions. High-quality genomic DNA was extracted from leaves using the Cetyl Trimethyl Ammonium Bromide (CATB) method (Doyle & Doyle, 1987). We also sampled *H. littoralis* fresh leaves, fruits, and flowers, as well as *H. fomes* leaves for RNA sequencing. To obtain data for phylogeny construction, we collected healthy *H. angustata*, *H. parvifolia*, and *Sterculia monosperma* leaves from the South China Botanical Garden for RNA sequencing (RNA-seq) (Table S23).

Young, intermediate, and mature *H. littoralis* fruits were sampled from nine randomly selected individuals from Qi'ao Island, Guangdong, China (Table S23). We dissected the fruit and kept epicarps for RNA extraction. Fresh epicarp tissues were first stored in liquid nitrogen, and then conserved at −80°C. RNA was

extracted following a modified CTAB method (Yang et al., 2008). We used fruit, leaf, flower, and root morphological characters for species diagnosis of the four *Heritiera* species (Figure S2 and Table S24).

Sequencing, genome assembling, and annotation

Here we briefly describe the methods used for genome sequencing, assembling, and annotation. Extended descriptions can be found in the Supplemental Note of the Supporting Information File S1.

Using an Illumina X-TEN platform (San Diego, CA, USA), we obtained 43Gb (~51×) and 48Gb (~55×) short reads for *H. fomes* and *H. littoralis*, respectively (Table S3). We used Jellyfish (Marçais & Kingsford, 2011) to estimate the genome sizes of *H. fomes* and *H. littoralis* based on these Illumina reads. SMRT sequencing of *H. fomes* was conducted on a PacBio Sequel II sequencing platform using HiFi Bundle v.2 sequencing reagents and SMRT Cell (8 M). 8 M SMRT Cells and V3.0 sequencing reagent were used for *H. littoralis* to perform sequencing on a PacBio Sequel II platform (PacBio, CA, USA). We obtained 440Gb (~475×) and 160Gb (~166×) of long reads for *H. fomes* and *H. littoralis*, respectively (Table S1).

H. fomes circular consensus (CCS) reads were assembled using HiFiasm v.0.14 (Cheng et al., 2021) while *H. littoralis* continuous long reads (CLRs) were assembled using FALCON (Chin et al., 2016). Redundant contigs were removed using Purge Haplotigs v1.1.2 (Roach et al., 2018) with parameters set as: -l 30 -m 120 -h 197. To assess the quality of the assembled genomes, raw Illumina paired-end reads (Table S3), RNA-seq reads (Table S3), SMRT long reads (Table S1), and conserved core genes from the Benchmarking Universal Single-Copy Ortholog (embryophyte_odb10 of BUSCO) database were mapped to the assemblies.

To facilitate the prediction of protein-coding genes, RNA extracted from *H. littoralis* flowers, fruits, and leaves was sequenced on a BGISEQ-500 platform with 100 bp reads by employing OligoDT, while RNA extracted from *H. fomes* leaves was sequenced on an Illumina Novaseq platform (Table S3). RepeatModeler v.2.0.2a and RepeatMasker v.4.1.2-p1 (http://repeatmasker.org/) were used to perform *de novo* identification and masking of repeat sequences in the *H. fomes* and *H. littoralis* genomes. We also predicted repeat sequences *de novo* and classified them following a previously described pipeline (Xie et al., 2023). *H. fomes, H. littoralis*, *T. cacao*, *D. zibethinus*, *D. turbinatus*, and *A. thaliana* genome sequences were used in this prediction.

Identification of protein-coding regions and gene prediction were performed using a combination of ab initio, homology-based, and transcriptome-based prediction methods (see detailed method descriptions in the Supplementary Note of Supporting Information S1). Functional annotations of the obtained gene sets were conducted using BLASTP against the NCBI-NR (https://www.ncbi.nlm.nih.gov/), SwissProt (https://www.uniprot.org/), and Trembl in UniProt (https://www.uniprot.org/) (Table S4). Protein domains were annotated by mapping to the InterPro v87 (https://www.ebi.ac.uk/interpro/) and EggNOG v5 databases (http://eggnog5.embl.de/) using InterProScan and EggNOG-mapper v2 (Cantalapiedra et al., 2021). KEGG annotations were assigned to genes by mapping them to the KEGG database (https://www.genome.jp/kegg/) with EggNOG-mapper v2 and Kofam_scan v1.3.0 (Aramaki et al., 2020). Gene Ontology (GO) terms for genes were correspondingly extracted from InterProscan or EggNOG-mapper v2 (Table S4).

