JIPB Journal of Integrative Plant Biology

Research Article
https://doi.org/10.1111/jipb.13625

Genomic analysis of *Nypa fruticans* elucidates its intertidal adaptations and early palm evolution[oo]

Weihong Wu[1†], Xiao Feng[1,2†], Nan Wang[1], Shao Shao[1], Min Liu[1], Fa Si[1], Linhao Chen[1], Chuanfeng Jin[1], Shaohua Xu[1], Zixiao Guo[1], Cairong Zhong[3], Suhua Shi[1] and Ziwen He[1*]

1. State Key Laboratory of Biocontrol and Guangdong Provincial Key Laboratory of Plant Resources, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China
2. Greater Bay Area Institute of Precision Medicine, School of Life Sciences, Fudan University, Guangzhou 511462, China
3. Hainan Academy of Forestry (Hainan Academy of Mangrove), Haikou 571100, China
†These authors contributed equally to this work.
*Correspondence: Ziwen He (heziwen@mail.sysu.edu.cn)

Weihong Wu

Ziwen He

ABSTRACT

Nypa fruticans (Wurmb), a mangrove palm species with origins dating back to the Late Cretaceous period, is a unique species for investigating long-term adaptation strategies to intertidal environments and the early evolution of palms. Here, we present a chromosome-level genome sequence and assembly for *N. fruticans*. We integrated the genomes of *N. fruticans* and other palm family members for a comparative genomic analysis, which confirmed that the common ancestor of all palms experienced a whole-genome duplication event around 89 million years ago, shaping the distinctive characteristics observed in this clade. We also inferred a low mutation rate for the *N. fruticans* genome, which underwent strong purifying selection and evolved slowly, thus contributing to its stability over a long evolutionary period. Moreover, ancient duplicates were preferentially retained, with critical genes having experienced positive selection, enhancing waterlogging tolerance in *N. fruticans*. Furthermore, we discovered that the pseudogenization of Early Methionine-labelled 1 (*EM1*) and *EM6* in *N. fruticans* underly its crypto-vivipary characteristics, reflecting its intertidal adaptation. Our study provides valuable genomic insights into the evolutionary history, genome stability, and adaptive evolution of the mangrove palm. Our results also shed light on the long-term adaptation of this species and contribute to our understanding of the evolutionary dynamics in the palm family.

Keywords: crypto-vivipary, genome stability, long-term adaptation, mangrove, *Nypa fruticans*, palm

Wu, W., Feng, X., Wang, N., Shao, S., Liu, M., Si, F., Chen, L., Jin, C., Xu, S., Guo, Z., et al. (2024). Genomic analysis of Nypa fruticans elucidates its intertidal adaptations and early palm evolution. J. Integr. Plant Biol. **66:** 824–843.

INTRODUCTION

Plant adaptation to extreme environments is an important topic in evolutionary biology. The intertidal zones at the interface between terrestrial and marine ecosystems are extreme environments. The plants growing there are subjected to periodic salinity stress, hypoxic stress, and other biotic and abiotic stresses (Giri et al., 2011). Mangroves are a group of 70 plant species around the world that inhabit these extreme environments, making them attractive model systems for evolutionary biology (Tomlinson, 2016). Previous studies have mainly focused on eudicot mangrove species and provided insights into their adaptive evolution, molecular convergence, and speciation (Xu et al., 2017a, 2017b; Lyu et al.,

2018; He et al., 2019, 2020; Feng et al., 2020, 2021, 2023; Hu et al., 2020; Miryeganeh et al., 2021; Zhu et al., 2023). Unlike all other mangrove species, *Nypa fruticans* is a monocot mangrove species and the sole species in its genus. This palm is one of the most ancient extant mangrove species, and has inhabited the intertidal zones since the Late Cretaceous (Schrank, 1987; Harley, 2006). The species has evolved several specialized traits to adapt to extreme environments, including crypto-vivipary and leaf bases with abundant aerenchyma. During vivipary, newly developed embryos germinate and grow out of the fruit coat (in viviparous plants, e.g., *Rhizophora, Bruguiera, Ceriops*, and *Kandelia*) or the seed coat (in crypto-viviparous plants, e.g., *Aegiceras, Avicennia*, and *Nypa*) before leaving the maternal plant, which can enhance their reproductive potential. This phenomenon is common in mangrove species but rare in other plants, and helps mangrove seedlings avoid high-saline environments during their earlier developmental stages (Tomlinson, 2016). Additionally, mangrove species living in intertidal environments must deal with a periodic lack of oxygen. Indeed, the thick stems of *N. fruticans* are buried in mud, leaving only its giant pinnate leaves exposed to the air. The leaf bases play the role of a "giant pneumatophore" to supply oxygen to the underground tissues instead of relying on the specialized aerating roots (pneumatophores) present in many other mangroves (Chomicki et al., 2014). These highly specialized phenotypic traits make *N. fruticans* a unique species for probing long-term plant adaptation strategies to extreme environments.

The genus *Nypa* branched from the other palms early, as supported by its extensive fossil records, primarily represented by pollen and fruit (Muller, 1968; Gregor and Hagn, 1982; Schrank, 1987; Gee, 2001; Harley, 2006), and molecular evidence (Yao et al., 2023; He et al., 2015). The appearance of the genus *Nypa* in the form of pollen (*Spinizonocolpites* Muller) dates back to the Late Cretaceous (~75 million years ago (mya)) (Muller, 1968; Schrank, 1987; Gee, 2001; Harley, 2006). The fruit and seed fossil records also date back to the upper Maastrichtian–Danian age (61–70 mya) (Gregor and Hagn, 1982; El-Soughier et al., 2011). The divergence times between *N. fruticans* and its relatives have been inferred from molecular data in various studies (Couvreur et al., 2011; He et al., 2015). Using DNA sequence data, restriction fragment length polymorphisms (RFLP) markers, and morphology, Couvreur et al. (2011) reconstructed a complete genus-level phylogeny of the palms (Arecaceae) and estimated that *N. fruticans* diverged from other palms 87–100 mya. He et al. (2015) estimated the divergence time between *N. fruticans* and other palm species at ~75.5 mya using transcriptome data. The age of the Arecaceae has been estimated in many studies (Janssen and Bremer, 2004; Couvreur et al., 2011; Silvestro et al., 2021) and has generally been taken as suggesting that the palms originated in the early Upper Cretaceous (85–100 mya). For instance, a recent study developed a Bayesian Brownian bridge model to estimate the ages of angiosperm families using the present diversity and ~15,000 fossils, reporting the age of the Arecaceae to be 92.3 (87.6–98.3) mya (Silvestro et al., 2021). By contrast, some genomic studies reported that the palms originated after the Cretaceous period (Zhao et al., 2018; Wang et al., 2021). Due to their economic importance, several studies have sequenced the genome of commercial palm crops, including the date palm (*Phoenix dactylifera*), oil palm (*Elaeis guineensis*), and coconut (*Cocos nucifera*), but few have focused on their origin and early evolution. Only limited physiological studies and transcriptome data have been reported for *N. fruticans* (Chomicki et al., 2014; He et al., 2015), and this lack of effective genomic data is a major obstacle to understanding the adaptive evolution of this species.

In this study, we presented a high-quality reference genome and assembly for *N. fruticans*, as part of a major sequencing project of mangrove genomes from around the world (He et al., 2022). We performed a comparative analysis among the genome of representative palm species and attempted to elucidate the low speciation rate and long-term sustained adaptation of *N. fruticans* in extreme environments. The *N. fruticans* genome will also provide a critical resource for studying palm and mangrove evolution.

RESULTS

The evolutionary history of the mangrove palm *N. fruticans*, informed by its genome

Genome sequencing, assembly, and annotation

We assembled a high-quality genome of *N. fruticans* (Figure 1A) by incorporating PacBio Single-Molecule Real-Time (SMRT) sequencing, Illumina-based short-read sequencing, and high-throughput chromosome conformation capture (Hi-C) technologies. We obtained a final assembly of *N. fruticans* of 489.38 Mb, with an N50 value of 25.77 Mb, which is consistent with an estimation of genome size (~486 Mb) based on *k*-mer analysis (Figure S1). The 17 pseudo-chromosomes cover 89.92% of the *N. fruticans* genome (Figures S2, S3; Table S1). We also examined the assembly integrity by aligning SMRT subreads and Illumina short reads to the genome, resulting in 94.24% of PacBio Continuous Long Reads (CLRs) and 94.41% of Illumina short reads successfully mapped back to the genome. A Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis using the embryophyta_odb10 lineage dataset showed that 97.0% of BUSCOs were confirmed as complete in the genome assembly (Table S2). These results suggested that our *N. fruticans* genome assembly was of high quality.

After masking repetitive sequences, we predicted 24,163 protein-coding genes in the *N. fruticans* genome using a combination of homology-based predictions, *de novo*-based predictions, and transcriptome data (Table 1). A BUSCO analysis showed that 1,303 (94.7%) of complete BUSCOs were identified in the gene models (Table S2). In addition, we functionally annotated 23,980 genes (99.24% of all protein-coding genes) using their sequences as queries against public databases (Tables S3, S4). We also predicted 635 transfer

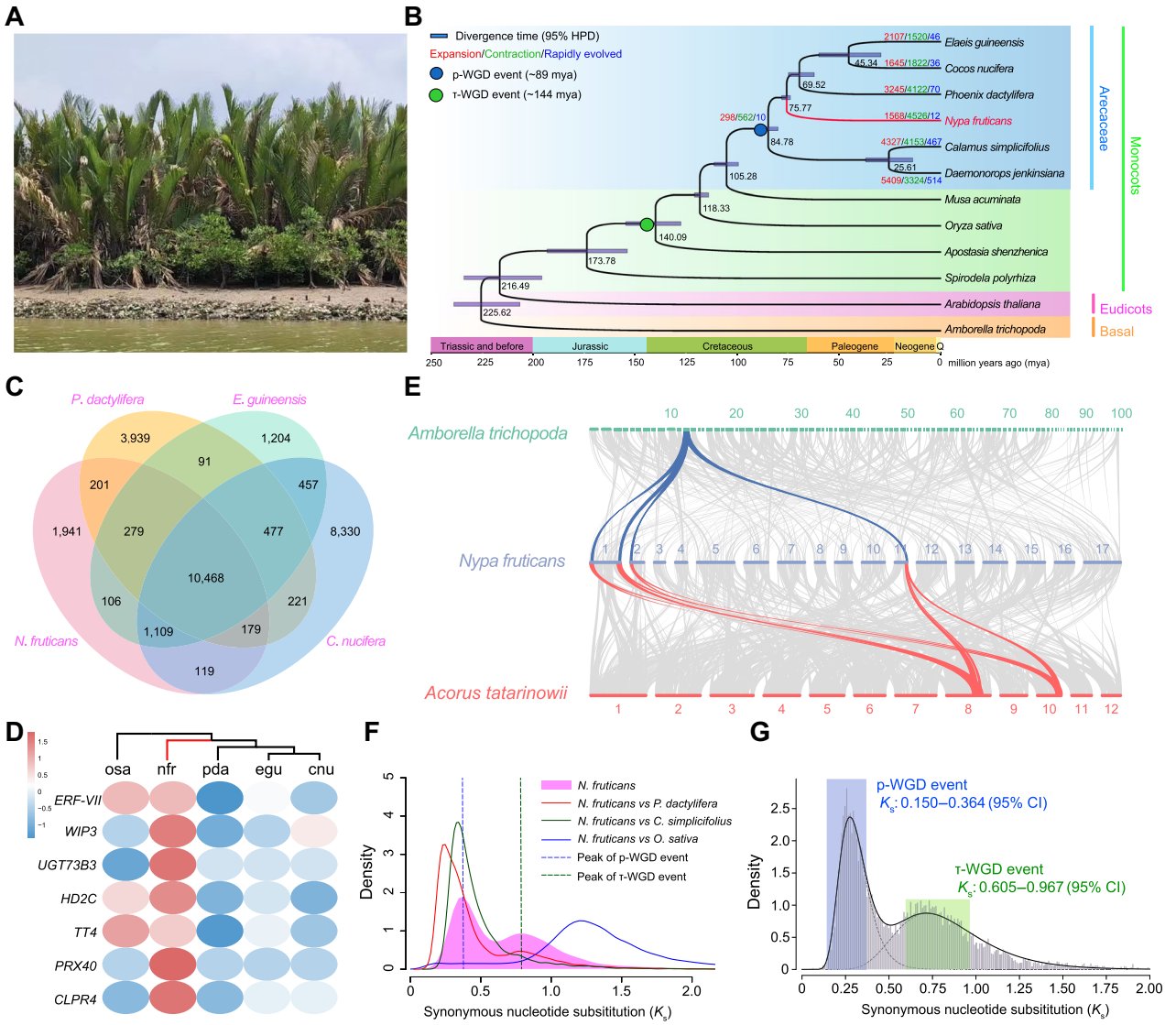

Figure 1. Phylogenetic relationship and whole-genome duplications (WGDs) of *Nypa fruticans* and other palms

(A) Representative photograph of *N. fruticans* in a mangrove. **(B)** Phylogenetic tree of 12 plant species with their estimated divergence time. Blue bars at the nodes indicate the 95% highest posterior density (HPD). Red, green, and blue numbers at each branch represent expanded, contracted, and rapidly evolved gene families, respectively, in each palm species and the common ancestor of these palms. The green circle and blue circle indicate the timing of the τ-WGD and p-WGD events. **(C)** Venn diagram showing the number of shared and unique gene families among four palms (*N. fruticans, Phoenix dactylifera, Elaeis guineensis*, and *Cocos nucifera*). **(D)** Heatmap representation of copy numbers for the *N. fruticans* expanded genes related to water-logging tolerance in five species (*Oryza sativa, N. fruticans, P. dactylifera, E. guineensis*, and *C. nucifera*), shown as *z*-score. The *z*-score was calculated as (gene copy number of each species − average gene copy number of all five species)/(standard deviation of the gene copy number of all five species). **(E)** Syntenic blocks among genomic regions in *N. fruticans, Amborella trichopoda*, and *Acorus tatarinowii*. **(F)** Synonymous substitution rate (K_s) distributions between the paralogs in *N. fruticans* and the orthologs in different species. **(G)** K_s distribution of the paralogs in *N. fruticans*. The distribution was split into two peaks by a fitted mixture model.

