





# Genomic Insights into Adaptation to Karst Limestone and Incipient Speciation in East Asian *Platycarya* spp. (Juglandaceae)

Yu Cao <sup>1,2,3</sup> Fabricio Almeida-Silva,<sup>2,3</sup> Wei-Ping Zhang <sup>1</sup> Ya-Mei Ding<sup>1</sup> Dan Bai,<sup>1</sup> Wei-Ning Bai,<sup>\*1</sup> Bo-Wen Zhang,<sup>\*1</sup> Yves Van de Peer <sup>\*,2,3,4,5</sup> and Da-Yong Zhang <sup>\*,1</sup>

<sup>1</sup>State Key Laboratory of Earth Surface Process and Resource Ecology and Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing, China

<sup>2</sup>Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium

<sup>3</sup>Center for Plant Systems Biology, VIB, Ghent, Belgium

<sup>4</sup>Center for Microbial Ecology and Genomics, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa

<sup>5</sup>College of Horticulture, Nanjing Agricultural University, Nanjing, China

\*Corresponding authors: E-mails: baiwn@bnu.edu.cn; zhangbw@mail.bnu.edu.cn; yves.vandeppeer@psb.ugent.be; zhangdy@bnu.edu.cn.

Associate editor: Michael Purugganan

## Abstract

When challenged by similar environmental conditions, phylogenetically distant taxa often independently evolve similar traits (convergent evolution). Meanwhile, adaptation to extreme habitats might lead to divergence between taxa that are otherwise closely related. These processes have long existed in the conceptual sphere, yet molecular evidence, especially for woody perennials, is scarce. The karst endemic *Platycarya longipes* and its only congeneric species, *Platycarya strobilacea*, which is widely distributed in the mountains in East Asia, provide an ideal model for examining the molecular basis of both convergent evolution and speciation. Using chromosome-level genome assemblies of both species, and whole-genome resequencing data from 207 individuals spanning their entire distribution range, we demonstrate that *P. longipes* and *P. strobilacea* form two species-specific clades, which diverged around 2.09 million years ago. We find an excess of genomic regions exhibiting extreme interspecific differentiation, potentially due to long-term selection in *P. longipes*, likely contributing to the incipient speciation of the genus *Platycarya*. Interestingly, our results unveil underlying karst adaptation in both copies of the calcium influx channel gene *TPC1* in *P. longipes*. *TPC1* has previously been identified as a selective target in certain karst-endemic herbs, indicating a convergent adaptation to high calcium stress among karst-endemic species. Our study reveals the genic convergence of *TPC1* among karst endemics and the driving forces underneath the incipient speciation of the two *Platycarya* lineages.

**Key words:** comparative genomics, ecological speciation, gene flow, molecular convergence, positive selection, population genetics, walnut family, *TPC1*.

## Introduction

Convergent evolution, in which distinct lineages independently evolve similar traits, is prevalent in both animals and plants and has fascinated evolutionary biologists for centuries (Darwin 2004). Some of the most well-known examples of convergent evolution include high-altitude adaptation in humans and domestic animals in the Tibetan plateau (Gou et al. 2014; Wu et al. 2020) and echolocation in bats and toothed whales (Li et al. 2010; Liu et al. 2010; Parker et al. 2013). Furthermore, many extremophile plants have been found to exhibit similar traits and adaptive strategies in response to intense selective pressures from a similar or shared habitat, such as low temperature

(Yeaman et al. 2016), low nutrients (Fukushima et al. 2017), high salinity (Xu et al. 2017; Lyu et al. 2018; He et al. 2020), and heavy metal toxicity (Preite et al. 2019). Whereas at the same time, adaptation to extreme habitats also frequently leads to the divergence among closely related lineages, which is described as a “relay race” that begins with physiological and epigenetic adaptations and concludes with slower, longer-lasting genetic adaptations (Yona et al. 2015). The accumulation of sufficient genetic variants eventually triggers the initiation of speciation (Kautt et al. 2020; Todesco et al. 2020; Nosil et al. 2021). Despite this understanding, molecular evidence for such, on the one hand, convergence in phenotype, and on the other hand, incipient speciation, remains elusive at the

© The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

Open Access

gene, pathway, or genetic level, especially in woody plants (Sackton and Clark 2019; Xu et al. 2020).

The karst habitats, known for their elevated levels of calcium (Ca) and magnesium (Mg), increased pH value, and reduced water storage capacity relative to non-karst soils (Nie et al. 2010; Hao et al. 2014; Geekiyanage et al. 2019) has long been recognized as “natural laboratories” for exploring adaptative evolution and speciation (Clements et al. 2006; Oliver et al. 2017). Plants living in karst habitats are characterized as high Ca and drought tolerant. Advances in comparative genomics and population genetics have revealed evidence of adaptive evolution in some karst plant taxa. For instance, Tao et al. (2016) demonstrated that *TPC1*, a major reactive oxygen species (ROS)-responsive  $\text{Ca}^{2+}$  channel, is potentially involved in the local adaptation to karst  $\text{Ca}^{2+}$ -rich environments of *Primulina* (Gesneriaceae). In the cave plant, *Primulina huaijiensis*, Feng et al. (2020) pointed out that the rapid expansion of the *WRKY* gene family facilitated adaptation to limestone karst. Other studies have also unveiled adaptive signals in genes from karst-inhabiting herbaceous taxa, such as *Marsdenia tenacissima* (Zhou et al. 2023), *Urophysa* (Xie et al. 2021), and *Lonicera confusa* (Jin et al. 2018). However, as the dominant and constructive species of karst forest (Geekiyanage et al. 2019), woody plants are poorly studied. Given that woody plants have longer generation times and generally smaller effective population sizes than herbs, thus requiring more time to amass significant genetic adaptations, it would be instructive to investigate how woody plants gradually adapt to harsh karst environments.