Phylogenomic analyses

We sequenced the transcriptomes of three species closely related to *H. fomes* and *H. littoralis*: *H. angustata*, *H. parvifolia*, and *S. monosperma* (Table S7). RNA extracted from leaves was used to construct cDNA libraries following the TruSeq standard mRNA Library Preparation Guide. The 350 bp insert libraries were sequenced using the Illumina Novaseq platform. Quality control of the output raw reads was performed using Trimmomatic v.0.35 (Bolger et al., 2014). Clean *H. angustata*, *H. parvifolia*, and *S. monosperma* reads were assembled into transcriptomes using TRINITY (Grabherr et al., 2011). Redundant contigs were removed using CD-HIT (Fu et al., 2012). BWA v.0.7.17 (Li & Durbin, 2009) was then used to map clean reads back to the transcriptome assemblies. SAMtools (Li et al., 2009) were used to obtain and sort bam files. Contigs with an average depth of <2 and length < 200 bp were further removed and isoforms transcribed from the same gene were unified by keeping the longest one as a unique transcript (Table S6). The TRINOTATE pipeline (https://github.com/Trinotate) was used to identify gene models (Table S25).

We used OrthoFinder v. 2.3.1 to identify 20 302 orthologous groups in *H. fomes, H. littoralis*, and seven related species (*H. angustata*, *H. parvifolia*, *S. monosperma*, *D. Zibethinus*, *T. cacao*, *D. turbinatus*, and *A. thaliana*, Table S7) (Emms & Kelly, 2019). We found 658 single-copy orthologous groups with a single representative from all nine species. Protein sequences of these single-copy orthologs were aligned using MUSCLE v. 3.8.31 (Edgar, 2004) and transformed into DNA codons using PAL2NAL v. 14.0 (Suyama et al., 2006). Alignment gaps were deleted under the "complete deletion" model using trimAl v. 1.4 (Capella-Gutiérrez et al., 2009), setting the "-nogaps" parameter. All genes with alignments longer than 250 bps were concatenated to reconstruct the phylogeny using RAxML v. 8.2.9 (parameters: -fa -m GTRGAMMA -p 12345 -N 1000 -k -x 12 345) (Stamatakis, 2014).

We further dated the constructed phylogeny using MCMCTREE implemented in the PAML v.4.9 package (Yang, 2007). Three reliable fossil calibrations were incorporated (Table S8). First, the root node of Malvales and Brassicales was constrained between 89.8 and 93.9 Mya (Gandolfo et al., 1998). Second, the *Bombacoxylon langstoni* (Dipterocarpaceae) wood fossil from the late Campanian was used to calibrate the divergence of Malvaceae species from *D. turbinatus* (Dipterocarpaceae) as no later than 72.1 Mya (Wheeler & Lehman, 2000). Third, *Durio*-type pollen recorded from the Paleocene and Eocene in India was used to calibrate the root age of the common ancestor of *Heritiera* genus, *S. monosperma*, and *D. Zibethinus* between 33.9 and 66.0 Mya (Ashton & Gunatilleke, 1987).

Based on the 658 single-copy gene groups, we also calculated Ks values for orthologs between *H. fomes* and each of *H. littoralis*, *H. angustata*, *H. parvifolia*, *S. monosperma*, *D. zibethinus*, and *T. cacao*. The distributions of these Ks values were used as supporting evidence to infer the divergence order of these species. Expansion and contraction of the orthologous gene groups were inferred using CAFE v.4.2.1 (Han et al., 2013). The program uses a birth and death process to model gene gain and loss on a phylogeny.

Analysis of genome synteny and whole-genome duplication

Collinear genome blocks were defined as pairs of genome regions harboring at least five pairs of collinear genes. Each protein-coding gene in *H. fomes*, *H. littoralis*, *D. zibethinus*, and *T. cacao* was compared to all other genes in the same genome using

BLASTP with an e-value cut-off of 1.0×10^{-10}. McscanX (Wang et al., 2012) was then used to identify intraspecific collinear genome blocks and intraspecific collinear gene pairs. We used the same method to identify interspecific collinear genome blocks and interspecific collinear gene pairs for *H. fomes* and *H. littoralis*, *H. fomes*, and *D. zibethinus*, as well as *H. fomes* and *T. cacao*.