RNAs (tRNAs), 1,403 ribosomal RNAs (rRNAs), 386 small non-coding RNAs (snRNAs), and 159 microRNAs (miRNAs) in the genome (Table S5).

Phylogenetic relationships, gene family evolution, and whole-genome duplication

To ascertain the origin of *N. fruticans* and palms, we reconstructed a phylogenetic tree using RAxML-NG with the GTR+GAMMA+I model, based on 545 single-copy orthologs from *N. fruticans*, five palm species (*Calamus simplicifolius, Daemonorops jenkinsiana, P. dactylifera, E. guineensis*, and *C. nucifera*), and six other angiosperm species (*Amborella trichopoda, Arabidopsis thaliana, Spirodela polyrhiza, Apostasia shenzhenica, Oryza sativa*, and *Musa acuminata*). We confirmed that *N. fruticans* is a basal species of the Arecaceae excluding the subfamily Calamoideae, based on the maximum-likelihood (ML) tree with full bootstrap supports (Figures 1B, S4). We estimated the origin time of palms and

Table 1. Statistics for the *Nypa fruticans* genome assembly and gene annotation

Assembly feature and Genome annotation	Statistics
Estimated genome size (by *k*-mer analysis) (Mb)	486.24
Scaffold N50 (Mb)	25.77
Longest scaffold (Mb)	38.19
Assembled genome size (Mb)	489.38
Benchmarking Universal Single-Copy Orthologs (BUSCO)	97.0%
Percentage of repeat region	42.88%
Number of predicted gene models	24,163
Average exons per gene	5.92

the divergence time between *N. fruticans* and the other palms using the MCMCTREE program with sequence data and reliable fossil records. We thus estimated the origin time of *N. fruticans* to be about 75.77 mya, while the diversification of extant lineages of palms started at approximately 84.78 mya (Figure 1B). The phylogenetic analysis suggested that the origin of palm plants probably occurred during the early Upper Cretaceous period. The time of palm origin obtained in this study was earlier than previous genomic studies (Zhao et al., 2018; Wang et al., 2021), but closer to the time estimated by the fossil record (Harley, 2006; Silvestro et al., 2021), therefore reconciling the paleontological and genomic studies concerning the origin time of this family.

We conducted a comparative analysis of gene family evolution among *N. fruticans* and other palm species: *P. dactylifera* (Hazzouri et al., 2019), *E. guineensis* (Singh et al., 2013), and *C. nucifera* (Lantican et al., 2019). Among the gene families examined, 10,468 gene families were found to be shared between *N. fruticans* and other palms, while 1,941 gene families only existed in *N. fruticans* (Figure 1C). These *N. fruticans*-specific gene families play a role in various functions, including the "peptide metabolic process," "amide biosynthetic process," and "translation," as informed by a Gene Ontology (GO) term enrichment analysis (Table S6). In comparison with its palm relatives, the *N. fruticans* genome exhibited an expansion in 1,568 gene families and contraction in 4,526 gene families. Notably, the expanded gene families of *N. fruticans* were significantly enriched in the GO functional categories "response to stress," "DNA-templated transcription," and "chromatin organization" (Table S7). As *N. fruticans* often experiences waterlogging stress during rising tides, we examined the waterlogging-response genes in *N. fruticans*. We identified expansions of genes including ETHYLENE RESPONSE FACTOR VII (*ERF-VII*), WOUND INDUCED POLYPEPTIDE 3 (*WIP3*), and the UDP-GLCOSYL TRANSFERASE 73B3 (*UGT73B3*) (Figure 1D).

Whole-genome duplications (WGDs) are prevalent in plants, especially in angiosperms. Previous studies have determined that two rounds of WGD events have occurred in all palms since the monocot–eudicot divergence (Al-Mssallem et al., 2013; Jiao et al., 2014). The first event was τ-WGD, shared by almost all monocots, excluding Alismatales and Acorales; the second event was p-WGD, which can only be detected in the palm species (Figure S5). The role of the p-WGD event in the evolution of the palm species and the time at which it occurred remains poorly understood. We first inferred and dated these two WGD events by combining syntenic, synonymous substitution rates (K_s) distribution, and phylogenetic methods. For the syntenic analysis, the 4:1 collinear relationship between *N. fruticans* and *A. trichopoda* suggested that *N. fruticans* experienced two rounds of WGD events after its divergence from *A. trichopoda* (Figure 1E). The 4:2 collinear relationship between *N. fruticans* and the monocot basal species *Acorus tatarinowii* (Shi et al., 2022) (Figure 1E) and the 4:4 collinear relationship between *N. fruticans* and the Poales species *Ananas comosus* (Ming et al., 2015) also supported this inference (Figure S6). The two peaks in the K_s distribution also indicated that *N. fruticans* underwent two WGD events in its evolutionary history. We noted that the K_s peak of the *N. fruticans* paralogs having arisen from the p-WGD event is larger than that of comparison between *N. fruticans* and *C. simplicifolius*, demonstrating that p-WGD event occurred before the divergence of the Calamoideae and Nypoideae (Figure 1F). The results confirm that p-WGD occurred in a common ancestor of all extant palms.

We then used a fitted mixture model to distinguish the K_s distributions of paralogs between τ-WGD and p-WGD (Figure 1G). Under a Gaussian Mixture Model (GMM), the K_s distribution of the p-WGD ranged from 0.150 to 0.364 (95% confidence interval (CI)), while the τ-WGD ranged from 0.605 to 0.967 (95% CI) in *N. fruticans*. Similarly, we inferred the K_s distributions of τ-WGD and p-WGD in the other three palm species (Figure S7). To precisely estimate the absolute timing of the two WGDs, we performed a molecular clock analysis by concatenating gene families that reflected p-WGD and τ-WGD in their gene trees, calibrated using between-species divergence times (see Materials and Methods). The τ-WGD occurred ~144 (95% CI 131–158) mya, while the p-WGD occurred ~89 (95% CI 82–96) mya.

Maintenance of genome stability in *N. fruticans*

Nypa fruticans genes show a lower mutation rate and stronger purifying selection

Despite the early origin of the Nypoideae, it has an extremely low level of species diversity with only one extant species, *N. fruticans*. This palm has a stable morphological structure and ecological preferences (Schrank, 1987; Plaziat et al., 2001; Tomlinson, 2016), leading us to investigate whether the genome of *N. fruticans* reflects its slow evolution over a long history (Figure S8).

We discovered that the substitution rates of palm species are lower than those of other species based on the ML phylogenetic tree reconstructed using 545 single-copy orthologs, indicating that the evolution of genes in palm species is much slower than in other related plants. We discovered that the protein-coding sequences in grasses

(Poaceae) evolved ~2.5-fold faster than those in palm species based on the ML phylogenetic tree reconstructed using fourfold degenerate (4d) sites (as shown in Figure S9). This finding concurs with previous studies that estimated the evolutionary rate of grass ALCOHOL DEHYDROGENASE (*ADH*) sequences to be about 2.5-fold faster than their palm orthologs (Gaut et al., 1996). A Tajima's relative rate test revealed a lower substitution rate in *N. fruticans* than in the other palm species (Table S8). The results support a low mutation rate for *N. fruticans*.

Additionally, we estimated the ω (K_a/K_s) among different palm lineages using CodeML with the free-ratio model and determined that *N. fruticans* has the lowest ω value (K_a/K_s = 0.266) among all palm species based on 2,680 single-copy orthologs (Figure 2A). This result suggests that genes in the *N. fruticans* genome have undergone stronger negative selection than those in other palm species. It has been well documented in previous studies that positive selection tends to be more prevalent in rapid evolutionary radiations (Nevado et al., 2016, 2019). Conversely, species-poor lineages are typically associated with a lower frequency of positive selection. To study whether the percentage of positive selection is significantly less common in the genus *Nypa*, we analyzed 2,680 single-copy orthologs among palm lineages using a branch-site model to determine whether any were under positive selection. Different frequencies of positive selection in each lineage may be caused by changes in the intensity of any negative selection (Chen et al., 2022a). We observed a significantly lower percentage of genes evolving under positive selection in the species-poor palm lineages (*Nypa* and *Elaeis*) than in the other palm lineages (Welch's *t*-test, *P*-value = 0.00259; Figure S10). These results indicated that the *N. fruticans* genome has undergone strong purifying selection in its evolutionary history.

Limited LTR-RT insertions and recent gene duplications contribute to maintaining a stable genome

Transposable elements (TEs) insertions and gene duplications are the major forces driving plant genome evolution (Wicker et al., 2007; Conant and Wolfe, 2008), shaping the genome architecture, and contributing to speciation (Brawand et al., 2014; Edger et al., 2015; Serrato-Capuchina and Matute, 2018). We first estimated the contents for TEs and long terminal repeat retrotransposons (LTR-RTs) in the genome of the palms *N. fruticans, C. simplicifolius* (Zhao et al., 2018), *D. jenkinsiana* (Zhao et al., 2018), *P. dactylifera* (Hazzouri et al., 2019), *E. guineensis* (Singh et al., 2013), and *C. nucifera* (Lantican et al., 2019). We discovered that LTR-RTs covered 22.38%–68.91% of their genomes (Table S9). *Nypa fruticans* has the lowest percentage of TEs and LTR-RTs among the palm plants (Figure 2A). To further explore the lower TE load in *N. fruticans*, we estimated the insertion time distributions of intact LTR-RTs (Figure 2A). *Nypa fruticans* exhibited a smooth distribution with a low level of insertion times, in contrast with other palm species, which showed a peak indicating a recent burst of LTR-RTs in their genomes. The lack of a recent burst of LTR-RTs results in a lower LTR content in the *N. fruticans* genome. In addition, we estimated the relative abundances of solo and intact LTRs (*S*:*I*) and the ratio of transition: transversion (T_s:T_v) for the sequences of LTRs (Figure S11). We obtained a higher *S*:*I* value and medium T_s:T_v value in *N. fruticans* than in other palm species. The higher *S*:*I* value in *N. fruticans* reflects a more efficient LTR-RTs removal mechanism within this genome (Hawkins et al., 2009). We also identified and classified all LTR-RTs into three groups: *Copia, Gypsy*, and unclassified. We detected a significantly lower relative content of *Copia* elements in the *N. fruticans* genome (comprising only ~2% of the genome) than in other species (16%–45% of other palm genomes) (Figure S11).

We also estimated the number of recent gene duplication events ($K_s < 0.3$) among six palm species to determine the gene duplications that have occurred since the branching of *N. fruticans* from other palm species. *Nypa fruticans* appears to have undergone significantly fewer recent gene duplication events than its relatives (Figure 2B). We then investigated the selection pressure of these recently duplicated genes among the palm species by calculating the K_a/K_s value between pairs of duplicates. The recently duplicated genes in *N. fruticans* had significantly lower K_a/K_s values than those in the other palm species (Figure 2C), suggesting that the recently duplicated genes in *N. fruticans* have undergone strong purifying selection.

Decreased loss and preferential retention of ancient WGD-derived genes in the *N. fruticans* genome

Fractionation and diploidization following WGD events cause most duplications to be converted back into a single-copy state (Lynch and Conery, 2000). We determined that the *N. fruticans* genome has comparatively fewer losses of ancient WGD genes (Figure 2D), suggesting that fractionation and diploidization may be suppressed in the *N. fruticans* genome. We first looked for orthologs based on the best hit from a reciprocal BLAST search and identified 11,502 orthogroups among four species (*N. fruticans, P. dactylifera, E. guineensis*, and *C. nucifera*), which were used to reconstruct the ancestral gene sets of palms. We then identified the genes retained from p-WGD in each genome based on the above GMM analysis. Only 1,252 duplicated genes from p-WGD were lost from the *N. fruticans* genome since the divergence between *N. fruticans* and the other palm species, while the other three palms lost approximately twice as many (Figure 2D).