*Platycarya longipes* (Juglandaceae), a common dominant species in karst forests, is confined to limestone substrates and hence considered to be Ca tolerant (Kuang and Lu 1979; Zhang et al. 2010). In contrast, its only congeneric species, *Platycarya strobilacea*, has a widespread distribution in the sunny mountain regions of East Asia with typical soil types (Manos and Stone 2001; Fukuhara and Tokumar 2014; Kozłowski et al. 2018). Some controversy over the taxonomic status of the *Platycarya* species persists, with some studies supporting the “two species” scenario based on the leaf, fruit, pollen, and leaf epidermis morphology, cytotaxonomy, and molecular systematics (fig. 1a) (Luo 2015; Wan et al. 2017), whereas others treating *P. longipes* as an ecological variant of *P. strobilacea* (Lu et al. 1999; Fang et al. 2011; Chen et al. 2012). Given their distinct differences in morphology and habitat, *P. longipes* and *P. strobilacea* represent a promising system for studying karst adaptation, convergent evolution, and speciation.

In this study, we aimed to address the following questions pertaining to the two *Platycarya* lineages: 1) What mechanisms enabled the adaptation of the woody tree species, *P. longipes*, to the challenging karst limestone environment? 2) Does evidence for convergent evolution exist between woody trees and herbs residing in karst habitats? 3) To what extent has such adaptation driven the differentiation and speciation of the two *Platycarya* lineages? To answer these questions, we assembled

high-quality chromosome-level reference genomes of both *P. longipes* and *P. strobilacea* and resequenced genomes of 207 individuals covering the entire distribution ranges of the two species (fig. 1b). Additionally, we sequenced the transcriptomes of 36 experimental seedlings of both species from four-time points under high Ca–Mg treatment and 36 *P. longipes* and 29 *P. strobilacea* wild individuals (5–10 years old) from three sympatric field populations. Our analyses detected signals of convergent evolution for the *TPC1* gene in the Ca-tolerant *P. longipes* and revealed, for the first time, the genomic changes and transcriptomic signatures for karst adaptation of woody trees.

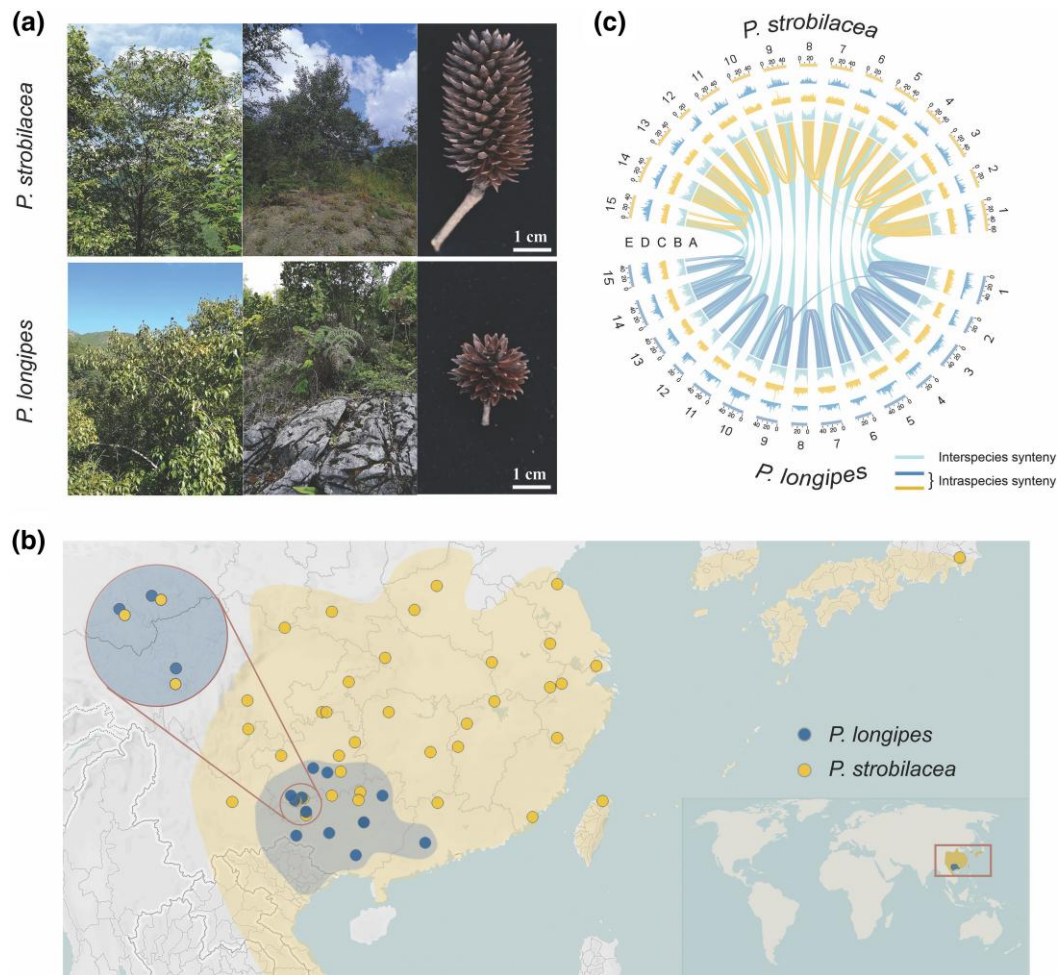
## Results

### Genomic Profiles of *P. longipes* and *P. strobilacea*

We have generated high-quality, chromosome-level reference genomes for the species *P. longipes* and *P. strobilacea*, with estimated genome size of 695.798 Mb and 703.509 Mb, respectively (figs. 1c and S1 and notes, Supplementary Material online). Using fluorescence in situ hybridization (FISH)-based karyotype analysis and flow cytometry, we grouped the contigs of both species into 15 pseudochromosomes (supplementary fig. S1, Supplementary Material online), resulting in ~99.94% and ~97.38% of the assembled sequences being properly anchored for *P. longipes* and *P. strobilacea*, respectively (supplementary table S1, Supplementary Material online). The predicted number of protein-coding genes was 29,525 and 29,330 for *P. longipes* and *P. strobilacea*, respectively (supplementary table S2, Supplementary Material online), in addition to repeats (supplementary table S3, Supplementary Material online) and noncoding RNAs (supplementary table S4, Supplementary Material online). Approximately 94% of 1,440 benchmarking universal single-copy (BUSCO) genes were recovered full-length in both genomes (supplementary table S1, Supplementary Material online). Additionally, 93.04% of the genes in *P. longipes* and 99.03% in *P. strobilacea* could be assigned to entries in six functional databases (supplementary table S5, Supplementary Material online). Further details regarding the genome assembly and annotation can be found in supplementary note 1, Supplementary Material online. Interspecies synteny analysis showed that the first half of chromosome 14 in *P. longipes* and *P. strobilacea* corresponds to chromosome 14 in *Juglans regia* (Zhu et al. 2019), whereas the second half corresponds to chromosome 15 in *J. regia*, indicating a chromosome fusion event occurred in the ancestral species of *Platycarya* through an end–end translocation (Mandakova and Lysak 2018) (supplementary fig. S1, Supplementary Material online).