As transcriptomes have no scaffolding information, we are unable to identify collinear blocks using transcriptome data. However, *H. angustata*, *H. parvifolia*, and *S. monosperma* transcripts were used to identify intraspecific homologs (paralogs). Based on the homologous gene groups constructed before, we extracted homologs from each of the three species. N genes in a species yield N*(N-1)/2 pairs of homolog candidates. To reduce computational complexity, we used only homologous groups with two to four paralogous genes, e.g., we required the group to contain between two and four *H. angustata* genes to test for paralog presence.

Synonymous substitution rates (Ks) between intraspecific collinear genes, interspecific collinear genes, and intraspecific paralogs were computed using KaKs_Calculator v. 2.0 under the YN model (Wang et al., 2010). To explore synteny between the *H. fomes* and *T. cacao* genomes, as well as between *H. fomes* and *D. zibethinus*, we used MCscan (https://github.com/tanghaibao/jcvi/wiki) to obtain pairwise synteny information and visualized it using a dot plot. Based on the Ks distributions and the phylogeny constructed above, we estimated the age of the WGD event following the method described previously (Guo et al., 2022).

Analyses of gene duplication and positive selection

We used DupGen_finder with default parameters (Qiao et al., 2019) to identify gene duplication events in *H. fomes*, *H. littoralis*, *D. zibethinus*, *T. cacao*, *D. turbinatus*, and *A. thaliana*. These events underlie gene gains. We further used DupGen_finder-unique to classify the gene duplications into five classes according to the duplication mechanisms: whole-genome (WGD), tandem (TD), proximal (less than 10 genes away on the same chromosome, PD), transposed (TRD), and dispersed (DSD). We then calculated the Ka (number of amino-acid changing per nonsynonymous site), Ks (number of synonymous substitutions per synonymous site), and Ka/Ks ratios for each pair of duplicated genes based on the YN model in KaKs_Calculator v. 2.0 (Wang et al., 2010). We also searched for *H. littoralis* and *H. fomes* genes included in both expanded gene groups identified using CAFE and any of the five classes of duplicated genes. We conducted gene ontology enrichment analysis for these gene group intersections using the ClusterProfiler package v.4.0.5 (Wu et al., 2021).

We identified homologous gene groups that include genes annotated as encoding proteins involved in lignin (Table S15), tannin (Table S16), and wax biosynthesis (Table S17) (Dong & Lin, 2021; Fich et al., 2016; Kong et al., 2020; Lewandowska et al., 2020; Xie et al., 2018; Yeats & Rose, 2013). For gene groups containing two or more paralogous duplicated copies, we used Notung 2.9 to reconcile gene trees. We manually split the paralogs into subgroups based on the reconciled gene trees. With these gene groups and subgroups, we used the branch-site model implemented in PAML (Yang, 2007) to detect genes under positive selection.

Comparative fruit transcriptome analyses

We captured gene expression in *H. littoralis* fruit epicarps and seeds at young, intermediate, and mature developmental stages. For each stage, we sequenced three individuals for biological replication. The library construction and RNA sequencing were conducted using the same method as described above for *H. angustata*, *H. parvifolia*, and *S. monosperma* RNA sequencing.

We quantified gene expression levels and identified differentially expressed genes (DEGs). Clean reads from each sample were aligned to the *H. littoralis* genome using HISAT2 version 2.2.1 (Kim et al., 2019). Feature counts from the package Rsubread v.2.6.1 were used to obtain gene expression values as raw read counts across all samples (Liao et al., 2014). Gene expression values were firstly converted to transcripts per kilobase of exon model per million mapped reads (TPM) using Deseq 2 (Love et al., 2014), and TPM values were then normalized as trimmed mean of M-values (TMM) using edgeR (Robinson et al., 2009).