Nypa fruticans showed much fewer p-WGD duplicate losses than the other palms, indicating that more of the duplicates were retained. In fact, different ecological pressures drive distinct preferential retention following WGD events (Van de Peer et al., 2017; Robertson et al., 2017). Since *N. fruticans* diverged from other palms, it has lived in intertidal zones for over 75 million years, which is different from the habitat of other palm species. We therefore investigated the p-WGD-derived duplicates among these four palm species. We observed that 736 gene families were commonly

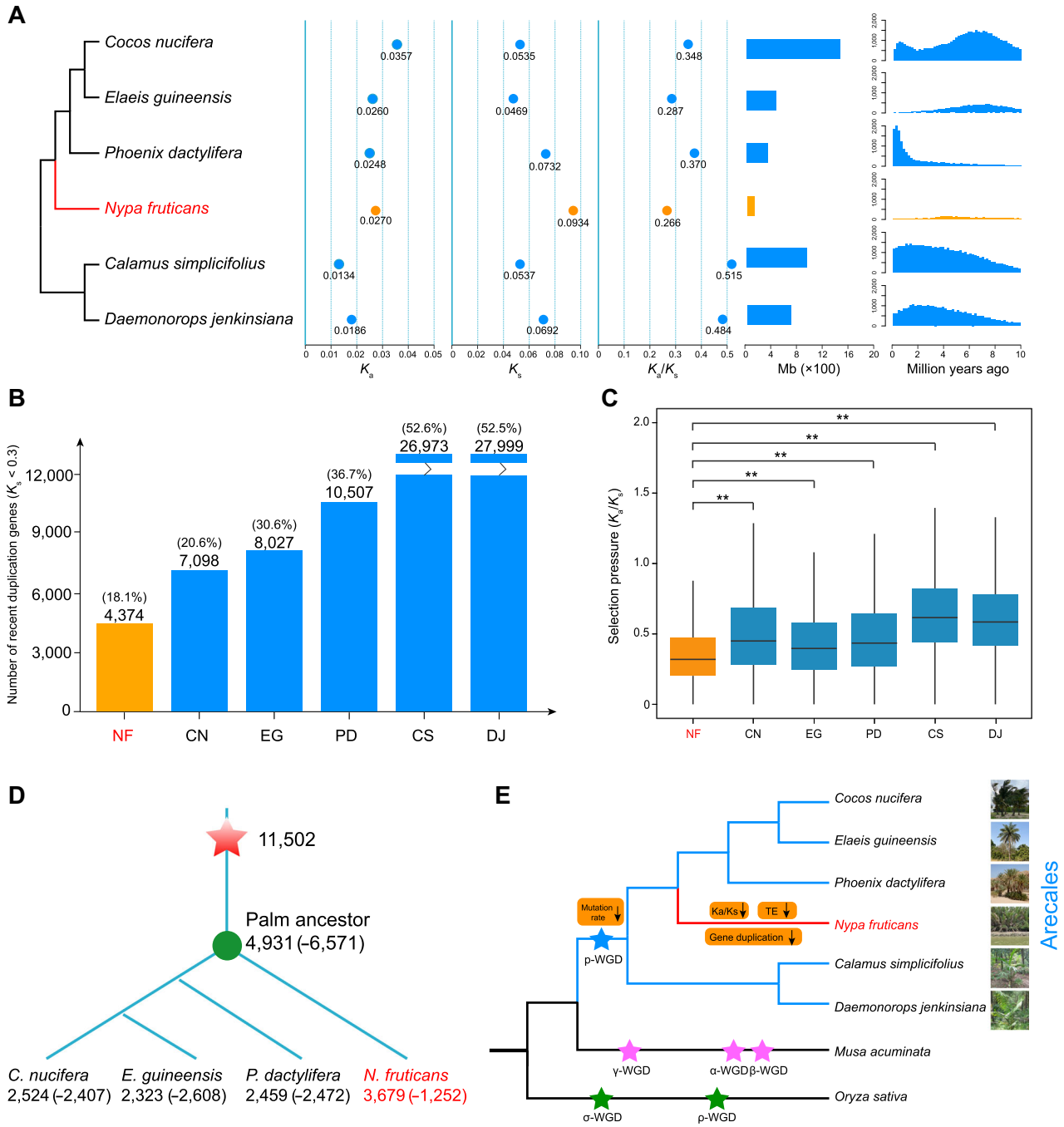

Figure 2. The *Nypa fruticans* genome underwent strong purifying selection and maintained its stability

(A) Summary of the evolutionary rate and fraction of the genome represented by long terminal repeat. retrotransposons (LTR-RTs). The first plot represents the nonsynonymous substitution rate (K_a) for each species. The second plot represents the nonsynonymous substitution rate (K_s) for each species. The third plot represents the ratio of the nonsynonymous mutation rate to the synonymous mutation rate (K_a/K_s) for each species. The fourth plot represents the overall cumulative length of LTR-RTs. The fifth plot represents the distribution of the insertion times of the intact LTR-RTs. **(B)** Numbers and proportions of recently duplicated genes in the six palm species. (NF, *N. fruticans*; CN, *Cocos nucifera*; EG, *Elaeis guineensis*; PD, *Phoenix dactylifera*; CS, *Calamus simplicifolius*; DJ, *Daemonorops jenkinsiana*). **(C)** Selective pressure. on the tandem duplicated genes in six palm species, shown as a box and whisker plot. A significant difference in the selective pressure on the tandem duplicated genes between species is marked with double asterisks (P-value < 0.001). **(D)** Model of the divergent diploidization in different palm lineages. The red star indicates the p-WGD event that occurred in the common ancestor of palms. The numbers under each species name represents the number of extant duplicated gene pairs and lost duplicated genes. **(E)** Summary of the key differences in the genomic evolution of *N. fruticans* compared with other palm species.

retained in all palm species, with 946 gene families being retained only in the *N. fruticans* genome (Figure S12). We further divided the pairs of p-WGD duplicates from *N. fruticans* into three clusters: NR (*N. fruticans*-specific gene retentions, representing only *N. fruticans* retained their p-WGD-derived duplicates in these gene families), PR (at least one of the other palm species but not all palm species retained their p-WGD-derived duplicates in these gene families), and AR (all palm species retained their p-WGD-derived duplicates in these gene families). The duplicated genes in the NR group had a significantly higher ω (K_a/K_s) value and sequence divergence than those in the PR and AR groups (Figure S13). These results indicate that p-WGD-derived genes retained only in *N. fruticans* underwent a relaxed purifying selection and evolved faster.

Overall, our findings show that *N. fruticans* has experienced limited LTR-RT insertions, gene duplications, and a lower loss of ancient WGD-derived genes since it diverged from other palm species (Figure 2E). We suggested that *N. fruticans* has maintained a stable genome during its long-term adaptation to the intertidal zone, which may also have contributed to its slow evolutionary pace.

Intertidal adaptation through phenotypic innovation

Genetic contribution of the palm-specific WGD events to the early evolution of the palms

The early evolution of the palms plays a role in shaping their specialized traits (Tomlinson, 2006; Couvreur et al., 2011). The p-WGD, the palm-specific WGD, may have contributed to their evolution and adaptation. We performed a GO term enrichment analysis of the retained duplicates derived from p-WGD and the earlier τ-WGD shared by most monocots to compare their function (Figure 3A; Table S10). The GO terms "biological regulation," "transcription regulator activity," and "kinase activity" were significantly enriched among these retained genes from both WGD events; however, the terms "signal transduction," "vesicle-mediated transport," and "gravitropism" were significantly enriched among the retentions from the p-WGD event only, not from the τ-WGD event. For the genes associated with the term "signal transduction," many retained genes were involved in pathways responsive to salt stress, osmotic stress, dehydration, and other abiotic and biotic stresses (Table S11). We focused on the gene family encoding calcineurin B-like protein-interacting protein kinase (CIPK), which functions in Ca^{2+}-dependent kinase-related signaling pathways related to abiotic stress responses (Luan, 2009; Chen et al., 2022b). We reconstructed a phylogenetic tree of the *CIPK* gene families in six species (*A. thaliana, O. sativa, N. fruticans, P. dactylifera, E. guineensis*, and *C. nucifera*) and classified them into 14 subgroups (Groups A–N). This phylogenetic analysis showed that eight subgroups of *CIPK*s were expanded by the p-WGD event in palm species (Figure 3B).

Furthermore, we investigated the role of p-WGD in the evolution of lignin biosynthesis genes in palms. We identified more copies of FERULATE 5-HYDROXYLASE (*F5H*) and CINNAMIC ACID 4-HYDROXYLASE (*C4H*) in palm species (Table S12), and that the *C4H*, COUMARATE 3-HYDROXYLASE (*C3H*), CINNAMOYL-COA REDUCTASE (*CCR*), and *F5H* genes retained their duplicates after p-WGD (Figure 3C). Woody plants have more syringyl (S)-unit lignin than herbaceous plants (Stewart et al., 2009); F5H is a key enzyme in the S-units lignin biosynthesis pathway (Figure 3C). All palm species retained two copies of *F5H* derived from the p-WGD event and another copy derived from tandem duplication. Two duplicates derived from p-WGD have undergone positive selection. Amino acid 232 in F5H was inferred as the positive selection site. The change of Pro-232 to Met might alter the folding of the resulting protein and affect its function (Figure 3D).

Preferential retention and positive selection of key gene families promoting waterlogging tolerance in *N. fruticans*

Nypa fruticans, living in intertidal environments, has evolved to overcome the challenge of oxygen deprivation. It forms leaf bases with abundant aerenchyma, creating a "giant pneumatophore" that supplies oxygen to its underground tissues (Chomicki et al., 2014). The genes specifically retained in *N. fruticans* underwent a relaxed purifying selection and evolved faster, which may have contributed to its adaptation. We performed a Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of the genes retained only in *N. fruticans* and determined that they were mainly enriched in the pathways "Tight junction," "Aminoacyl-tRNA biosynthesis," "Gap junction," and "Sulfur metabolism." The pathways "Transcription factors," "Apoptosis," and "Metabolism of xenobiotics by cytochrome P450" were also overrepresented (Table S13). Glutathione *S*-transferase (GST) in this latter pathway plays a role in the detoxification of reactive oxygen species (ROS). Genes in the "Transcription factors" pathway, such as ETHYLENE RESPONSE FACTOR (*ERF*), *MYB*, and SALT TOLERANCE ZINC FINGER (*STZ*), are crucial for the response to waterlogging stress (Figure 4A, B; Table S14).

Genes from Group VII *ERF* (*ERF-VII*) are involved in sensing the hypoxia stress (Gibbs et al., 2011). These genes include SUBMERGENCE 1 A (*SUB1A*) in rice (Xu et al., 2006) and RELATED TO APETALA 2.2 (*RAP2.2*) in Arabidopsis (Hinz et al., 2010). We searched for *ERF-VII* family genes among palm genomes based on the AP2 domain and a conserved motif (MCGG) at the N-terminal region. After a strict alignment, we identified 14 *ERF-VII* genes in *N. fruticans*, more than almost all other palm species except for *P. dactylifera*. Half of the *ERF-VII* duplicates were derived from the p-WGD event and the others were derived from the τ-WGD and tandem duplication events (Figures 4C, S14). The biosynthesis of suberin can facilitate the formation of the radial oxygen loss barrier, contributing to waterlogging tolerance in plants (Watanabe et al., 2013; Vishwanath et al., 2015). We summarized the suberin biosynthesis genes reported in the literature (Vishwanath et al., 2015). We identified

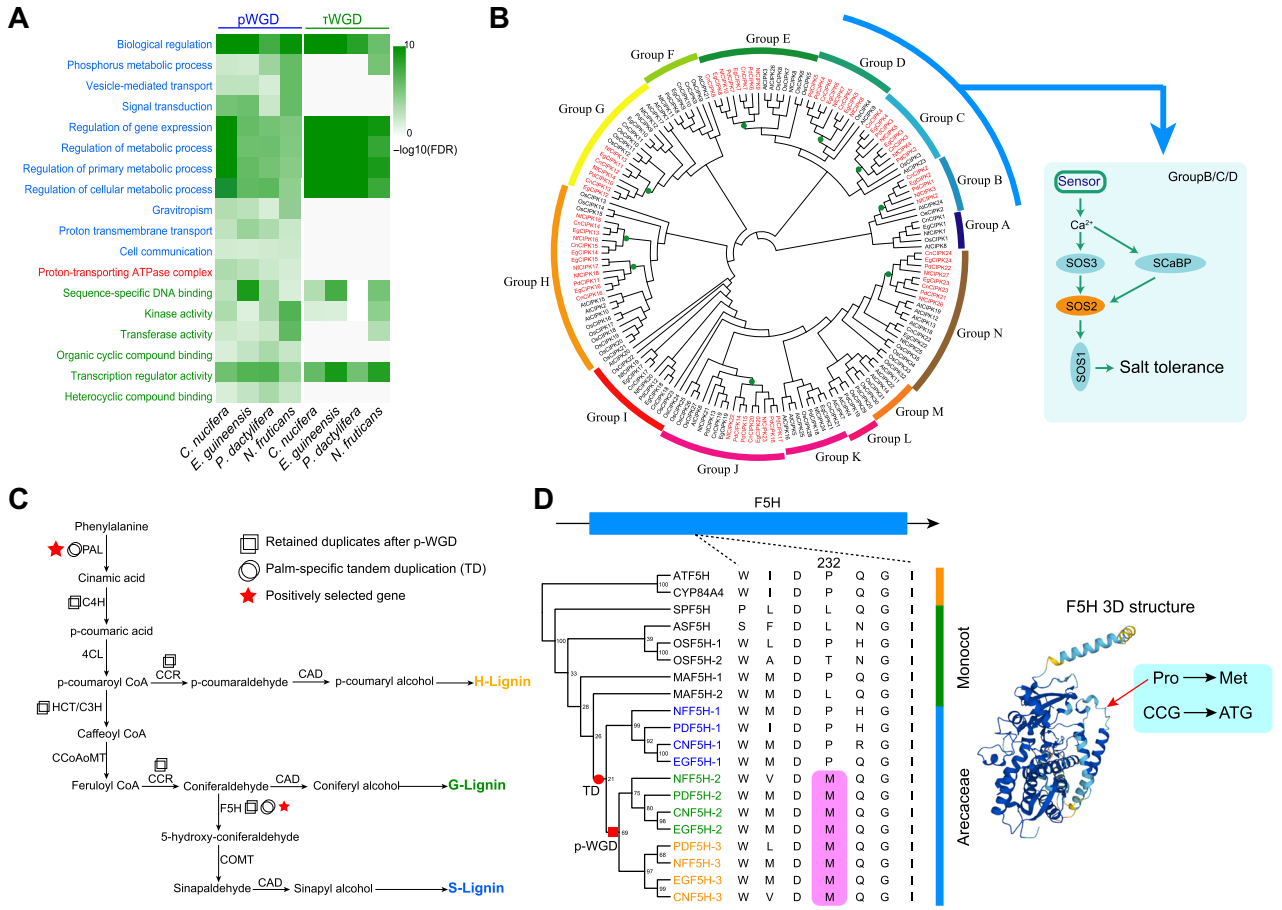

Figure 3. Genetic contribution of the palm-specific p-WGD event to the early evolution of the palm species

(A) Heatmap representation of Gene Ontology (GO) enrichment analysis for genes retained following the two whole-genome duplication (WGD) events. The blue, red, and green terms are Biological Process, Cellular Component and Molecular Function terms, respectively. **(B)** Phylogenetic tree of the CBL-INTERACTING PROTEIN KINASE (*CIPK*) gene families among the six species, including (Arabidopsis, rice, and four palm species). The *CIPK*s were classified into 14 groups (Groups A–N). The green circles at the nodes indicated the duplications generated by the p-WGD event. The *CIPK* members in the B, C, and D groups function in salt tolerance. **(C)** Diagram of the major lignin biosynthesis pathway, showing the three kinds of lignin (H-lignin, G-lignin, and S-lignin). The expansion of key genes is related to p-WGD and palms-specific tandem duplications. *PAL* and *F5H* underwent positive selection. **(D)** Phylogenetic tree of *F5H*, encoding a key enzyme in S-lignin biosynthesis. The p-WGD and tandem duplication contributed to the expansion of *F5H* among palm species. The pairs of duplicates are postulated to be under positive selection after p-WGD and the P232M variant underwent positive selection (P-value < 0.05). The protein structure of F5H in Arabidopsis was predicted by AlphaFold. The Pro-232 site of F5H plays a crucial role in its folding, while the variant Met-232 in palm species might have altered protein folding and functions.

two key gene pairs (*Nfr009150* and *Nfr019342*, which encode orthologs of the CYTOCHROME P450 FAMILY 86 A; *Nfr009721* and *Nfr015937*, which are 3-KETOACYL-COA SYNTHASE 2 orthologs) in the suberin biosynthesis pathway that were specifically retained in *N. fruticans* following the p-WGD event (Figure S15; Table S14). Additionally, we inferred that nine genes responding to waterlogging stress have probably undergone positive selection in *N. fruticans* (Figure 4D; Table S15). Among them, *Nfr011854* (GLUTATHIONE REDUCTASE, *GR*), *Nfr013963* (SUPEROXIDE DISMUTASE, *SOD*), and *Nfr011627* (CATALASE, *CAT*) are genes encoding key enzymes in ROS detoxification. *Nfr022064* (*ERF-VII*), a critical transcription factor gene in sensing waterlogging stress, underwent positive selection in *N. fruticans*.