### Population History Revealed Asymmetric Migration in the Differentiation of the Two Species

The results of the population structure analysis based on 2,182 neutral and independent genome-wide single nucleotide polymorphisms (SNPs) were consistent with



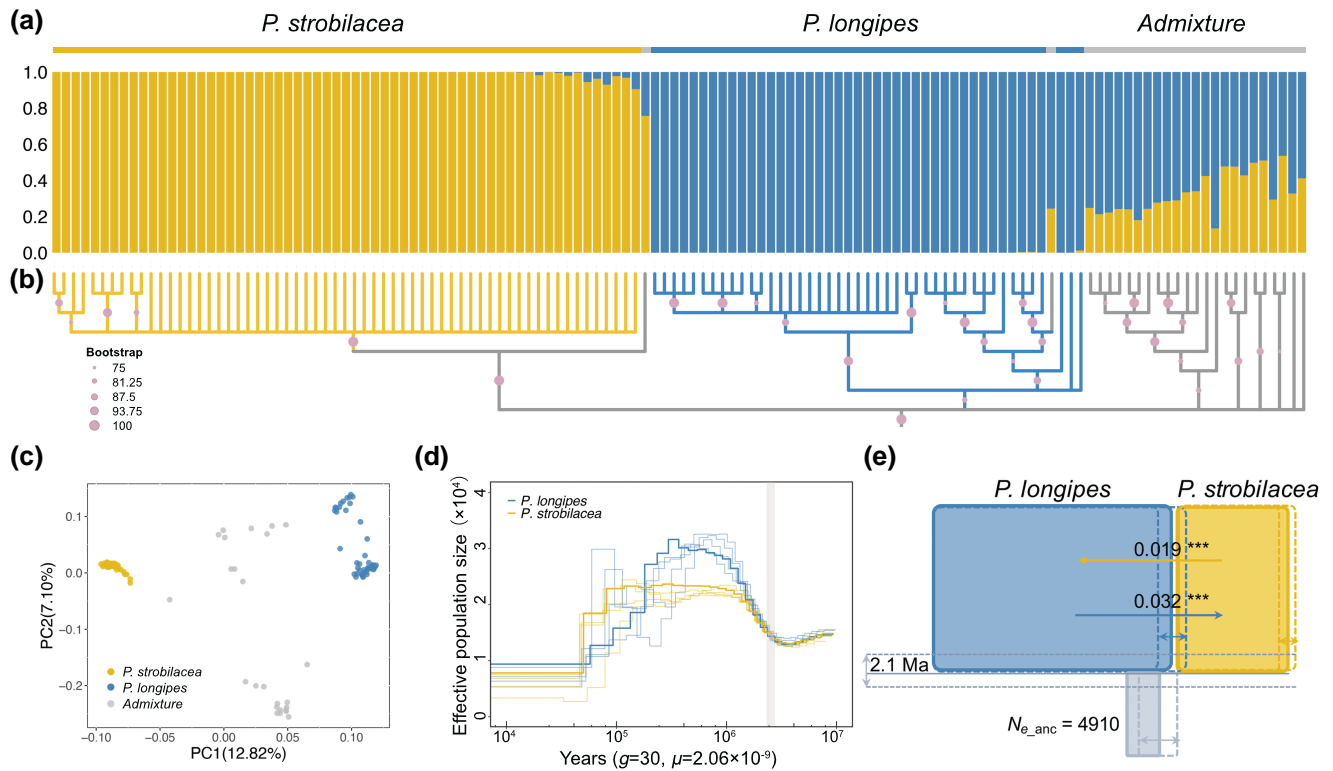
**Fig. 1.** Morphology, habitat, geographic distribution and genome features of *P. longipes* and *P. strobilacea*. (a) Shape of trees, habitat, and shape of fruits of *P. longipes* and *P. strobilacea*. (b) Sample distribution in this study. The circles indicate where the 85 individuals of *P. longipes* and the 122 individuals of *P. strobilacea* were sampled. The enlarged box in the upper left corner represents where the field transcriptome samples were collected. (c) Genome structure of *P. longipes* and *P. strobilacea*. Different tracks (moving outward) denote identified (A) syntenic blocks, either intraspecific or interspecific; (B) gene density in 100 kb sliding windows; (C) copia density in 10 kb sliding windows; (D) gypsy density in 10 kb sliding windows; and (E) chromosomes.

morphological features. The optimal number of genetic clusters ( $K$ ) was two, with all sampled individuals clearly subdivided into two species-specific groups and some individuals exhibiting admixture (fig. 2a). The SNP-based phylogenetic trees also revealed two primary clades, *P. longipes* and *P. strobilacea*, and one admixture clade, rather than a nested topology (fig. 2b). When  $K = 3$ , the admixed individuals were aggregated into a separate cluster (supplementary fig. S2, Supplementary Material online). Furthermore, we found that only 11% of the genome-wide polymorphisms were shared between species, with the remainder ascribed to either *P. strobilacea* (37%) or *P. longipes* (51%) (supplementary fig. S2, Supplementary Material online). These results suggest that *P. longipes* and *P. strobilacea* are separate species with some cross-species admixtures, rather than one being ancestral to the other. In addition, principal component analysis (PCA) of the independent SNPs also revealed *P. longipes* and *P. strobilacea* as two distinct groups in the first principal component (PC1), with a small number of admixed individuals in between (fig. 2c).