DEGs were identified for three comparisons: mature vs. intermediate, mature vs. young, and intermediate vs. young stage. For each comparison, the transcripts with adjusted *P*-values (false discovery rate-adjusted *P*-values) below 5% and a fold change greater than two were considered DEGs. Heatmaps of the expression profiles were produced using the pheatmap package. Mean expression values from three biological replicates were used as the expression values of a gene.

GO and KEGG analysis of the DEGs was performed using the ClusterProfiler package v.4.0.5 (Wu et al., 2021). Hypergeometric tests were performed to determine whether specific functional categories from GO and KEGG were significantly overrepresented in *H. fomes* and *H. littoralis* gene sets compared to the whole genome.

ACCESSION NUMBERS

The genome sequences and raw reads have been deposited in the National Center for Biotechnology Information (NCBI). Accession numbers of the genome sequences for *Heritiera fomes* and *H. littoralis* are JAPDPM000000000 and JAPDPK000000000, respectively. Accession numbers of the RNA-seq reads for *H. angustata*, *H. parvifolia*, and *Sterculia monosperma* are SRR24791438, SRR24791437, and SRR24791435, respectively. Accession numbers for the RNA-seq reads of *H. littoralis* fruits at different development stages are provided in Table S19 in the Supporting Information.

CONFLICT OF INTEREST

The authors have no competing interest to declare.

ACKNOWLEDGMENTS

We thank Prof. Chung-I Wu at Sun Yat-sen University for his helpful suggestions and Prof. Jianquan Liu at Sichuan University for providing *Heritiera fomes* plant material. We thank Ms. Lanjun Liu for her constructive discussions on plant root development. We thank Dr. Luzhen Chen at Xiamen University, Yechun Xu at Environmental Horticulture Research Institute of Guangdong Academy of Agriculture Sciences, and Wenbing Yu at Xishuangbanna Tropical Botanic Gargen, Chinese Academy of Sciences for providing the fruit photos of *Heritiera fomes*, *Heritiera angustata*, *Heritiera parvifolia, Durio zibethinus*, *Theobroma cacao*, and *Dipterocarpus turbinatus* in Figure 2. This study was supported by the Guangdong Basic and Applied Basic Research Foundation (grant nos. 2023A1515012406 and 2021A1515011160); National Natural Science Foundation of China (grant nos. 42276159, 31830005, and 32330005); Guangzhou Basic and Applied Basic Research

Foundation (2023A04J1949); the Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (no. 311021006); and the Chang Hungta Science Foundation of Sun Yat-Sen University.

AUTHOR CONTRIBUTIONS

ZG and S. Shi conceived the study. S. Shi, JW, ZG, and XW collected the samples and conducted the sequencing. JW, WX, ZG, FS, ZH, XW, and S. Shao analyzed the data. ZG and JW wrote the manuscript. All authors read and approved the manuscript.

SUPPORTING INFORMATION

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

Figure S1. Genome size estimation by jellyfish.

Figure S2. Morphological and distributional information of *Heritiera fomes*, *H. littoralis*, and two congeneric relatives.

Figure S3. The phylogeny of eight species of Malvales, with *A. thaliana* as outgroup.

Figure S4. The Ka/Ks values distributions of gene pairs derived from different modes of duplication.

Figure S5. Duplicated genes are under differential natural selection.

Figure S6. Gene expression correlations among all samples.

Figure S7. Volcano plots of gene expression profiles in the epicarps and seeds of *H. littoralis*.

Figure S8. KEGG enrichment analysis with the differentially expressed genes.

Figure S9. Detailed biosynthetic pathways of lignin and condensed tannins in *H. littoralis*.

Figure S10. Expression levels of the putative candidate genes involved in the lignin biosynthetic pathway.

Figure S11. Expression levels of the putative candidate genes involved in the transcription factors regulating lignin biosynthesis in *H. littoralis*.

Figure S12. A scheme of transcriptional regulation of lignin biosynthetic genes in plants.

Figure S13. Expression levels of the putative candidate genes involved in the condensed tannins biosynthetic pathway.

Figure S14. Expression levels of the putative candidate genes involved in the cutin and wax biosynthetic pathway.

Table S1. Statistics of data obtained from Single Molecule Real Time (SMRT) sequencing.

Table S2. Summary of BUSCO genome completeness assessment.

Table S3. Statistics of the sequences obtained from second-generation sequencing.