To investigate the mechanism underlying waterlogging tolerance in *N. fruticans*, we conducted transcriptome deep sequencing (RNA-seq) of leaves and roots at four time points during waterlogging: 0 h (T_1), 12 h (T_2), 24 h (T_3), and 72 h (T_4) (Table S16). We compared the transcriptomes at the T_2, T_3, and T_4 time points to those of T_1, resulting in the identification of 3,271 differentially expressed genes (DEGs) in leaves and 2,713 DEGs in roots. Many of the upregulated genes play crucial roles in the waterlogging-response pathways. In leaves, genes associated with ethylene signaling, response to hypoxia, and ROS detoxification were upregulated in the waterlogging treatment groups (Figure 4E). In roots, genes involved in anaerobic metabolism, response to hypoxia, response to oxidative stress, and the biogenesis of secondary cell walls were upregulated in response to waterlogging treatment (Figure 4F).

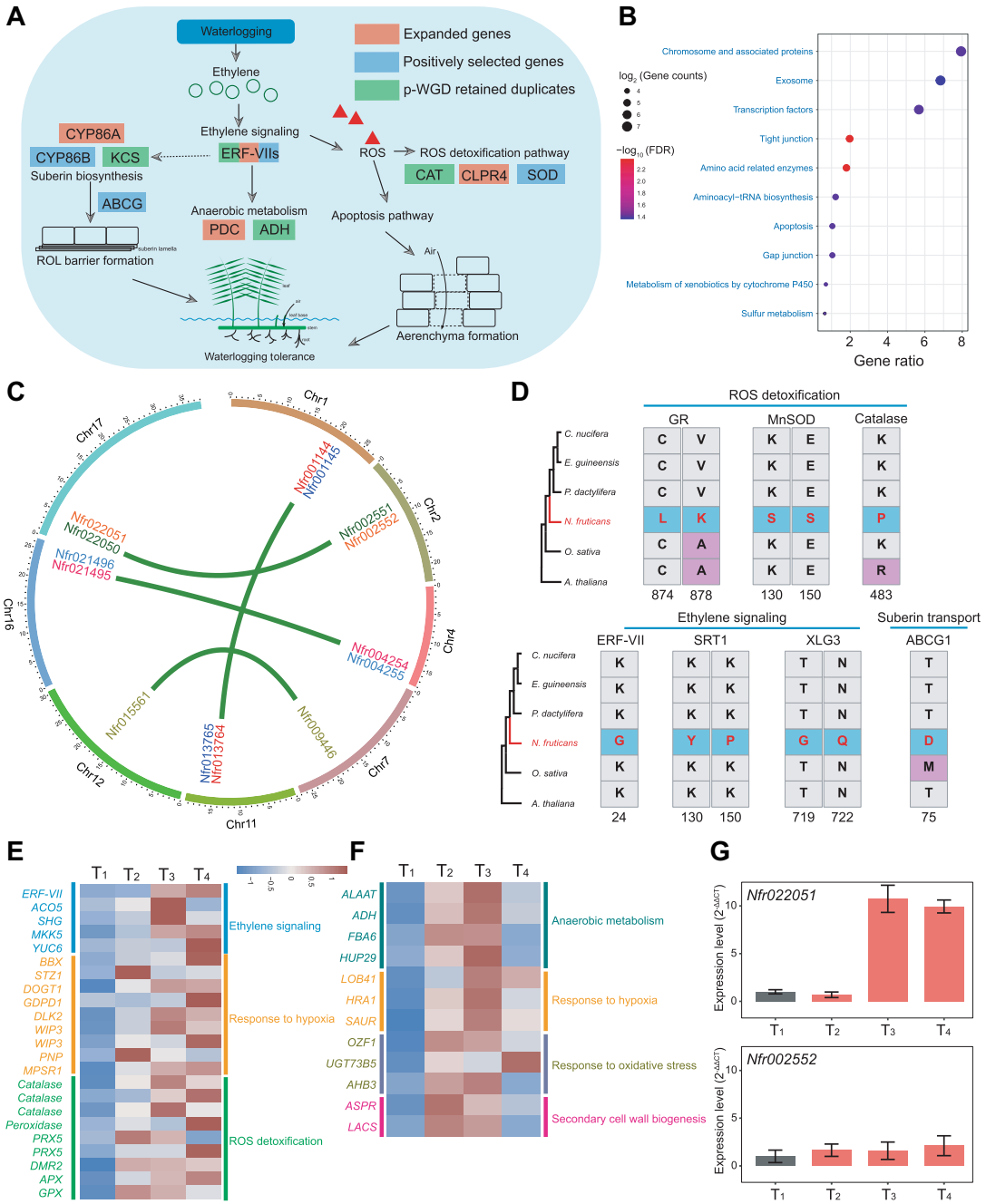

Figure 4. Distinct gene retention preference and adaptive genes in *Nypa fruticans* under intertidal pressures

(A) Diagram summarizing the major pathways involved in waterlogging tolerance in *N. fruticans*: ethylene signaling, reactive oxygen species (ROS) detoxification, hypoxia response, and suberin biosynthesis. The expanded genes are marked by red boxes, positively selected genes by blue boxes, and p-WGD-retained duplicates by green boxes. **(B)** Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of the retained genes belonging to the NR group (duplicate genes only present in *N. fruticans*). **(C)** Position and collinear relationship of *ERF-VII* genes in the *N. fruticans* genome. Green lines indicate gene pairs derived from the p-WGD event and gene names in the same color represent the p-WGD paralogs. **(D)** Positively selected genes in the waterlogging tolerance pathway and their putative selected sites **(E)** Heatmap representation of expression levels for the indicated genes, showing that the genes upregulated in leaves subjected to waterlogging treatment are involved in ethylene signaling, response to hypoxia, and ROS detoxification (T_1, control group; T_2, 12 h of waterlogging; T_3, 24 h of waterlogging; T_4, 72 h of waterlogging). The expression levels are shown as *z*-score normalized transcript per million (TPM) values. **(F)** Heatmap representation of expression levels for the indicated genes, showing that the genes upregulated in roots subjected to waterlogging are involved in anaerobic metabolism, response to hypoxia, response to oxidative stress, and secondary cell wall biogenesis. **(G)** Quantitative PCR with reverse transcription (RT-qPCR) analysis of *Nfr022051* (an *ERF-VII* coding gene) and its p-WGD paralog *Nfr002552* transcript levels, using the $2^{-\Delta\Delta CT}$ method. Values are means ± *SD* from three biological replicates.

Among these DEGs, we identified 1,461 examples of gene retention from the p-WGD event in leaves and roots. Notably, the expression levels of the expanded gene families, such as *ERF-VII* and *WIP3*, and the positively selected genes, including *ERF-VII* and *CAT* were also upregulated during waterlogging treatment. The expression level of *Nfr022051* (one of the *ERF-VII* genes), a positively selected p-WGD retention gene, increased in T_3 and T_4 relative to T_0 (Figure 4E). By contrast, its p-WGD paralog, *Nfr002552*, exhibited a different expression pattern and was not upregulated by waterlogging. The results of a reverse transcription quantitative PCR analysis (RT-qPCR) further supported this inference (Figure 4G). The distinct expression pattern of *Nfr022051* and its p-WGD paralog, *Nfr002552*, points to the neofunctionalization of these p-WGD retention genes. Collectively, these findings suggest that the neofunctionalization of preferential WGD retentions and the positive selection of key gene families may contribute to the adaptation of *N. fruticans* to waterlogging stress.

Pseudogenization of EARLY METHIONINE-LABELLED 1 (*EM1*) and *EM6* is related to crypto-vivipary in *N. fruticans*

Seed dormancy and germination are controlled by complex genetic networks and environmental factors (Graeber et al., 2012; Tuan et al., 2018). The absence of seed dormancy allows mangrove seeds to germinate before they leave their parents, improving seedling survival rates in stressful intertidal environments. To investigate the genetic factors underlying this crypto-vivipary, we searched for genes specifically lost from its genome. We used the *O. sativa* gene annotation data as a reference to detect gene losses from palm genomes. We identified 226 genes that were specifically lost in *N. fruticans* (Table S17), five of which were involved in seed development (Table S18).

Among these genes, we focused on the homologs of Early Methionine-labelled (*EM*), encoding a crucial seed dormancy regulator (Gaubier et al., 1993; Finkelstein and Lynch, 2000; Manfre et al., 2009). We found homologs for two *EM* genes (*EM1* and *EM6* in *A. thaliana*) in each of the other palm genomes (*P. dactylifera, E. guineensis*, and *C. nucifera*) but none in *N. fruticans*. We aligned the sequences of the *EM1* and *EM6* coding regions in other palms to the *N. fruticans* genome and identified three putative pseudogene regions (Table S18). *EM1* and *EM6* homologous Region 2 (*NfEM2*) and Region 3 (*NfEM3*) in the *N. fruticans* genome showed significant pseudogenization, with premature termination codon mutations or large fragment deletions (Figure S16). The termination codon of *NfEM1* was converted into "TAC," which can delay translation termination and potentially disrupt gene function (Figure 5A). If *NfEM1* had undergone pseudogenization, it would have evolved under neutral. We aligned *NfEM1* to its orthologs and performed an evolutionary rate analysis using RELAX. The results revealed a significant relaxation of natural selection, supporting the hypothesis that *NfEM1* has been pseudogenized (Figure 5B). To eliminate potential biases stemming from sampling, sequencing, and assembly, we mapped the resequencing data of 12 individuals to the reference genome, confirming the fixation of *EM1* and *EM6* pseudogenization within the *N. fruticans* population (Figure S16). Overall, these results indicated no functional EM1 and EM6 proteins being translated in *N. fruticans* that could underlie the crypto-vivipary in this species (Figure 5C).

DISCUSSION

As the only monocot mangrove species and the sole species in its genus, *N. fruticans* has developed unique adaptive traits for surviving its intertidal habitat over the 75 million years since its divergence from other palms in the Late Cretaceous (Figures S17, S18; Table S19). The genome sequencing of *N. fruticans* has shed light on its evolutionary history, genome stability, and intertidal adaptation. Since its origin approximately 75 mya, *N. fruticans* has maintained morphological and ecological stability, exhibiting a low level of species diversity, and is regarded as a living fossil. The stable genome of *N. fruticans* is intriguing, particularly as it is the only living species in the Nypoideae subfamily. In our comparative genomic analyses between *N. fruticans* and its relatives, we discovered that *N. fruticans* had a low mutation rate and had evolved slowly over its long history. Furthermore, the genome of *N. fruticans* has undergone strong purifying selection, with limitations on its LTR-RT insertions and recent gene duplications, and preferential retention of ancient p-WGD-derived genes, which contributed to its genome's stability.

As the basal species among woody palms (Asmussen et al., 2006; Yao et al., 2023), *N. fruticans* is a unique and pivotal species for exploring the woody characteristics in this family. Moreover, due to its early origin and stable genome evolution, *N. fruticans* has emerged as a critical species for studying the early genomic evolution of palms. This species also possesses many unique mangrove adaptative traits worthy of study; for instance, its leaf bases with abundant aerenchyma allow it to adapt to the waterlogging stress of intertidal zones (Chomicki et al., 2014). As the sole monocot mangrove species, studying the adaptive evolution of *N. fruticans* is crucial for understanding plant adaptations to intertidal environments.