Followed by a massive drop of explained variation (from 12.82% to 7.10% and 6.91%), the PC2 and PC3 further divided *P. longipes* into different subpopulations (supplementary fig. S2, Supplementary Material online). As for the whole-genome population statistics, the weighted mean of the fixation index ( $F_{ST}$ ) between the two species was  $0.479 \pm 0.161$  (supplementary table S6, Supplementary Material online), and the absolute pairwise nucleotide divergence between species ( $D_{XY}$ ) was  $0.009 \pm 0.003$  (supplementary table S6, Supplementary Material online). In addition, our analysis revealed that *P. longipes* manifests reduced levels of genome-wide heterozygosity per individual, denoted as  $H_{EV}$  and a higher inbreeding coefficient,  $F_{IS}$ , as compared with *P. strobilacea* (supplementary table S6, Supplementary Material online). These observations lend support to the notion that the karst adaptation of *P. longipes* has resulted in a limited and fragmented range for this species.

The pairwise sequentially Markovian coalescent (PSMC) (Li and Durbin 2011) analysis revealed that the





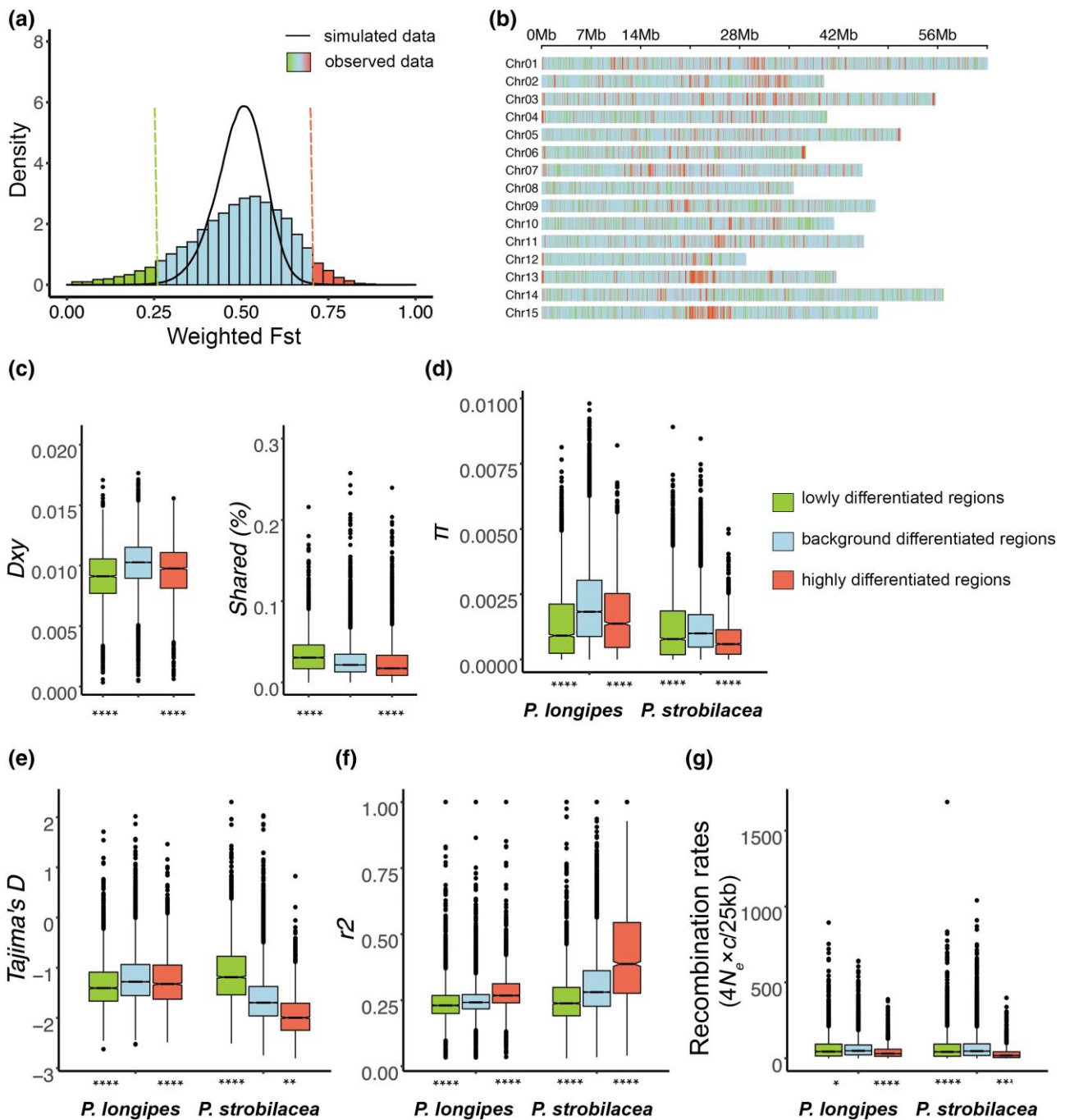
**Fig. 2.** Genetic structure and demographic history of *P. longipes* and *P. strobilacea*. (a) Genetic structure of the two species inferred using STRUCTURE v. 2.3.4 (Pritchard et al. 2000). The y-axis quantifies subgroup membership, and the x-axis shows each individual. (b) The phylogenetic tree depicting 130 unrelated individuals was reconstructed using SVDquartets (Chifman and Kubatko 2014) and rooted by *J. regia*. Clades exhibiting bootstrap values below 75 were collapsed. (c) PCA plot based on genetic covariance among all individuals of *P. longipes* and *P. strobilacea*. The first two PCs are shown. (d) PSMC estimates of the effective population size ( $N_e$ ) changes for *P. longipes* and *P. strobilacea*. The time scale on the x-axis is calculated assuming a neutral mutation rate per year ( $\mu$ ) =  $2.06 \times 10^{-9}$  and generation time ( $g$ ) = 30 years. The vertical bar indicates the divergence time inferred by IMA3. (e) The best-fitting demographic model inferred by IMA3. Boxes represent populations, with widths proportional to estimated effective population sizes (ancestral  $N_e$  is given for scale). Confidence intervals are indicated as dashed-line boxes aligned with the corresponding population's box on the left side. Estimated population migration rates ( $2N_e m$ , indicated by a directional arrow) that are associated with a migration rate significantly  $> 0$  based on a marginal likelihood ratio test (Nielsen and Wakeley 2001) are shown together with their estimated  $2N_e m$  values (\*\* $P < 0.001$ ). The inferred demographic parameters are described in the text and shown in [supplementary table S7, Supplementary Material](#) online.