Table S4. Functional annotations of the predicted genes of *H. fomes* and *H. littoralis*.

Table S5. Repeat sequence content in the genomes of the six species with genome sequence.

Table S6. Summary of *H. parvifolia*, *H. angustata* and *Sterculia monosperma* transcriptome assemblies.

Table S7. Data resources of the genomes and transcriptomes used in this study.

Table S8. Fossil calibrations were used in this study.

Table S9. Collinear genes of the six species with genome sequence.

Table S10. Functional enrichment of genes in the expanded gene families in the most recent common ancestor of *H. fomes* and *H. littoralis*.

Table S11. Genes duplicated from different mechanisms of the six species with the genome sequence.

Table S12. Genes duplicated from different mechanisms in the expanded gene groups.

Table S13. Functional enrichment of duplicated genes in the expanded gene families in *H. littoralis*.

Table S14. Functional enrichment of duplicated genes in the expanded gene families in *H. fomes*.

Table S15. Genes encoding key enzymes of the lignin biosynthesis pathway.

Table S16. Genes encoding key enzymes of the condensed tannins biosynthesis pathway.

Table S17. Genes encoding key enzymes of the cutin and wax biosynthetic pathways.

Table S18. Detailed annotation of the genes under positive selection.

Table S19. Overview of *H. littoralis* fruit RNA-seq read mapping.

Table S20. Differentially expressed genes in epicarp and seed.

Table S21. KEGG pathways significantly enriched in the differentially expressed genes.

Table S22. Transcription factors regulating lignin biosynthesis.

Table S23. Sampling locations and characteristics of the four *Heritiera* species and *Sterculia monosperma*.

Table S24. Morphological characters of the four *Heritiera* species.

Table S25. Functional annotation of the *de novo* assembled transcriptomes.

Supplemental Note S1. Extended method description for sequencing, genome assembly, and annotation; 2. Description of the working model of cutin and wax biosynthesis.

REFERENCES

Aramaki, T., **Blanc-Mathieu, R.**, **Endo, H.**, **Ohkubo, K.**, **Kanehisa, M.**, **Goto, S.** ***et al.*** (2020) KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*, **36**, 2251–2252.

Ashton, P.S. & **Gunatilleke, C.V.S.** (1987) New light on the plant geography of Ceylon. I. Historical Plant Geography. *Journal of Biogeography*, **14**, 249–285.

Banerjee, A.K., **Feng, H.**, **Guo, W.**, **Harms, N.E.**, **Xie, H.**, **Liang, X.** ***et al.*** (2022) Glacial vicariance and oceanic circulation shape population structure of the coastal legume Derris trifoliata in the Indo-West Pacific. *American Journal of Botany*, **109**, 1016–1034.

Berg, M.P., **Toby Kiers, E.**, **Driessen, G.**, **Van Der Heijden, M.**, **Kooi, B.W.**, **Kuenen, F.** ***et al.*** (2010) Adapt or disperse: understanding species persistence in a changing world. *Global Change Biology*, **16**, 587–598.

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

Bolger, A.M., **Lohse, M.** & **Usadel, B.** (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Cantalapiedra, C.P., **Hernández-Plaza, A.**, **Letunic, I.**, **Bork, P.** & **Huerta-Cepas, J.** (2021) eggNOG-mapper v2: functional annotation, Orthology assignments, and domain prediction at the metagenomic scale K. Tamura, ed. *Molecular Biology and Evolution*, **38**, 5825–5829.

Capella-Gutiérrez, S., **Silla-Martínez, J.M.** & **Gabaldón, T.** (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.

Chen, I.C., **Hill, J.K.**, **Ohlemüller, R.**, **Roy, D.B.** & **Thomas, C.D.** (2011) Rapid range shifts of species associated with high levels of climate warming. *Science*, **333**, 1024–1026.

Cheng, H., **Concepcion, G.T.**, **Feng, X.**, **Zhang, H.** & **Li, H.** (2021) Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, **18**, 170–175. Available from: https://doi.org/10.1038/s41592-020-01056-5

Chin, C.S., **Peluso, P.**, **Sedlazeck, F.J.**, **Nattestad, M.**, **Concepcion, G.T.**, **Clum, A.** ***et al.*** (2016) Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods*, **13**, 1050–1054. Available from: https://pubmed.ncbi.nlm.nih.gov/27749838/

Das, A.B., **Mukherjee, A.K.** & **Das, P.** (2001) Molecular phylogeny of Heritiera Aiton (Sterculiaceae), a tree mangrove: variations in RAPD markers and nuclear DNA content. *Botanical Journal of the Linnean Society*, **136**, 221–229.