The high-quality assembly genome of *N. fruticans* presented here provides valuable insights into the early evolution of palm plants and opens new avenues for further research into the genetics and biology of these important species. After the monocot–eudicot divergence, all palms underwent two WGD events, τ-WGD and p-WGD. WGDs are prevalent in angiosperms and usually promote adaptative evolution in plants (Soltis and Soltis, 2016; Van de Peer et al., 2017). Our comparative genomic analysis suggested that the palm-specific WGD event, known as p-WGD, significantly contributed to the evolution and adaptation of palm plants. While GO terms related to biological regulation, transcription regulator activity, and kinase activity were significantly

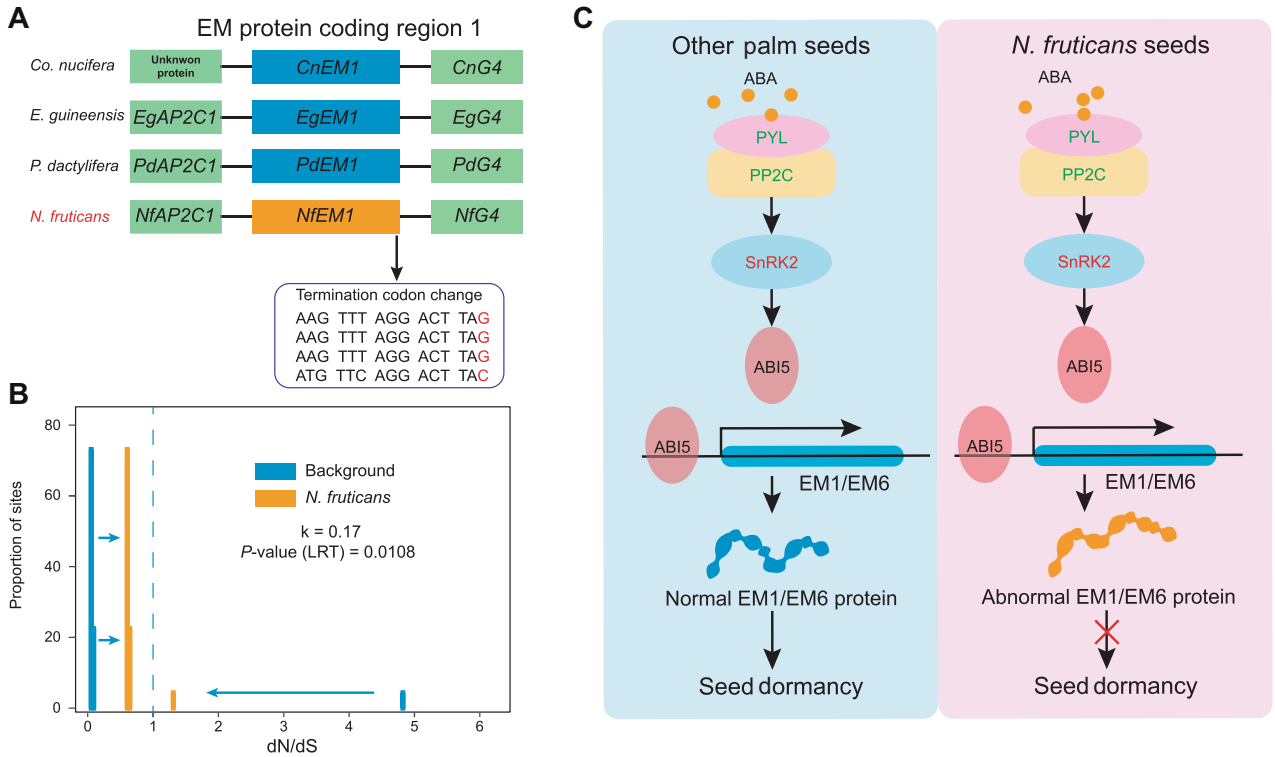

Figure 5. The pseudogenization of *EM1* and *EM6* could underlie *Nypa fruticans* crypto-vivipary

(A) The termination codon of *EM* protein-coding Region 1 (*NfEM1*) has been changed from TAG to TAC in *N. fruticans*. **(B)** Distribution of *dN*/*dS* values across sites in *NfEM1* (orange) and its orthologs (blue). The blue dotted line represents the neutral value (*dN*/*dS* = 1). The arrows represent the direction of change in *dN*/*dS* between *NfEM1* (orange) and its orthologs in other palms. *P*-values were calculated using the likelihood ratio test (LRT). And $k < 1$ ($k = 0.17$) indicates the relaxation of selection in *NfEM1*. **(C)** Proposed model for the control of seed dormancy induced by *EM1* and *EM6* in *N. fruticans* and other palms. Left, regulation of seed dormancy in a non-viviparous seed, such as in *Phoenix dactylifera*. Right, regulation of seed dormancy in *N. fruticans*. The pseudogenization of *EM1* and *EM6* in *N. fruticans* results in an abnormal EM1 and EM6 protein, which might decrease seed dormancy and lead to crypto-vivipary in *N. fruticans*.

enriched among the retained genes from both WGD events, the GO terms related to signal transduction, vesicle-mediated transport, and gravitropism were significantly enriched only among the retentions from the p-WGD event. In signal transduction, the *CIPK* gene family functions in Ca^{2+}-dependent kinase-related signaling pathways and is related to multiple abiotic stress responses (Liu et al., 2000; Kim et al., 2003; Zhang et al., 2020). We found that the p-WGD event promoted the expansion of eight *CIPK* subgroups, including subgroups B–D, known as SALT OVERLY SENSITIVE 2 (*SOS2*) or *SOS2*-like genes related to salt tolerance (Liu et al., 2000), and Group E members, which function in drought tolerance (Kim et al., 2003). Lignin biosynthesis plays a critical role in plant growth, development, lodging resistance, and responses to multiple biotic and abiotic stresses, which is important for the structural integrity of palms and protection from stress. We, therefore, examined the role of p-WGD in the evolution of lignin biosynthesis in palm plants. Palm plants have more copies of *F5H*, encoding a key enzyme in the S-units lignin biosynthesis pathway (Yoon et al., 2015). Two copies of *F5H* were derived from the p-WGD event, and another copy was derived from tandem duplication. We showed here that two of the duplicates derived from p-WGD had undergone positive selection. These findings suggested that the p-WGD event played a significant role in shaping the characteristics of palm plants, particularly their signal transduction pathways, abiotic stress response, and lignin biosynthesis evolution.

Nypa fruticans is the only true mangrove species among all monocotyledonous plants. We noted similarities and distinctions between *N. fruticans* and the eudicot mangroves. In terms of genomic evolution, the TE content of the *N. fruticans* genome is significantly lower than that of its relatives. This pattern is reminiscent of findings in eudicot mangroves, suggesting a potential convergent strategy among mangrove species to adapt to the dynamic and unstable intertidal environment (Lyu et al., 2018). *Nypa fruticans* also showcases both shared and distinct adaptive traits in comparison to eudicot mangroves; for instance, periodic waterlogging due to tidal fluctuations poses a threat to the survival of mangroves. Most eudicot mangrove species have evolved aerial roots to respond to these environmental challenges (Tomlinson, 2016); however, unlike eudicot mangroves, *N. fruticans* developed leaf bases with abundant aerenchyma, supplying oxygen to its underground and underwater tissues (Chomicki et al., 2014). In addition to waterlogging tolerance, the crypto-vivipary of

embryos is another specialized trait found in several mangrove species, including *Nypa, Aegiceras, Avicennia*, and *Aegialitis*, which increases the survival rates of seedlings (Tomlinson, 2016). Notably, the genetic basis for crypto-vivipary may differ between *N. fruticans* and eudicot mangroves due to the unique evolutionary pathways these species have taken.

In this study, we further investigated the genetic evolution of the *N. fruticans* genomic regions underlying its waterlogging tolerance and crypto-vivipary traits. For waterlogging tolerance, we identified a series of specific and preferentially retained genes in *N. fruticans* that had undergone relaxed purifying selection and were evolving faster than others, which could enhance plant waterlogging tolerance and contribute to adaptive evolution. By comparing the transcriptomes of plants subjected to waterlogging to those of the control group, we noted that the neofunctionalization of p-WGD retentions in ethylene signaling (*ERF-VIIs*), anaerobic metabolism (*ADH*), and ROS detoxification (*CAT*) may contribute to the waterlogging stress response in *N. fruticans*. For the crypto-vivipary trait, previous studies have shown that DELAY OF GERMINATION 1 (*DOG1*) was missing or defunct in true viviparous plants and had lost its heme-binding ability in crypto-viviparous plants (Qiao et al., 2020; Feng et al., 2021). In this study, we screened for gene loss or pseudogenization in the *N. fruticans* genome and revealed that some of the 226 genes specifically lost in *N. fruticans* were involved in seed development. Among these genes, we discovered loss-of-function mutations in the *EM1 and EM6* genes in the *N. fruticans* genome. Previous studies have revealed that EM1 and EM6 proteins are crucial seed dormancy regulators that can be activated by ABSCISIC ACID INSENSITIVE 5 (*ABI5*) and control seed dormancy (Gaubier et al., 1993; Finkelstein and Lynch, 2000). The loss of EM1 and EM6 function in *N. fruticans* may result in the precocious germination of the seed and enable crypto-vivipary, thus improving seedling survival in stressful intertidal environments.

In summary, we assembled a high-quality genome of *N. fruticans* by incorporating SMRT long-read sequencing, highly accurate short-read sequencing, and Hi-C technologies. This genome provided valuable insights into the evolutionary history, genome stability, and intertidal adaptation of this species. These findings also can inform further studies on the early evolution and adaptation of palm plants and their contribution to the species diversity of plants.

MATERIALS AND METHODS

Plant materials and genome sequencing

We sampled the *N. fruticans* plant from the Dongzhai Harbor National Nature Reserve nursery in Hainan with permission and extracted genomic DNA and RNA using the modified cetyltrimethylammonium bromide (CTAB) method (Stewart and Via, 1993). We combined Illumina short-read sequencing, SMRT long-read sequencing, and high-throughput chromosome conformation capture (Hi-C) technologies to generate the whole-genome sequence of *N. fruticans*.

For Illumina short-read sequencing, we prepared 150-bp paired-end libraries with 500-bp insert sizes and sequenced them on the Illumina NovaSeq. 6000 platform. It generated 52.03 Gb of short reads.

For SMRT long-read sequencing, we performed them on the PacBio sequel II platform. An SMRT-bell library was constructed from fragment DNA by performing a template library preparation procedure. After binding polymerase to SMRT-bell templates, we sequenced the library on PacBio SMRT cells 8 M. In total, 67.39 Gb of PacBio sequencing reads were yielded after filtering and preprocessing.

Finally, we performed Hi-C sequencing on Hi-C libraries, which were prepared by *N. fruticans* tender leaves and ligation technology. After fixing the leaves with formaldehyde and lysed, we used *Mbo*I to digest the cross-linked DNA. Subsequently, we biotinylated and ligated the restriction fragment ends. The prepared DNA was fragmented to ~400 bp and sequenced on the Illumina NovaSeq. 6000 platform.

Genome assembly and quality assessment

We assembled the *de novo* genome based on PacBio long reads using wtdbg2 (Ruan and Li, 2020). We used Quiver (SMRT Analysis v2.3.0) to polish the assembly (Chin et al., 2013). To improve the accuracy of primary assembly, we applied Pilon to correct the remaining errors using short reads (Walker et al., 2014).

To refine and generate pseudo-chromosome-level genomes, we first used HiC-Pro to evaluate and qualify the Hi-C data (Servant et al., 2015). Hi-C contact maps were generated by Juicer using filtered Hi-C data (Durand et al., 2016b). The residual duplicate contigs were removed and incorrectly connected scaffolds were split by Juicebox (Durand et al., 2016a). Finally, we anchored the scaffolds to the chromosomes using the 3D-DNA pipeline (Dudchenko et al., 2017).

We estimate the genome size based on *k*-mer frequency distribution using Jellyfish (Marcais and Kingsford, 2011) and Genomescope (Vurture et al., 2017) based on the short-read data. The completeness of the *de novo* assembly was evaluated by the Benchmarking Universal Single-Copy Orthologs (BUSCO) version 3.1 (Simao et al., 2015) based on the embryophyta_odb10 database (1,375 BUSCO groups). In addition, the PacBio subreads were mapped back to the assembly using Minimap2 (Li, 2018) and the Illumina short reads were mapped using BWA (Li and Durbin, 2010) to examine the assembly integrity.

Genome annotation

We identified the repetitive sequences of the *N. fruticans* genome by integrating both homology-based and de novo-based methods. First, we annotated the repeats based on the Repbase TE library (RepBase-20181026) (Bao et al., 2015) and Dfam (v3.1) (Hubley et al., 2016) using RepeatMasker (v4.0.9) (Tarailo-Graovac and Chen, 2009). Subsequently, we performed a *de novo*-based method to identify the repeats by

constructing a repeats library of the *N. fruticans* genome using RepeatModeler (v2.0.1) (Flynn et al., 2020) with the "LTRStruct" parameter and identifying the repeats based on this library. Finally, the homology-based gff and *de novo*-based gff were merged into the result of the final repeat.

For the structural annotation of protein-coding genes in the *N. fruticans* genome, we integrate *de novo*-based, homology-based, and RNA-seq-based predictions. For the *de novo*-based method, we performed Augustus (Stanke et al., 2006) and GlimmerHMM (Majoros et al., 2004) based on the repeat-masked genome generated by the above analysis. We used BLAT to compare the protein sequences of related species with the *N. fruticans* genome for the homology-based method (Kent, 2002). Based on the filtered BLAT results, we conducted the homology-based prediction using GeneWise (Birney et al., 2004). For RNA-seq-based prediction, we used the HISAT2 (Kim et al., 2015) and StringTie (Pertea et al., 2015) pipelines to assemble the *N. fruticans* transcripts. We combined the above gene prediction results using EvidenceModeler (Haas et al., 2008) with the weight as follows (GeneWise:10; StringTie:5; AUGUSTUS:1; GlimmerHMM:1). Eventually, we predicted a total of 24,163 protein-coding genes in the *N. fruticans* genome and functionally annotated them by comparing with the non-redundant protein database (NR), GO, and the KEGG. Moreover, we annotated the transcription factors (TFs) and non-coding RNAs in the *N. fruticans* genome by iTAK (Zheng et al., 2016), tRNAscan-SE (Lowe and Eddy, 1997), and INFERNAL (Nawrocki et al., 2009).