demographic histories of the two species were very similar until  $\sim 2$  million years ago (Mya), suggesting a possible time of divergence between them. Subsequently, the effective population size ( $N_e$ ) of *P. strobilacea* stabilized at  $\sim 1$  Mya, whereas the  $N_e$  of *P. longipes* continued to increase and reached its peak at  $\sim 0.5$  Mya (fig. 2d). Overall, our analysis reveals a comparable  $N_e$  of 10,000–40,000 for both species. This is supported by similar levels of population genetic diversity ( $\pi$ ) observed for both species (supplementary table S6, Supplementary Material online). To further investigate the population demographic history, we employed the isolation-with-migration (IM) model in IMA3 (Hey et al. 2018), using 200 independent and noncoding loci. IMA3 results indicated that the two species diverged at  $\sim 2.09$  Mya (95% highest posterior density [HPD]: 1.89–2.31 Mya), with asymmetric bidirectional gene flow between the two species (fig. 2e and supplementary table S7, Supplementary Material online). The estimated  $N_e$  of *P. longipes* and *P. strobilacea* were 36,329 (95% HPD: 34,101–38,580) and 16,591 (95% HPD:

15,279–17,887), respectively, both being higher than the  $N_e$  of their common ancestor ( $N_{e\_anc} = 4,923$  [95% HPD: 1,899–7,852]).

### Natural Selection Dominated the Formation of Highly Differentiated Regions

Most nonoverlapping 25-kb windows throughout the genome showed high genetic differentiation between *P. longipes* and *P. strobilacea*, as evidenced by a weighted mean  $F_{ST}$  of  $0.479 \pm 0.161$  (fig. 3a). To access the contribution of neutral historical demography to this pattern, we conducted 500,000 coalescent simulations using ms software (Hudson 2002) based on the isolation with migration (IM) model estimated from IMA3. Comparison of the observed  $F_{ST}$  distribution with the expected null distribution from the simulations showed a flatter distribution with a greater presence of extremely high and low values (fig. 3a). Furthermore, we identified 2,347 and 3,240 outlier windows exhibiting significantly (false discovery rate



**FIG. 3.** Genomic islands of divergence. (a) Distribution of genetic differentiation ( $F_{ST}$ ) between *P. longipes* and *P. strobilacea* from the observed and simulated data sets based on the best-fitting demographic model inferred by *Ima3*. The dashed lines indicate the thresholds for determining significantly ( $FDR < 1\%$ ) high and low interspecific differentiation based on coalescent simulations. (b) Chromosomal distribution of genetic differentiation ( $F_{ST}$ ) between *P. longipes* and *P. strobilacea*. Comparisons of (c)  $D_{xy}$  and the proportion of interspecific shared polymorphisms, (d) nucleotide diversity ( $\pi$ ), (e) Tajima's  $D$ , (f) LD measured in squared correlation coefficients  $r^2$ , and (g) recombination rate, scaled by population sizes in 25 kb size windows, among regions displaying significantly high and low differentiation versus the genomic background. Asterisks designate significant differences between outlier windows and the rest of genomic regions by Mann-Whitney  $U$  test (\* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; \*\*\*\* $P < 0.0001$ ).

[ $FDR < 0.01$ ] high and low interspecific  $F_{ST}$  compared with the expected null distribution (fig. 3a). These windows, with the most extremely high  $F_{ST}$  values, were identified as genomic islands of divergence and were found to be dispersed across the genome with a higher concentration in the middle and towards the end of chromosomes (fig. 3b).

To explore how various evolutionary processes have shaped the patterns of genomic divergence, we further quantified and compared interspecific genomic differentiation between two unions of outlier windows and the rest of the genome. Utilizing absolute pairwise nucleotide divergence between species ( $D_{XY}$ ) and the proportion of

interspecific shared polymorphisms, we also compared the nucleotide diversity ( $\pi$ ), Tajima's  $D$ , recombination rate ( $\rho = 4Nc$ ), and linkage disequilibrium (LD, measured by squared correlation coefficients,  $r^2$ , between pairs of SNPs within each species). The results indicated that compared with the genomic background averages, the  $D_{XY}$  and the proportion of interspecific shared polymorphisms were significantly lower ( $P < 0.0001$  and  $P < 0.01$ , Mann–Whitney  $U$  test) in the highly differentiated regions (fig. 3c). Moreover, these regions exhibited significantly reduced levels of polymorphisms, skewed allele frequency spectra towards rare alleles (more negative Tajima's  $D$ ), and stronger signals of LD (higher  $r^2$ ) ( $P < 0.0001$ , Mann–Whitney  $U$  test) (fig. 3d–f). These findings demonstrated that the elevated genetic differentiation of genomic divergence islands resulted from reduced genetic diversity within species rather than increased absolute genetic distance between species. Additionally, recombination rates were significantly suppressed in outlier regions displaying exceptionally high or low interspecific differentiation relative to the genomic background (fig. 3g). Our analysis revealed a significant negative correlation between relative divergence, represented by  $F_{ST}$ , which depends on genetic diversity within species, and the recombination rates in *P. longipes* (Spearman's  $\rho = -0.057$ ,  $P = 6e^{-14}$ ) (supplementary fig. S3, Supplementary Material online). Conversely, we found a positive correlation between absolute divergence  $D_{XY}$  and recombination rates in *P. longipes* (Spearman's  $\rho = 0.016$ ,  $P = 0.035$ ) (supplementary fig. S3, Supplementary Material online). Collectively, these results suggest that natural selection, particularly long-term linked selection in *P. longipes*, may have a dominant role in the formation of genome islands of species divergence.

### Gene Copy Number Variation as the Source for Karst Adaptation in *P. longipes*

DNA copy number variations (CNVs) play a crucial role in the evolutionary adaptation and trait innovations of organisms (Ohno 2013; Wang et al. 2019). The mutation rate of CNVs is much higher than that of single base mutations (Lynch and Conery 2003; Lynch and Walsh 2007; Katju and Bergthorsson 2013). Here, we explored the role of the expansion of gene families such as whole-genome duplication (WGD) and tandem duplications in the adaptation of *P. longipes* to karst limestone.