Dong, N.Q. & **Lin, H.X.** (2021) Contribution of phenylpropanoid metabolism to plant development and plant–environment interactions. *Journal of Integrative Plant Biology*, **63**, 180–209.

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

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

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

Fang, L., **Xu, X.**, **Li, J.**, **Zheng, F.**, **Li, M.**, **Yan, J.** ***et al.*** (2020) Transcriptome analysis provides insights into the non-methylated lignin synthesis in Paphiopedilum armeniacum seed. *BMC Genomics*, **21**, 524.

Fich, E.A., **Segerson, N.A.** & **Rose, J.K.C.** (2016) The plant polyester Cutin: biosynthesis, structure, and biological roles. *Annual Review of Plant Biology*, **67**, 207–233.

Friess, D.A., **Rogers, K.**, **Lovelock, C.E.**, **Krauss, K.W.**, **Hamilton, S.E.**, **Lee, S.Y.** ***et al.*** (2019) The state of the World's mangrove forests: past, present, and future. *Annual Review of Environment and Resources*, **44**, 89–115.

Fu, L., **Niu, B.**, **Zhu, Z.**, **Wu, S.** & **Li, W.** (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, **28**, 3150–3152.

Gandolfo, M.A., **Nixon, K.C.** & **Crepet, W.L.** (1998) A new fossil flower from the turonian of New Jersey: Dressiantha bicarpellata gen. et sp. nov. (Capparales). *American Journal of Botany*, **85**, 964–974.

Grabherr, M.G., **Haas, B.J.**, **Yassour, M.**, **Levin, J.Z.**, **Thompson, D.A.**, **Amit, I.** ***et al.*** (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology*, **29**, 644–652.

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

Guo, Z., **Xu, S.**, **Xie, W.**, **Shao, S.**, **Feng, X.**, **He, Z.** ***et al.*** (2022) Adaptation to a new environment with pre-adaptive genomic features – evidence from woody plants colonizing the land–sea interface. *The Plant Journal*, **111**, 1411–1424.

Han, M.V., **Thomas, G.W.C.**, **Lugo-Martinez, J.** & **Hahn, M.W.** (2013) Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Molecular Biology and Evolution*, **30**, 1987–1997.

Hao, Z. & **Mohnen, D.** (2014) A review of xylan and lignin biosynthesis: foundation for studying Arabidopsis irregular xylem mutants with pleiotropic phenotypes. *Critical Reviews in Biochemistry and Molecular Biology*, **49**, 212–241.

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

Hu, M.J., **Sun, W.H.**, **Tsai, W.C.**, **Xiang, S.**, **Lai, X.K.**, **Chen, D.Q.** ***et al.*** (2020) Chromosome-scale assembly of the Kandelia obovata genome. *Horticulture Research*, **7**, 75. Available from: https://doi.org/10.1038/s41438-020-0300-x

Kandil, F.E., **Grace, M.H.**, **Seigler, D.S.** & **Cheeseman, J.M.** (2004) Polyphenolics in Rhizophora mangle L. leaves and their changes during leaf development and senescence. *Trees—Structure and Function*, **18**, 518–528.

Kim, D., **Paggi, J.M.**, **Park, C.**, **Bennett, C.** & **Salzberg, S.L.** (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, **37**, 907–915. Available from: https://pubmed.ncbi.nlm.nih.gov/31375807/

Kong, L., **Liu, Y.**, **Zhi, P.**, **Wang, X.**, **Xu, B.**, **Gong, Z.** ***et al.*** (2020) Origins and evolution of cuticle biosynthetic machinery in land plants. *Plant Physiology*, **184**, 1998–2010.

Lewandowska, M., **Keyl, A.** & **Feussner, I.** (2020) Wax biosynthesis in response to danger: its regulation upon abiotic and biotic stress. *The New Phytologist*, **227**, 698–713.