Phylogenetic analyses and gene family analyses

We used 12 plants with available genomes in phylogenetic analyses, including *A. trichopoda* (Albert et al., 2013), *A. thaliana* (Lamesch et al., 2012), *S. polyrhiza* (Wang et al., 2014), *A. shenzhenica* (Zhang et al., 2017), *O. sativa* (Ouyang et al., 2007), *M. acuminata* (D'Hont et al., 2012), *C. simplicifolius* (Zhao et al., 2018), *D. jenkinsiana* (Zhao et al., 2018), *N. fruticans, P. dactylifera* (Hazzouri et al., 2019), *E. guineensis* (Singh et al., 2013), and *C. nucifera* (Lantican et al., 2019). We identified 545 single-copy orthologs gene sets using OrthoFinder (v2.3.7) (Emms and Kelly, 2019). To perform multiple sequence alignments and generate the coding sequences, we utilized MAFFT (Katoh and Toh, 2008) and PAL2NAL (Suyama et al., 2006). Subsequently, we employed Gblocks to trim the alignment (Castresana, 2000). We inferred a ML phylogenetic tree based on the above alignment by RAxML-NG (Kozlov et al., 2019) with GTR (General time reversible)+GAMMA+I model. Meanwhile, 1,000 bootstraps were run to calculate the support values. To date the origin of *Nypa* and other palms, we conducted a molecular clock analysis by applying a Bayesian phylogenetic program MCMCTREE in PAML (Yang, 2007). A previous study found that *Nypa* first appeared nearly 75 mya in Egypt (Schrank, 1987). This fossil record and the other two time points were used to calibrate the divergence time estimation. Therefore, we set three times calibrations, including *N. fruticans*–*P. dactylifera* (>75 mya), Poaceae–Arecaceae (110–120 mya), and the time for the root (173–234 mya). We ran the Markov chain Monte Carlo process for 10 million steps with samples drawn every 500 steps after a discarded burn-in of 1 million steps. Based on the cluster size of ortholog genes inferred from OrthoFinder and the ultrametric tree inferred from MCMCTREE, we identified the expansion and contraction of gene families among 12 species by CAFÉ (v4.2.1) (Han et al., 2013). Gene families that have significant expansion or contraction with a P-value < 0.01 were regarded as rapidly evolving families. The heatmap of the waterlogging-response gene numbers with z-score

$$\left(\frac{\text{The gene copy number of each species} - \text{The average gene copy number of all species}}{\text{The standard deviation of gene copy number of six species}}\right),$$

normalization was plotted using the *pheatmap* R package.

Analyses of whole-genome duplication

To determine and date the WGDs in *N. fruticans* and other palm species, we performed three kinds of analysis, including syntenic analysis, K_s distribution analysis, and phylogenetic analysis. For syntenic analysis, we identified and exhibited the syntenic relationship using MCScanX (Wang et al., 2012). For K_s distribution analysis, we calculated the synonymous substitution rates (K_s) between paralogs or orthologs in the syntenic regions by KaKs_Calculator (Wang et al., 2010). Subsequently, we used a fitted mixture model (GMM model in the wgd program) (Zwaenepoel and Van de Peer, 2019) to distinguish paralogs between τ-WGD and p-WGD with 95% probability.

To re-estimate the timing of τ-WGD and p-WGD, we applied the phylogenetic methods mentioned by Clark and Donoghue (Clark and Donoghue, 2017). First, we integrated WGDs generating duplications and their orthologous genes from other species (*P. dactylifera, O. sativa, S. polyrhiza*, and *A. thaliana*) into a gene family. Orthologous genes were identified by reciprocal BLAST best hits. Then, we constructed the phylogenetic trees of those gene families using RAxML-NG and rooted the trees by Notung (Chen et al., 2000). After filtering the gene trees that do not reflect the correct WGD positions or relationships between species, we concatenated gene families into supergene and estimated the WGD event's timing using MCMCTREE with the divergence times of species as calibration.

To reveal the pattern of fractionation and diploidization following p-WGD among palms, we first used the reciprocal BLASTP best-hit method to identify the orthologs among four palms, including *N. fruticans, P. dactylifera, E. guineensis*, and *C. nucifera*. We identified a total of 11,502 one-to-one orthogroups among the four species (Table S20). Based on the GMM analysis results mentioned above, we ascertained

whether their p-WGD paralogs were present in each genome. We further divided the pair of p-WGD duplicates of *N. fruticans* into three clusters (Table S21), including NR (duplicates only retained in *N. fruticans*), PR (duplicates at least retained in one of the other palm species), and AR (duplicates retained in all four palm species). We calculated the selection pressure (K_a/K_s) and genetics using KaKs_Calculator with the YN model and divergence using the Kimura two-parameter method between each pair of p-WGD retention.

Mutation rate and selection pressure estimation

We extracted 4d sites in the above 545 single-copy orthologs gene alignments and constructed the ML phylogenetic tree to compare the mutation rate (mutation rate = branch length/divergent time) among palms and other angiosperms. To compare the mutation rate between *N. fruticans* and other palms, we performed a Tajima's relative rate test (Tajima, 1993) using MEGA-X (Kumar et al., 2018) and set the *O. sativa* as the outgroup species. Tajima's relative rate test was based on the χ^2 test to judge whether the two ingroup species have significantly different mutation rates.

We conducted selection pressure analysis on the 2,680 single-copy orthologs gene sets, retrieving from seven species genomes, including *O. sativa, C. simplicifolius, D. jenkinsiana, N. fruticans, P. dactylifera, E. guineensis*, and *C. nucifera* using OrthoFinder. We used MAFFT to align the gene sets and PAL2NAL to generate the coding sequences. We trimmed the alignments using Gblocks and transformed the alignments into PAML input format. First, we merged the 2680 gene alignments into a supergene alignment. We estimated the ω (K_a/K_s) of the supergene among different palm lineages using CodeML with the free-ratio model. Then, we used the branch-site model (setting model = 2, Nssites = 2) to determine the frequency of positive selection in each palm lineage. We fitted two models (alternative model with fix_omega = 0 and omega = 2; null model with fix_omega = 1 and omega = 1) and compared the two models with a likelihood ratio test (LRT) with one degree of freedom. We corrected the *P*-values by applying the Benjamini–Hochberg (BH) method and defined the genes with a false discovery rate (FDR) below 5% as positively selected genes (PSGs).

Long terminal RTs and recent gene duplication analyses

We identified intact LTR-RTs using LTRharvest (Ellinghaus et al., 2008) and LTRdigest (Steinbiss et al., 2009) workflow. We required the length of LTR-RTs to range from 100 to 3,000 bp and the length between a pair of putative LTRs ranged from 1,000 to 15,000 bp, with a similarity of more than 80%. We annotated and analyzed protein domains by searching against the GyDB 2.0 database (Llorens et al., 2011). An intact LTR-RT should have a complete *Gag-Pol* sequence. Then, we identified the solo-LTRs (similarity and coverage >80%) and classified them into three superfamilies (*Copia, Gypsy*, and unknown) of LTRs using LTR_retriever (Ou and Jiang, 2018). Furthermore, we estimated the timing of LTR-RT insertion based on the nucleic acid diversity between the 5′ and 3′ LTR sequences of intact LTR-RTs. We aligned the sequences of LTR-RTs two terminals and employed the Kimura two-parameter method (Kimura, 1980) to calculate the insertion date with the mutation rate of 1.8×10^{-8} substitutions per site per year ($T = K/2r$, K is genetic distance and r is the mutation rate). The ratio of transition: transversion (T_s:T_v) was also estimated using the Kimura two-parameter model ($T_S = -\frac{1}{2} \times \ln(1 - 2P - Q) + \frac{1}{4} \times \ln(1 - 2Q)$, $T_V = -\frac{1}{2} \times \ln(1 - 2Q)$, P is the frequency of transition and Q is the frequency of transversion).

To estimate the number of recent gene duplication events since the *N. fruticans* diverged from other palm species, we used BLASTP (with an e-value <1E−10) to find paralogs in each palm genome and estimated synonymous substitution rates (K_s) between the paralogs using KaKs_Calculator with the YN model. The peak of K_s distribution of the orthologs between *N. fruticans* and other palms was around 0.3. Thus, we set a threshold ($K_s < 0.3$) to detect the recent gene duplication events. We also estimated the selection pressure (K_a/K_s) between the paralogs using KaKs_Calculator with the YN model.

Gene functional analyses

We applied the "BiNGO" application in Cytoscape (Shannon et al., 2003) to perform GO enrichment analysis for four palm species WGD retentions (*N. fruticans, P. dactylifera, E. guineensis*, and *C. nucifera*), with each species complete gene annotation as a control. Moreover, we performed KEGG enrichment analysis for *N. fruticans*-specific retained genes based on Fisher's exact test with the whole genome as a control. We corrected *P*-values in both GO and KEGG enrichment analyses using the Benjamini–Hochberg (BH) method.

For *CIPK* gene family identification, we used the reported *CIPK* protein-coding sequences in *A. thaliana* and *O. sativa* to blast against palm genomes using BLASTP (with an e-value <1E−10, identity >60%, alignment length >60%). In addition, the candidate *CIPK* sequences should contain the NAF domain (PF03822) and the kinase domain (PF00069). For *ERF-VII*s, we searched for the *ERF-VII*s family among all palm genomes based on the AP2 domain and a conserved motif at the N-terminal (MCGG). We collected genes in the lignin biosynthesis pathway and suberin biosynthesis pathway that summed up based on previous literature (Vishwanath et al., 2015; Liu et al., 2018). We used the protein sequences of the genes in *A. thaliana* and *O. sativa* to blast against other species with the e-value < 1E-10, identity > 50%, and coverage > 50%. We summarized all identified genes in Tables S22–S24. We constructed the ML gene trees of *CIPK*s, *ERF-VII*s, and genes in the lignin biosynthesis pathway using RAxML-NG with the GTR+G model. We identified the PSGs in lignin biosynthesis and waterlogging tolerance pathways based on the branch-site model (Yang and Nielsen, 2002) in the program CodeML as described above.

Waterlogging treatment and transcriptome analysis

Two-year-old *N. fruticans* seedlings, ranging in height from 60 to 80 cm, were collected from the Coastal Garden Nursery

Farm in Qionghai, Hainan. These seedlings were transplanted into plastic containers filled with nutrient-rich soil and grown for 2 weeks to acclimate to the new environments. Before commencing the waterlogging treatment, all seedlings received an equal amount of water to standardize their conditions. Subsequently, at three time points, specifically 12 h (T_2), 24 h (T_3), and 72 h ago (T_4) prior to sample collection, we respectively flooded the seedlings and kept the water surface to 3 cm above the soil surface during the entire duration of the treatment. We collected the leaf and root tissues from four groups (control Group T_1 and the three waterlogging treatment Groups T_2, T_3, and T_4) at the same time, each with three biological replicates. These tissues were promptly frozen in liquid nitrogen and then stored at −80°C. We extracted the total RNA of the samples using the CTAB method (Stewart and Via, 1993). After preparing the RNA-seq libraries, sequencing was performed at the MGISEQ-2000 platform and 154.47 Gb RNA-seq data were yielded.

We used fastp (v.0.23.4) (Chen et al., 2018) to generate clean reads with default parameters and subsequently mapped clean reads to the reference genome using HISAT2 (v.2.2.1) (Kim et al., 2019). The number of RNA-seq reads mapped to each gene was counted by HTSEQ (v.0.12.4) (Anders et al., 2015). We used the R package DESEQ. 2 (Love et al., 2014) to identify the DEGs with the threshold (FDR < 0.05 and $|\text{Fold change}| > 2$). We further performed quantitative PCR with reverse transcription (RT-qPCR) to validate the expression level of key genes related to the waterlogging response. Primers for RT-qPCR were designed by Primer Premier 5 (Table S25). For reverse transcription, 1 μg of RNA was used, and this was achieved using HiScript IV RT SuperMix for qPCR, including the gDNA wiper (Vazyme). The expression of *EIF4a* was employed as the internal control for RT-qPCR. The experiment was carried out with ChamQ Universal SYBR qPCR Master Mix (Vazyme) and StepOne Software v.2.3 (Applied Biosystems) with 40 cycles of amplification. Three biological replicates for each waterlogging treatment group and three technical replicates for each biological replicate were conducted. The expression level of each gene was analyzed using the $2^{-\Delta\Delta C_T}$method (Livak and Schmittgen, 2001).

Identification of pseudogenization related to crypto-vivipary

To screen the genes that were specifically lost or became pseudogenized in *N. fruticans*, we set the *O. sativa* gene annotations as a reference and run BLASTP (e-value <1E−10) to search whether the homologs in *O. sativa* exist in palm genomes. To exclude the annotation defects, we performed BLAT to search for the putative pseudogene regions and manually checked whether the pseudogene regions indeed have the loss-of-function mutations (mutations that cause premature termination codon, frameshift, start codon loss, or large fragment deletion (>50 bp)) (Xu and Guo, 2020). We further mapped more accurate Illumina short reads to the putative pseudogene regions using BWA to exclude the possibility of assembly error. To eliminate potential biases stemming from sampling, sequencing, and assembly, we mapped the resequencing data of 12 individuals to the reference genome, thereby confirming the fixation of the pseudogenization of *EM1* and *EM6* within the *N. fruticans* population.

To detect whether the *EM1* and *EM6* pseudogene in *N. fruticans* suffered relaxation of natural selection, we aligned *NfEM1* to its orthologs and performed evolutionary rate analyses by running RELAX (Wertheim et al., 2015) implemented in HYPHY (Pond et al., 2020). RELAX generated three different *dN*/*dS* rate categories for both the test and background branches. Here, relaxation of selection indicates that small *dN*/*dS* values (<1) increase toward one and large *dN*/*dS* values (>1) decrease toward 1 in the test branch. RELAX also inferred a selection intensity parameter (*k*) to assess the degree of change in the test branch's selective strength ($k < 1$, relaxed; $k > 1$, intensified). To test whether the changes in selective strength are significant, the RELAX performs an LRT by comparing the alternative model (*k* is the free parameter) to the null model (*k* is constrained to 1).

Data availability statement

The genome assembly sequences have been deposited in the Genome Warehouse (GWH, https://bigd.big.ac.cn/gwh) in the National Genomics Data Center (NGDC) under accession number GWHBCKC00000000.1 with BioProject ID PRJCA022416 and BioSample ID SAMC3279992.

ACKNOWLEDGEMENTS

We thank Li Shichuan and Zhu Tao for their support. This project was supported by the National Natural Science Foundation of China (32170230, 31971540, 31830005, 42276159); the Guangdong Basic and Applied Basic Research Foundation (2023B1515020083); and the Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (311021006).

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

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

Edited by: Ya-Long Guo, Institute of Botany, CAS, China.

Received May 30, 2023; **Accepted** Jan. 28, 2024; **Published** Feb. 19, 2024

OO: OnlineOpen

REFERENCES

Al-Mssallem, I.S., Hu, S.N., Zhang, X.W., Lin, Q., Liu, W.F., Tan, J., Yu, X.G., Liu, J.C., Pan, L.L., Zhang, T.W., et al. (2013). Genome sequence of the date palm *Phoenix dactylifera* L. Nat. Commun. **4:** 2274.