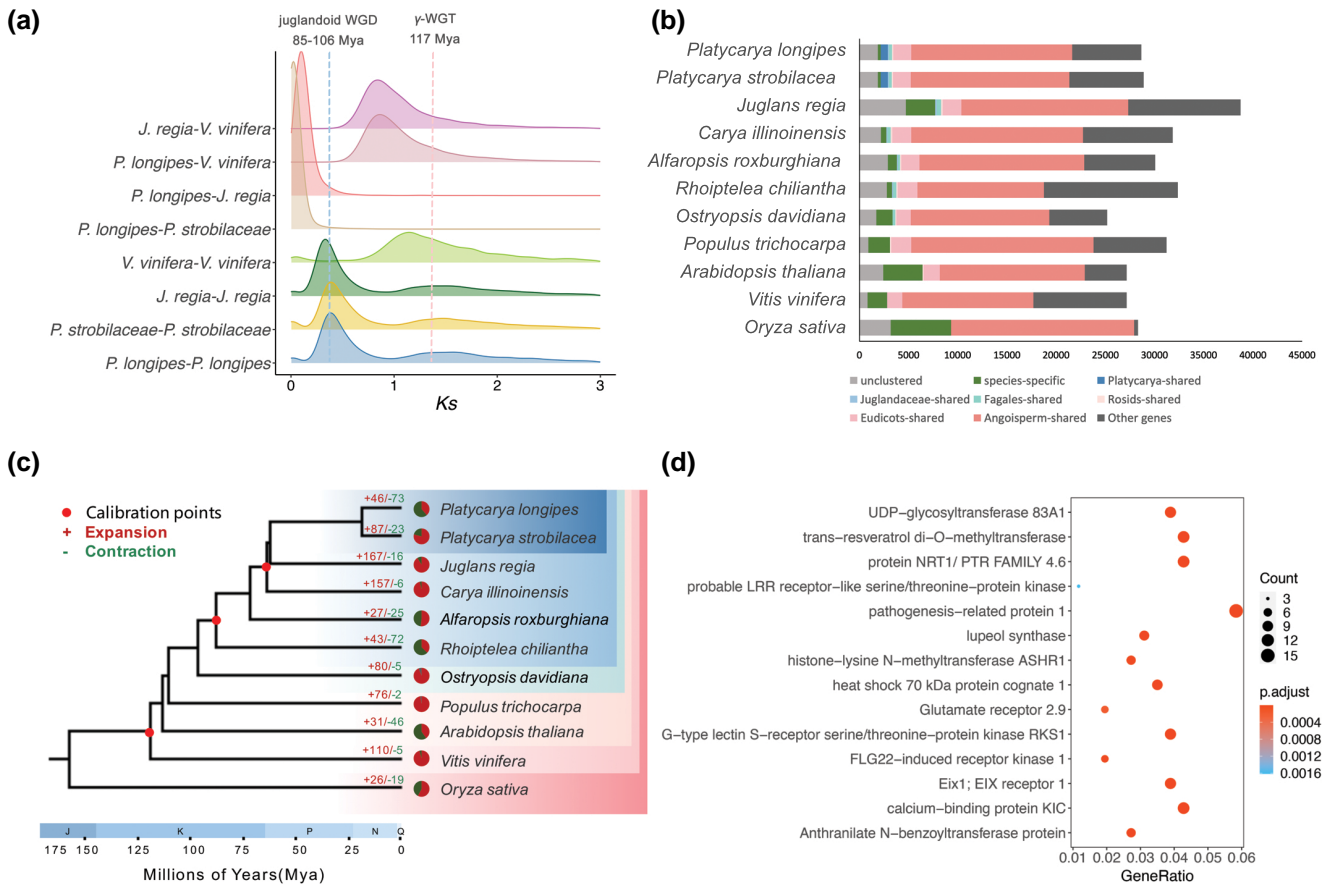
Analysis of synonymous substitutions per synonymous site ( $K_s$ ) between syntenic paralogs within the genomes shows that the genus *Platycarya* shared one ancient WGD event with the walnut family (fig. 4a), as previously suggested in several studies (Huang et al. 2019; Zhu et al. 2019; Ding et al. 2023). Based on sequence homology, a total of 327,237 genes were clustered into 23,452 families, of which 8,868 families were shared by all 11 angiosperm species included in the analysis (see Materials and Methods), and 130 were specific to *P. longipes* (fig. 4b). Using CAFÉ v4.2 (De Bie et al. 2006; Han et al. 2013), we inferred the ancestral gene content at each node of the species tree

covering 11 taxa across the angiosperms and modeled significant changes along each branch (fig. 4c). This analysis indicated that *P. longipes* had 46 significantly expanded gene families ( $P < 0.01$ ) containing 305 genes, compared with the inferred ancestral *Platycarya* genome (supplementary table S8, Supplementary Material online). Furthermore, our investigations into the origin of the expanded gene families revealed that the majority (79.67%) of these gene families were derived from tandem duplications, supporting the recent divergence of the two species. Additionally, 73 gene families were found to be significantly contracted in *P. longipes*, with only 186 genes remaining, whereas in *P. strobilacea*, there were 768 genes in these 73 gene families (supplementary table S9, Supplementary Material online).

Gene Ontology (GO) enrichment analysis revealed 14 molecular function terms and four biological process terms for the significantly expanded gene families (supplementary fig. S4 and table S10, Supplementary Material online). These terms include metabolism-related enzymes (such as sucrose synthase activity, terpene synthase activity, and transferase activity), ion channel and binding (such as extracellular glutamate-gated ion channel activity and manganese ion binding), response to wounding, recognition of pollen, protein translation, and sucrose metabolic process. Additionally, in the Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis, the expanded gene families were significantly enriched in 14 pathways. Among these pathways, three related to environmental adaptation were identified (fig. 4d and supplementary table S11, Supplementary Material online), including Ca-binding protein, *CML10* (K02183), heat shock 70 kDa protein (K03283), and glutamate receptor (*GLR*) 2.9 (K05387). Given that CNVs are frequently associated with molecular phenotypes, it is hypothesized that these pathways play a key role in the adaptation of *P. longipes*.