Li, H. & **Durbin, R.** (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Li, H., **Handsaker, B.**, **Wysoker, A.**, **Fennell, T.**, **Ruan, J.**, **Homer, N.** ***et al.*** (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Liao, Y., **Smyth, G.K.** & **Shi, W.** (2014) FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.

Liu, Y., **Zhang, X.**, **Han, K.**, **Li, R.**, **Xu, G.**, **Han, Y.** ***et al.*** (2021) Insights into amphicarpy from the compact genome of the legume Amphicarpaea edgeworthii. *Plant Biotechnology Journal*, **19**, 952–965.

Love, M.I., **Huber, W.** & **Anders, S.** (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**, 550.

Ma, D., **Guo, Z.**, **Ding, Q.**, **Zhao, Z.**, **Shen, Z.**, **Wei, M.** ***et al.*** (2021) Chromosome-level assembly of the mangrove plant Aegiceras corniculatum genome generated through Illumina, PacBio and Hi-C sequencing technologies. *Molecular Ecology Resources*, **21**, 1593–1607.

Marçais, G. & **Kingsford, C.** (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, **27**, 764–770.

Miller, K.G., **Kominz, M.A.**, **Browning, J.V.**, **Wright, J.D.**, **Mountain, G.S.**, **Katz, M.E.** ***et al.*** (2005) The phanerozoic record of global sea-level change. *Science (80-.)*, **310**, 1293–1298. Available from: http://www.sciencemag.org/content/310/5752/1293.abstract%5Cnhttp://www.sciencemag.org/content/310/5752/1293.full.pdf%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/16311326

Moritz, C. & **Agudo, R.** (2013) The future of species under climate change: resilience or decline? *Science*, **341**, 504–508.

Murat, F., **Armero, A.**, **Pont, C.**, **Klopp, C.** & **Salse, J.** (2017) Reconstructing the genome of the most recent common ancestor of flowering plants. *Nature Genetics*, **49**, 490–496.

Parmesan, C. & **Yohe, G.** (2003) A globally coherent fingerprint of climate change. *Nature*, **421**, 37–42.

Pierce, S., **Spada, A.**, **Caporali, E.**, **Ceriani, R.M.** & **Buffa, G.** (2019) Enzymatic scarification of Anacamptis morio (Orchidaceae) seed facilitates lignin degradation, water uptake and germination. *Plant Biology*, **21**, 409–414.

Poloczanska, E.S., **Brown, C.J.**, **Sydeman, W.J.**, **Kiessling, W.**, **Schoeman, D.S.**, **Moore, P.J.** ***et al.*** (2013) Global imprint of climate change on marine life. *Nature Climate Change*, **3**, 919–925.

Pootakham, W., **Sonthirod, C.**, **Naktang, C.**, **Kongkachana, W.**, **Sangsrakru, D.**, **U-thoomporn, S.** ***et al.*** (2022) A chromosome-scale reference genome assembly of yellow mangrove (*Bruguiera parviflora*) reveals a whole genome duplication event associated with the Rhizophoraceae lineage. *Molecular Ecology Resources*, **22**, 1939–1953.

POWO. (2022) Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Publ. Internet. http://www.plantsoftheworldonline.org/

Qiao, X., **Li, Q.**, **Yin, H.**, **Qi, K.**, **Li, L.**, **Wang, R.** ***et al.*** (2019) Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biology*, **20**, 38.

Roach, M.J., **Schmidt, S.A.** & **Borneman, A.R.** (2018) Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, **19**, 1–10.

Robinson, M.D., **McCarthy, D.J.** & **Smyth, G.K.** (2009) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

Sale, P.F., **Agardy, T.**, **Ainsworth, C.H.**, **Feist, B.E.**, **Bell, J.D.**, **Christie, P.** ***et al.*** (2014) Transforming management of tropical coastal seas to cope

with challenges of the 21st century. *Marine Pollution Bulletin*, **85**, 8–23. Available from: https://doi.org/10.1016/j.marpolbul.2014.06.005

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

Stone, M.L., **Anderson, E.M.**, **Meek, K.M.**, **Reed, M.**, **Katahira, R.**, **Chen, F.** ***et al.*** (2018) Reductive catalytic fractionation of C-Lignin. *ACS Sustainable Chemistry & Engineering*, **6**, 11211–11218.