Albert, V.A., Barbazuk, W.B., dePamphilis, C.W., Der, J.P., Leebens-Mack, J., Ma, H., Palmer, J.D., Rounsley, S., Sankoff, D., Schuster, S.C., et al. (2013). The *Amborella* genome and the evolution of flowering plants. Science **342:** 1467.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics **31:** 166–169.

Asmussen, C.B., Dransfeld, J., Deickmann, V., Barfod, A.S., Pintaud, J. C., and Baker, W.J. (2006). A new subfamily classification of the palm family (Arecaceae): Evidence from plastid DNA phylogeny. Bot. J. Linn. Soc. **151:** 15–38.

Bao, W.D., Kojima, K.K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob. DNA **6:** 11.

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and genomewise. Genome Res. **14:** 988–995.

Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A.Y., Lim, Z.W., Bezault, E., et al. (2014). The genomic substrate for adaptive radiation in African cichlid fish. Nature **513:** 375–381.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. **17:** 540–552.

Chen, K., Durand, D., and Farach-Colton, M. (2000). NOTUNG: A program for dating gene duplications and optimizing gene family trees. J. Comput. Biol. **7:** 429–447.

Chen, Q.P., Yang, H., Feng, X., Chen, Q.J., Shi, S.H., Wu, C.I., and He, Z.W. (2022a). Two decades of suspect evidence for adaptive molecular evolution-negative selection confounding positive-selection signals. Natl. Sci. Rev. **9:** nwab217.

Chen, S.F., Zhou, Y.Q., Chen, Y.R., and Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics **34:** 884–890.

Chen, Y., Jin, Y.F., Wang, Y.S., Gao, Y.S., Wang, Q., and You, X. (2022b). Diverse roles of the *CIPK* gene family in transcription regulation and various biotic and abiotic stresses: A literature review and bibliometric study. Front. Genet. **13:** 1041078.

Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods **10:** 563–569.

Chomicki, G., Bidel, L.P.R., Baker, W.J., and Jay-Allemand, C. (2014). Palm snorkelling: Leaf bases as aeration structures in the mangrove palm (*Nypa fruticans*). Bot. J. Linn. Soc. **174:** 257–270.

Clark, J.W., and Donoghue, P.C.J. (2017). Constraining the timing of whole genome duplication in plant evolutionary history. Proc. R. Soc. B-Biol. Sci. **284:** 20170912.

Conant, G.C., and Wolfe, K.H. (2008). Turning a hobby into a job: How duplicated genes find new functions. Nat. Rev. Genet. **9:** 938–950.

Couvreur, T.L.P., Forest, F., and Baker, W.J. (2011). Origin and global diversification patterns of tropical rain forests: Inferences from a complete genus-level phylogeny of palms. BMC Biol. **9:** 1–12.

D'Hont, A., Denoeud, F., Aury, J.M., Baurens, F.C., Carreel, F., Garsmeur, O., Noel, B., Bocs, S., Droc, G., Rouard, M., et al. (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature **488:** 213–217.

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., et al. (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science **356:** 92–95.

Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016a). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. **3:** 99–101.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016b). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. **3:** 95–98.

Edger, P.P., Heidel-Fischer, H.M., Bekaert, M., Rota, J., Gloeckner, G., Platts, A.E., Heckel, D.G., Der, J.P., Wafula, E.K., Tang, M., et al. (2015). The butterfly plant arms-race escalated by gene and genome duplications. Proc. Natl. Acad. Sci. U.S.A. **112:** 8362–8366.

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinform. **9:** 1–14.

El-Soughier, M.I., Mehrotra, R.C., Zhou, Z.Y., and Shi, G.L. (2011). *Nypa* fruits and seeds from the Maastrichtian–Danian sediments of Bir Abu Minqar, South Western Desert, Egypt. Palaeoworld **20:** 75–83.

Emms, D.M., and Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. Genome Biol. **20:** 238.

Feng, X., Li, G.H., Wu, W.H., Lyu, H.M., Wang, J.X., Liu, C., Zhong, C.R., Shi, S.H., and He, Z.W. (2023). Expansion and adaptive evolution of the *WRKY* transcription factor family in *Avicennia* mangrove trees. Mar. Life Sci. Technol. **5:** 155–168.

Feng, X., Li, G.H., Xu, S.H., Wu, W.H., Chen, Q.P.A., Shao, S., Liu, M., Wang, N., Zhong, C.R., He, Z.W., et al. (2021). Genomic insights into molecular adaptation to intertidal environments in the mangrove *Aegiceras corniculatum*. New Phytol. **231:** 2346–2358.

Feng, X., Xu, S.H., Li, J.F., Yang, Y.C., Chen, Q.P., Lyu, H.M., Zhong, C. R., He, Z.W., and Shi, S.H. (2020). Molecular adaptation to salinity fluctuation in tropical intertidal environments of a mangrove tree *Sonneratia alba*. BMC Plant Biol. **20:** 178.

Finkelstein, R.R., and Lynch, T.J. (2000). The *Arabidopsis abscisic* acid response gene *ABI5* encodes a basic leucine zipper transcription factor. Plant Cell **12:** 599–609.

Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. U.S. A. **117:** 9451–9457.

Gaubier, P., Raynal, M., Hull, G., Huestis, G.M., Grellet, F., Arenas, C., Pages, M., and Delseny, M. (1993). Two different Em-like genes are expressed in *Arabidopsis thaliana* seeds during maturation. Mol. Gen. Genet. **238:** 409–418.

Gaut, B.S., Morton, B.R., McCaig, B.C., and Clegg, M.T. (1996). Substitution rate comparisons between grasses and palms: Synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proc. Natl. Acad. Sci. U.S.A. **93:** 10274–10279.

Gee, C.T. (2001). The mangrove palm *Nypa* in the geologic past of the New World. Wetl. Ecol. Manag. **9:** 181–194.

Gibbs, D.J., Lee, S.C., Isa, N.M., Gramuglia, S., Fukao, T., Bassel, G.W., Correia, C.S., Corbineau, F., Theodoulou, F.L., Bailey-Serres, J., et al. (2011). Homeostatic response to hypoxia is regulated by the N-end rule pathway in plants. Nature **479:** 415–U172.

Giri, C., Ochieng, E., Tieszen, L.L., Zhu, Z., Singh, A., Loveland, T., Masek, J., and Duke, N. (2011). Status and distribution of mangrove forests of the world using earth observation satellite data. Glob. Ecol. Biogeogr. **20:** 154–159.

Graeber, K., Nakabayashi, K., Miatton, E., Leubner-Metzger, G., and Soppe, W.J. (2012). Molecular mechanisms of seed dormancy. Plant Cell Environ. **35:** 1769–1786.

Gregor, H.J., and Hagn, H. (1982). Fossil fructifications from the Cretaceous–Palaeocene boundary of SW-Egypt (Danian, Bir Abu Munqar). Tertiary Res. **4:** 121–147.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EvidenceModeler and the program to assemble spliced alignments. Genome Biol. **9:** R7.

Han, M.V., Thomas, G.W., Lugo-Martinez, J., and Hahn, M.W. (2013). Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE3. Mol. Biol. Evol. **30:** 1987–1997.

Harley, M.M. (2006). A summary of fossil records for Arecaceae. Bot. J. Linn. Soc. **151:** 39–67.

Hawkins, J.S., Proulx, S.R., Rapp, R.A., and Wendel, J.F. (2009). Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. Proc. Natl. Acad. Sci. U.S.A. **106:** 17811–17816.

Hazzouri, K.M., Gros-Balthazard, M., Flowers, J.M., Copetti, D., Lemansour, A., Lebrun, M., Masmoudi, K., Ferrand, S., Dhar, M.I., Fresquez, Z.A., et al. (2019). Genome-wide association mapping of date palm fruit traits. Nat. Commun. **10:** 4680.

He, Z.W., Feng, X., Chen, Q.P., Li, L.W., Li, S., Han, K., Guo, Z.X., Wang, J.Y., Liu, M., Shi, C.C., et al. (2022). Evolution of coastal forests based on a full set of mangrove genomes. Nat. Ecol. Evol. **6:** 738–749.

He, Z.W., Li, X.N., Yang, M., Wang, X.F., Zhong, C.R., Duke, N.C., Wu, C.I., and Shi, S.H. (2019). Speciation with gene flow via cycles of isolation and migration: Insights from multiple mangrove taxa. Natl. Sci. Rev. **6:** 275–288.

He, Z.W., Xu, S.H., Zhang, Z., Guo, W.X., Lyu, H.M., Zhong, C.R., Boufford, D.E., Duke, N.C., The International Mangrove Consortium, and Shi, S.H. (2020). Convergent adaptation of the genomes of woody plants at the land-sea interface. Natl. Sci. Rev. **7:** 978–993.

He, Z.W., Zhang, Z., Guo, W.X., Zhang, Y., Zhou, R.C., and Shi, S.H. (2015). 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.

Hinz, M., Wilson, I.W., Yang, J., Buerstenbinder, K., Llewellyn, D., Dennis, E.S., Sauter, M., and Dolferus, R. (2010). *Arabidopsis RAP2.2*: An ethylene response transcription factor that is important for hypoxia survival. Plant Physiol. **153:** 757–772.

Hu, M.J., Sun, W.H., Tsai, W.C., Xiang, S., Lai, X.K., Chen, D.Q., Liu, X. D., Wang, Y.F., Le, Y.X., Chen, S.M., et al. (2020). Chromosome-scale assembly of the *Kandelia obovata* genome. Hortic. Res. **7:** 75.

Hubley, R., Finn, R.D., Clements, J., Eddy, S.R., Jones, T.A., Bao, W.D., Smit, A.F.A., and Wheelers, T.J. (2016). The Dfam database of repetitive DNA families. Nucleic Acids Res. **44:** D81–D89.

Janssen, T., and Bremer, K. (2004). The age of major monocot groups inferred from 800+ *rbcL* sequences. Bot. J. Linn. Soc. **146:** 385–398.

Jiao, Y.N., Li, J.P., Tang, H.B., and Paterson, A.H. (2014). Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. Plant Cell **26:** 2792–2802.

Katoh, K., and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. **9:** 286–298.

Kent, W.J. (2002). BLAT - The BLAST-like alignment tool. Genome Res. **12:** 656–664.

Kim, D., Landmead, B., and Salzberg, S.L. (2015). HISAT: A fast spliced aligner with low memory requirements. Nat. Methods **12:** 357–U121.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. **37:** 907–915.

Kim, K.N., Cheong, Y.H., Grant, J.J., Pandey, G.K., and Luan, S. (2003). *CIPK3*, a calcium sensor-associated protein kinase that regulates abscisic acid and cold signal transduction in *Arabidopsis*. Plant Cell **15:** 411–423.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111–120.

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics **35:** 4453–4455.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. **35:** 1547–1549.

Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., et al. (2012). The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. Nucleic Acids Res. **40:** D1202–D1210.

Lantican, D.V., Strickler, S.R., Canama, A.O., Gardoce, R.R., Mueller, L.A., and Galvez, H.F. (2019). De novo genome sequence assembly of dwarf coconut (*Cocos nucifera* L. "catigan green dwarf") provides insights into genomic variation between coconut types and related palm species. G3-Genes Genom. Genet. **8:** 2377–2393.

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics **34:** 3094–3100.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics **26:** 589–595.

Liu, J.P., Ishitani, M., Halfter, U., Kim, C.S., and Zhu, J.K. (2000). The *Arabidopsis thaliana SOS2* gene encodes a protein kinase that is required for salt tolerance. Proc. Natl. Acad. Sci. U.S.A. **97:** 3730–3734.

Liu, Q.Q., Luo, L., and Zheng, L.Q. (2018). Lignins: biosynthesis and biological functions in plants. Int. J. Mol. Sci. **19:** 355.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods **25:** 402–408.

Llorens, C., Futami, R., Covelli, L., Dominguez-Escriba, L., Viu, J.M., Tamarit, D., Aguilar-Rodriguez, J., Vicente-Ripolles, M., Fuster, G., Bernet, G.P., et al. (2011). The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. Nucleic Acids Res. **39:** D70–D74.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq. 2. Genome Biol. **15:** 550.

Lowe, T.M., and Eddy, S.R. (1997). TRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25:** 955–964.

Luan, S. (2009). The *CBL-CIPK* network in plant calcium signaling. Trends Plant Sci. **14:** 37–42.

Lynch, M., and Conery, J.S. (2000). The evolutionary fate and consequences of duplicate genes. Science **290:** 1151–1155.

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

Majoros, W.H., Pertea, M., and Salzberg, S.L. (2004). TigrScan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. Bioinformatics **20:** 2878–2879.

Manfre, A.J., LaHatte, G.A., Climer, C.R., Marcotte Jr., W.R. (2009). Seed dehydration and the establishment of desiccation tolerance during seed maturation is altered in the *Arabidopsis thaliana* mutant *atem6-1*. Plant Cell Physiol. **50:** 243–253.

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

Ming, R., VanBuren, R., Wai, C.M., Tang, H.B., Schatz, M.C., Bowers, J. E., Lyons, E., Wang, M.L., Chen, J., Biggers, E., et al. (2015). The pineapple genome and the evolution of CAM photosynthesis. Nat. Genet. **47:** 1435–1442.

Miryeganeh, M., Marletaz, F., Gavriouchkina, D., and Saze, H. (2021). De novo genome assembly and in natura epigenomics reveal salinity-induced DNA methylation in the mangrove tree *Bruguiera gymnorhiza*. New Phytol. **233:** 2094–2110.

Muller, J. (1968). Palynology of the Pedawan and Plateau sandstone formations (Cretaceous – Eocene) in Sarawak, Malaysia. Micropaleontology **14:** 1–37.

Nawrocki, E.P., Kolbe, D.L., and Eddy, S.R. (2009). Infernal 1.0: Inference of RNA alignments. Bioinformatics **25:** 1335–1337.