### Positive Selection of *TPC1A* Underlies the Karst Adaptation in *P. longipes*

To investigate the genetic basis of karst adaptation in *P. longipes*, we performed a genome-wide screen to identify potential molecular targets underlying stress response. Despite the complex demographic history between *P. longipes* and *P. strobilacea*, our analysis revealed multiple signatures of positive selection in *P. longipes*. After applying strict filtering criteria (see Materials and Methods), we identified 178 25-kb windows that are likely to be under positive selection in *P. longipes*, as indicated by a significant decorrelated composite of multiple signals (DCMS) score ( $P < 0.05$ ; fig. 5a). A functional enrichment analysis was conducted on the 232 genes within these candidate regions utilizing KEGG database, revealing a significant enrichment of genes involved in Ca influx channel (e.g., cyclic nucleotide-gated ion channel 2, *CNGC2*) and genes involved in stress-related processes (quinone oxidoreductase-like protein and polyadenylate-binding protein *RBP45*-like,  $FDR < 0.1$ , supplementary table S12,



**FIG. 4.** Comparative genomic analysis of *P. longipes* and *P. strobilacea*. (a) The  $K_s$  distribution for paralogous and orthologous gene pairs of *P. longipes* and *P. strobilacea*, *J. regia*, and *V. vinifera*. (b) Clusters of orthologous and paralogous gene families in *P. longipes* and *P. strobilacea* and nine other species of angiosperms. (c) Phylogenetic tree of 11 species based on a concatenated sequence alignment of 792 single-copy genes. Estimated divergence times and time scales are shown at the bottom. The solid circle represents the calibration node: the ancestor node of core eudicots (Aptian, 117 Ma; Jiao et al. 2012); the node age of the ancestor of the walnut family (Late Turonian to Santonian, 89.8–83.6 Ma; Heřmanová et al. 2011) based on the oldest fossil of *Rhoiptelea*; and ancestor node of Juglandoideae (Danian, 64 Ma; Zhang et al. 2013). The pie graphs represent the proportion of gene families that underwent expansion or contraction, compared with their most recent common ancestor. (d) KEGG enrichment of 305 genes involved in 46 significantly expanded gene families in *P. longipes*.

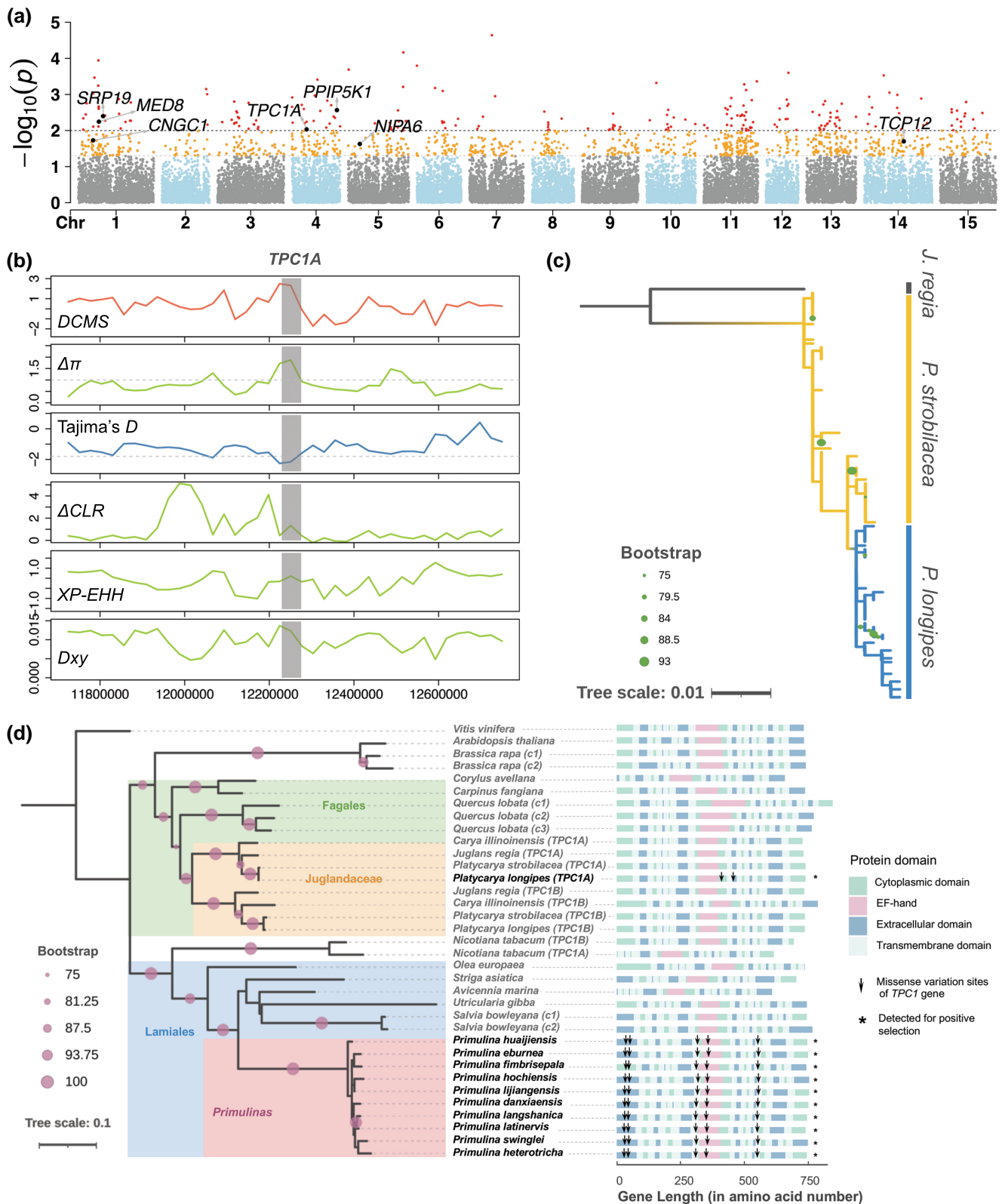
Supplementary Material online). Further examination of the positive selection signals and neighborhoods of these candidate genes led to the identification of seven genes exhibiting the strongest evidence of positive selection, *TPC1A*, *CNGC1*, *NIPA6*, *MED8*, *PPIP5K1*, *SRP19*, and *TCP12*, encompassing both stress-related and reproductive/developmental functional categories (fig. 5a and supplementary table S13, Supplementary Material online).