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.

Thomas, N., **Lucas, R.**, **Bunting, P.**, **Hardy, A.**, **Rosenqvist, A.** & **Simard, M.** (2017) Distribution and drivers of global mangrove forest change, 1996-2010. *PLoS One*, **12**, e0179302.

Tomlinson, P.B. (2016) *The botany of mangroves*, Second edition. London: Cambridge University Press.

Valladares, F., **Matesanz, S.**, **Guilhaumon, F.**, **Araújo, M.B.**, **Balaguer, L.**, **Benito-Garzón, M.** ***et al.*** (2014) The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change. *Ecology Letters*, **17**, 1351–1364.

Van Der Stocken, T., **Carroll, D.**, **Menemenlis, D.**, **Simard, M.** & **Koedam, N.** (2019) Global-scale dispersal and connectivity in mangroves. *Proceedings of the National Academy of Sciences of the United States of America*, **116**, 915–922.

Van Der Stocken, T., **Wee, A.K.S.**, **De Ryck, D.J.R.**, **Vanschoenwinkel, B.**, **Friess, D.A.**, **Dahdouh-Guebas, F.** ***et al.*** (2019) A general framework for propagule dispersal in mangroves. *Biological Reviews*, **94**, 1547–1575.

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 and Bioinformatics*, **8**, 77–80.

Wang, J., **Yuan, J.**, **Yu, J.**, **Meng, F.**, **Sun, P.**, **Li, Y.** ***et al.*** (2019) Recursive paleohexaploidization shaped the durian genome. *Plant Physiology*, **179**, 209–219.

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

Wang, Y., **Zhu, H.** & **Tam, N.F.Y.** (2014) Polyphenols, tannins and antioxidant activities of eight true mangrove plant species in South China. *Plant and Soil*, **374**, 549–563.

Wang, Z.Z., **Guo, Z.X.**, **Zhong, C.R.**, **Lyu, H.M.**, **Li, X.N.**, **Duke, N.C.** ***et al.*** (2022) Genomic variation patterns of subspecies defined by phenotypic criteria: analyses of the mangrove species complex, *Avicennia marina*. *Journal of Systematics and Evolution*, **60**, 835–847.

Wheeler, E.A. & **Lehman, T.M.** (2000) Late Cretaceous woody dicots from the Aguja and javelina formations, big bend National Park, Texas, USA. *IAWA Journal*, **21**, 83–120.

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. *Innovation (China)*, **2**, 100141. Available from: https://doi.org/10.1016/j.xinn.2021.100141

Xie, M., **Zhang, J.**, **Tschaplinski, T.J.**, **Tuskan, G.A.**, **Chen, J.G.** & **Muchero, W.** (2018) Regulation of lignin biosynthesis and its role in growth-defense tradeoffs. *Frontiers in Plant Science*, **9**, 1427.

Xie, W., **Guo, Z.**, **Wang, J.**, **He, Z.**, **Li, Y.**, **Feng, X.** ***et al.*** (2023) Evolution of woody plants to the land-sea interface–the atypical genomic features of mangroves with atypical phenotypic adaptation. *Molecular Ecology*, **32**, 1351–1365.

Xu, S., **Guo, Z.**, **Feng, X.**, **Shao, S.**, **Yang, Y.**, **Li, J.** ***et al.*** (2023) Where whole-genome duplication is most beneficial: adaptation of mangroves to a wide salinity range between land and sea. *Molecular Ecology*, **32**, 460–475.

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.

Yang, G., **Zhou, R.**, **Tang, T.** & **Shi, S.** (2008) Simple and efficient isolation of high-quality total RNA from Hibiscus tiliaceus, a mangrove associate and its relatives. *Preparative Biochemistry & Biotechnology*, **38**, 257–264.

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

Yeats, T.H. & **Rose, J.K.C.** (2013) The formation and function of plant cuticles. *Plant Physiology*, **163**, 5–20.

Zhang, R., **Guo, Z.**, **Fang, L.**, **Zhong, C.**, **Duke, N.C.** & **Shi, S.** (2022) Population subdivision promoted by a sea-level-change-driven bottleneck: a glimpse from the evolutionary history of the mangrove plant Aegiceras corniculatum. *Molecular Ecology*, **31**, 780–797.