Nevado, B., Atchison, G.W., Hughes, C.E., and Filatov, D.A. (2016). Widespread adaptive evolution during repeated evolutionary radiations in New World lupins. Nat. Commun. **7:** 12384.

Nevado, B., Wong, E.L.Y., Osborne, O.G., and Filatov, D.A. (2019). Adaptive evolution is common in rapid evolutionary radiations. Curr. Biol. **29:** 3081–3086.e5.

Ou, S.J., and Jiang, N. (2018). LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiol. **176:** 1410–1422.

Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R.L., Lee, Y., Zheng, L., et al. (2007). The TIGR rice genome annotation resource: Improvements and new features. Nucleic Acids Res. **35:** D883–D887.

Van de Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. Nat. Rev. Genet. **18:** 411–424.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. **33:** 290–295.

Plaziat, J.C., Cavagnetto, C., Koeniguer, J.C., and Baltzer, F. (2001). History and biogeography of the mangrove ecosystem, based on a critical reassessment of the paleontological record. Wetl. Ecol. Manag. **9:** 161–179.

Pond, S.L.K., Poon, A.F.Y., Velazquez, R., Weaver, S., Hepler, N.L., Murrell, B., Shank, S.D., Magalis, B.R., Bouvier, D., Nekrutenko, A., et al. (2020). HyPhy 2.5-a customizable platform for evolutionary hypothesis testing using phylogenies. Mol. Biol. Evol. **37:** 295–299.

Qiao, H.M., Zhou, X.X., Su, W.Y., Zhao, X., Jin, P.F., He, S.S., Hu, W., Fu, M.P., Yu, D.T., Hao, S.Q., et al. (2020). The genomic and transcriptomic foundations of viviparous seed development in mangroves. bioRxiv, https://doi.org/10.1101/2020.10.19.346163

Robertson, F.M., Gundappa, M.K., Grammes, F., Hvidsten, T.R., Redmond, A.K., Lien, S., Martin, S.A.M., Holland, P.W.H., Sandve, S.R., and Macqueen, D.J. (2017). Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. Genome Biol. **18:** 111.

Ruan, J., and Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. Nat. Methods **17:** 155–158.

Schrank, E. (1987). Palaeozoic and Mesozoic palynomorphs from northeast Africa (Egypt and Sudan) with special reference to Late Cretaceous pollen and dinoflagellates. Berl. Geowiss. Abh. Reihe A. Geol Palaeontol. **75:** 249–310.

Serrato-Capuchina, A., and Matute, D.R. (2018). The role of transposable elements in speciation. Genes **9:** 254.

Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J. P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. Genome Biol. **16:** 259.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. **13:** 2498–2504.

Shi, T., Huneau, C., Zhang, Y., Li, Y., Chen, J.M., Salse, J., and Wang, Q.F. (2022). The slow-evolving *Acorus tatarinowii* genome sheds light on ancestral monocot evolution. Nat. Plants **8:** 764–777.

Silvestro, D., Bacon, C.D., Ding, W.N., Zhang, Q.Y., Donoghue, P.C.J., Antonelli, A., and Xing, Y.W. (2021). Fossil data support a pre-Cretaceous origin of flowering plants. Nat. Ecol. Evol. **5:** 449–457.

Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31:** 3210–3212.

Singh, R., Ong-Abdullah, M., Low, E.T., Manaf, M.A., Rosli, R., Nookiah, R., Ooi, L.C., Ooi, S.E., Chan, K.L., Halim, M.A., et al. (2013). Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. Nature **500:** 335–339.

Soltis, P.S., and Soltis, D.E. (2016). Ancient WGD events as drivers of key innovations in angiosperms. Curr. Opin. Plant Biol. **30:** 159–165.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: Ab initio prediction of alternative transcripts. Nucleic Acids Res. **34:** W435–W439.

Steinbiss, S., Willhoeft, U., Gremme, G., and Kurtz, S. (2009). Fine-grained annotation and classification of de novo predicted LTR retrotransposons. Nucleic Acids Res. **37:** 7002–7013.

Stewart Jr., C.N., and Via, L.E. (1993). A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. Biotechniques **14:** 748–750.

Stewart, J.J., Akiyama, T., Chapple, C., Ralph, J., and Mansfield, S.D. (2009). The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid poplar. Plant Physiol. **150:** 621–635.

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. **34:** W609–W612.

Tajima, F. (1993). Simple methods for testing the molecular evolutionary clock hypothesis. Genetics **135:** 599–607.

Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinf. **25:** 4.10.1–4.10.14.

Tomlinson, P.B. (2006). The uniqueness of palms. Bot. J. Linn. Soc. **151:** 5–14.

Tomlinson, P.B. (2016). *The Botany of Mangroves*. 2nd ed. (Cambridge, UK: Cambridge University Press).

Tuan, P.A., Kumar, R., Rehal, P.K., Toora, P.K., and Ayele, B.T. (2018). Molecular mechanisms underlying abscisic acid/gibberellin balance in the control of seed dormancy and germination in cereals. Front. Plant Sci. **9:** 668.

Vishwanath, S.J., Delude, C., Domergue, F., and Rowland, O. (2015). Suberin: Biosynthesis, regulation, and polymer assembly of a protective extracellular barrier. Plant Cell Rep. **34:** 573–586.

Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., and Schatz, M.C. (2017). GenomeScope: Fast reference-free genome profiling from short reads. Bioinformatics **33:** 2202–2204.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q.D., Wortman, J., Young, S.K., et al. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE **9:** e112963.

Wang, D.P., Zhang, Y.B., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: A toolkit incorporating gamma-Series methods and sliding window strategies. Genom. Proteom. Bioinform. **8:** 77–80.

Wang, S.C., Xiao, Y., Zhou, Z.W., Yuan, J.Q., Guo, H., Yang, Z., Yang, J., Sun, P.C., Sun, L.S., Deng, Y., et al. (2021). High-quality reference

genome sequences of two coconut cultivars provide insights into evolution of monocot chromosomes and differentiation of fiber content and plant height. Genome Biol. **22:** 304.

Wang, W., Haberer, G., Gundlach, H., Glasser, C., Nussbaumer, T., Luo, M.C., Lomsadze, A., Borodovsky, M., Kerstetter, R.A., Shanklin, J., et al. (2014). The *Spirodela polyrhiza* genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. Nat. Commun. **5:** 3311.

Wang, Y.P., Tang, H.B., DeBarry, J.D., Tan, X., Li, J.P., Wang, X.Y., Lee, T.H., Jin, H.Z., Marler, B., Guo, H., et al. (2012). MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. **40:** e49.

Watanabe, K., Nishiuchi, S., Kulichikhin, K., and Nakazono, M. (2013). Does suberin accumulation in plant roots contribute to waterlogging tolerance? Front. Plant Sci. **4:** 178.

Wertheim, J.O., Murrell, B., Smith, M.D., Pond, S.L.K., and Scheffler, K. (2015). RELAX: Detecting relaxed selection in a phylogenetic framework. Mol. Biol. Evol. **32:** 820–832.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., et al. (2007). A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. **8:** 973–982.

Xu, K., Xu, X., Fukao, T., Canlas, P., Maghirang-Rodriguez, R., Heuer, S., Ismail, A.M., Bailey-Serres, J., Ronald, P.C., and Mackill, D.J. (2006). *Sub1A* is an ethylene-response-factor-like gene that confers submergence tolerance to rice. Nature **442:** 705–708.

Xu, S.H., He, Z.W., Guo, Z.X., Zhang, Z., Wyckoff, G.J., Greenberg, A., Wu, C.I., and Shi, S.H. (2017a). Genome-wide convergence during evolution of mangroves from woody plants. Mol. Biol. Evol. **34:** 1008–1015.

Xu, S.H., He, Z.W., Zhang, Z., Guo, Z.X., Guo, W.X., Lyu, H.M., Li, J.F., Yang, M., Du, Z.L., Huang, Y.L., et al. (2017b). The origin, diversification and adaptation of a major mangrove clade (Rhizophoreae) revealed by whole-genome sequencing. Natl. Sci. Rev. **4:** 721–734.

Xu, Y.C., and Guo, Y.L. (2020). Less is more, natural loss-of-function mutation is a strategy for adaptation. Plant Commun **1:** 100103.

Yang, Z.H. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24:** 1586–1591.

Yang, Z.H., and Nielsen, R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol. Biol. Evol. **19:** 908–917.

Yao, G., Zhang, Y.Q., Barrett, C., Xue, B., Bellot, S., Baker, W.J., and Ge, X.J. (2023). A plastid phylogenomic framework for the palm family (Arecaceae). BMC Biol. **21:** 50.

Yoon, J., Choi, H., and An, G. (2015). Roles of lignin biosynthesis and regulatory genes in plant development. J. Integr. Plant Biol. **57:** 902–912.

Zhang, G.Q., Liu, K.W., Li, Z., Lohaus, R., Hsiao, Y.Y., Niu, S.C., Wang, J.Y., Lin, Y.C., Xu, Q., Chen, L.J., et al. (2017). The *Apostasia* genome and the evolution of orchids. Nature **549:** 379–383.

Zhang, X.X., Li, X.X., Zhao, R., Zhou, Y., and Jiao, Y.N. (2020). Evolutionary strategies drive a balance of the interacting gene products for the CBL and CIPK gene families. New Phytol. **226:** 1506–1516.

Zhao, H.S., Wang, S.B., Wang, J.L., Chen, C.H., Hao, S.J., Chen, L.F., Fei, B.H., Han, K., Li, R.S., Shi, C.C., et al. (2018). The chromosome-level genome assemblies of two rattans (*Calamus simplicifolius* and *Daemonorops jenkinsiana*). Gigascience **7:** giy097.

Zheng, Y., Jiao, C., Sun, H.H., Rosli, H.G., Pombo, M.A., Zhang, P.F., Banf, M., Dai, X.B., Martin, G.B., Giovannoni, J.J., et al. (2016). iTAK: A program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. Mol. Plant **9:** 1667–1670.

Zhu, R.R., Shao, S., Xie, W., Guo, Z.X., He, Z.W., Li, Y.L., Wang, W.Q., Zhong, C.R., Shi, S.H., and Xu, S.H. (2023). High-quality genome of a pioneer mangrove *Laguncularia racemosa* explains its advantages for intertidal zone reforestation. Mol Ecol Resour. In press. https://doi.org/10.1111/1755-0998.13863

Zwaenepoel, A., and Van de Peer, Y. (2019). Wgd-simple command line tools for the analysis of ancient whole-genome duplications. Bioinformatics **35:** 2153–2155.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article: http://onlinelibrary.wiley.com/doi/10.1111/jipb.13625/suppinfo

Figure S1. Genome size estimation of *Nypa fruticans* based on *k*-mer distribution analysis
Figure S2. Hi-C contact map of the *N. fruticans* genome assembly
Figure S3. Features of the *N. fruticans* genome
Figure S4. Maximum-likelihood tree of 12 angiosperm species
Figure S5. The positions of two whole-genome duplications (WGDs) (τ-WGD and p-WGD)
Figure S6. Syntenic blocks between genomic regions in *Nypa fruticans* and *Ananas comosus*
Figure S7. The K_s distributions for paralogs in four palm species
Figure S8. The stability of *N. fruticans* over long-term evolution
Figure S9. Maximum-likelihood (ML) phylogenetic tree inferred using fourfold degenerate (4d) sites
Figure S10. The bar plot shows the frequency of genes evolved under positive selection
Figure S11. The statistical information of long terminal repeat retrotransposons (LTR-RTs)
Figure S12. A Venn diagram of the shared and unique gene families among the p-WGD retained genes in four palm species
Figure S13. The selective pressure (K_a/K_s) and the sequence divergence (*d*) between the pairs of p-WGD-derived duplicates
Figure S14. The synteny and phylogeny analyses of *ERF-VIIs* in the *N. fruticans* genome
Figure S15. Overview of the suberin biosynthetic pathway
Figure S16. Loss-of-function mutations in three *EM* putative pseudogene regions in *N. fruticans* genome
Figure S17. The natural distribution of *N. fruticans*
Figure S18. The morphological structure of *N. fruticans*
Table S1. Statistics for the *N. fruticans* genome assembly
Table S2. Genome assembly and annotation completeness
Table S3. Functional annotation of predicted genes in the *N. fruticans* genome
Table S4. Transcription factor (TF) identification (top 15 TF families in *N. fruticans* and total numbers of TF)
Table S5. Non-coding RNA identification
Table S6. The Gene Ontology (GO) function enrichment of the *N. fruticans*-specific gene families
Table S7. The Gene Ontology (GO) function enrichment of the *N. fruticans* expanded gene families
Table S8. Results of Tajima's relative rate test
Table S9. Prediction of repetitive sequences in the six palm genomes using RepeatModeler and RepeatMasker
Table S10. The Gene Ontology (GO) function enrichment of the two whole-genome duplication (WGD) events in four palm species
Table S11. Function annotation of gene families retained from the p-WGD involved in the "signal transduction" term
Table S12. The numbers of genes involved in the lignin biosynthesis pathway among palms and five other angiosperm species
Table S13. KEGG enrichment of *N. fruticans*-specific p-WGD-derived gene retentions
Table S14. *N. fruticans*- specific p-WGD-derived gene retentions involved in the waterlogging tolerance pathway
Table S15. Positively selected genes related to waterlogging tolerance in *N. fruticans*
Table S16. The numbers of differentially expressed genes in leaf and root
Table S17. Specifically lost genes in *N. fruticans*
Table S18. Pseudogene and gene loss in *N. fruticans* genome related to seed development.

Table S19. Earlier studies regarding the age of *Nypa* and palm species
Table S20. Identified syntenic ortholog-paralog gene clusters between four palm species
Table S21. Gene list of three groups of p-WGD duplicate pairs in *N. fruticans*
Table S22. The summary of *CIPK* genes identification
Table S23. The summary of lignin biosynthesis pathway genes identification
Table S24. The summary of *ERF-VII* and suberin biosynthesis pathway genes identification
Table S25. Primer sequences for real-time quantitative PCR analysis

Scan using WeChat with your smartphone to view JIPB online

Scan with iPhone or iPad to view JIPB on Twitter