Among the seven candidate genes, *TPC1A* was found to be one of the two copies of *TPC1*, located on the recessive subgenome of *Platycarya* (Ding et al. 2023). *TPC1A* and *TPC1B* were generated after a WGD event, which was shared by all members of Juglandaceae (supplementary fig. S5, Supplementary Material online). Our analysis provided robust evidence of selection in the *TPC1A* region, as supported by all statistical tests conducted (fig. 5b). The genomic region of *TPC1A* was differentiated between *P. longipes* and *P. strobilacea* ( $F_{ST} = 0.522$ ) with a decreased genetic diversity in *P. longipes* relative to *P. strobilacea* ( $\pi_P$ ,  $\pi_{longipes}/\pi_{P. strobilacea} = 0.54$ ). Tajima's *D* showed that this

region had a significant negative value in *P. longipes* (−2.209) but not in *P. strobilacea* (−0.888) (fig. 5b). To further investigate the evolutionary relationship of *TPC1A* in *Platycarya*, a phylogenetic tree was reconstructed using the entire protein sequence of the gene for sequenced individuals of two *Platycarya* lineages, rooted with *J. regia*. The results showed that all accessions of *P. longipes* formed a monophyletic group nested within *P. strobilacea* (fig. 5c), implying that the *TPC1A* haplotypes of *P. longipes* originated from *P. strobilacea*. These findings suggest that *TPC1A* underwent positive selection in *P. longipes*.

*TPC1* is a Ca influx channel that is ubiquitous in all land plants and has a conserved function between monocots and dicots (Hedrich and Marten 2011; Dadacz-Narloch et al. 2013). This channel is presumed to be involved in the high-Ca stress response, and it is highly conserved among different core eudicots. Previous studies have reported that the *TPC1* gene underwent positive selection for high-Ca adaptation in the karst herb genus *Primulina* (Tao et al. 2016) (Gesneriaceae), implying convergent





**FIG. 5.** Signatures of positive selection and convergent adaptation in *P. longipes*. (a) Manhattan plot of DCMS values over the entire genome in a comparison of *P. longipes* versus *P. strobilacea*. Thresholds for significance at  $P = 0.05$  (bottom line) and  $P = 0.01$  (top line) are indicated. The x-axis shows each chromosome. (b) A zoom-in on five population genetic statistic scores, and DCMS scores, for the TPC1A region and its upstream and downstream 500 kb extension. Each data point is based on a sliding window analysis using nonoverlapping 25-kb windows. (c) A phylogenetic tree based on the entire length of the TPC1A gene, for the two *Platycarya* lineages and *J. regia* as the outgroup. (d) A phylogenetic tree (left panel) and structure of the TPC1 gene protein domain (right panel) highlight gene-level convergent evolution of the TPC1 gene in plants living in karst habitats. The arrows show the location of missense variations. The stars mark the gene copy that is under positive selection. The suffixes *c1*, *c2*, *c3*, TPC1A, and TPC1B behind species names indicate different copies of the TPC1 gene family in the specific genome.



adaptation of this gene in two distantly related karst taxa. To elucidate the origin of the adaptive evolution of *TPC1* in *P. longipes* and *Primulina*, we reconstructed a core eudicot scale phylogeny of the *TPC1* gene family (fig. 5d). Our results showed that Lamiales, including *Primulina* from different soil types, clustered into a highly conserved monophyletic group, and *Platycarya* clustered into another monophyletic group along with other Fagales. This indicates that the adaptive evolution of *TPC1* occurs independently in the two karst taxa. Furthermore, we compared the mutation sites on the *TPC1* sequence between *P. longipes* and *Primulina*. Haplotype analysis of *TPC1A* revealed a high degree of differentiation between *P. longipes* and *P. strobilacea*, as evidenced by 396 SNPs along the entire gene length (47,529 bp) that have an interspecific  $F_{ST}$  value exceeding 0.5 (supplementary fig. S5, Supplementary Material online). Of these SNPs, four were coding variations, with two being missense mutations at amino acid sites 412 and 447. They were identified in the cytoplasmic and transmembrane domains of *TPC1*, respectively, with the former located close to the EF-hand, which senses cytoplasmic  $Ca^{2+}$  levels. Tao et al. (2016) identified five missense variations under positive selection in karst adaptive herbs *Primulina*, including the one located in the domain of EF-hand and one adjacent to it (fig. 5d). These variations could potentially result in functional changes to the *TPC1(A)* gene and might have contributed to the karst adaptation observed in corresponding taxa. As the adaptive signals were found in the same gene but for different sites in two distantly related taxa, we conclude that the *TPC1* gene represents an example of gene-level convergent evolution for plants living in karst habitats, such as *P. longipes* and *Primulina*.

In addition to *TPC1A*, four additional genes related to ion transport also showed significant signals of selection (supplementary fig. S6, Supplementary Material online): *CNGC1*, a member of the *CNGCs* involved in the influx of cellular  $Ca^{2+}$  (Leng et al. 2002); *NIPA6*, an  $Mg^{2+}$  transporter; *PPIP5K1*, which participates in inward rectifier potassium channel activity; and *SRP19*, which is involved in the signaling pathway for G protein-coupled receptors. The positive selection of these genes may have contributed to the adaptation of *P. longipes* to the stress of high Ca, Mg, and low potassium levels in the karst environments. Furthermore, *MED8*, a component of the mediator complex, and a transcriptional regulator associated with plant defense, flowering time, and pollen tube growth in *Arabidopsis* (Lalanne et al. 2004; Kidd et al. 2009), was identified as one of the genes under positive selection that is related to reproduction (supplementary fig. S6, Supplementary Material online). *TCP12*, another positively selected gene (figs. 5a and 5b, Supplementary Material online), is a transcription factor that prevents axillary bud outgrowth and is probably involved in the auxin-induced control of apical dominance (Aguilar-Martínez et al. 2007). Interestingly, we found that *TCP12* was expressed in trace amounts in *P. strobilacea* but not in *P. longipes* across all three field-sampled populations (supplementary fig. S6,

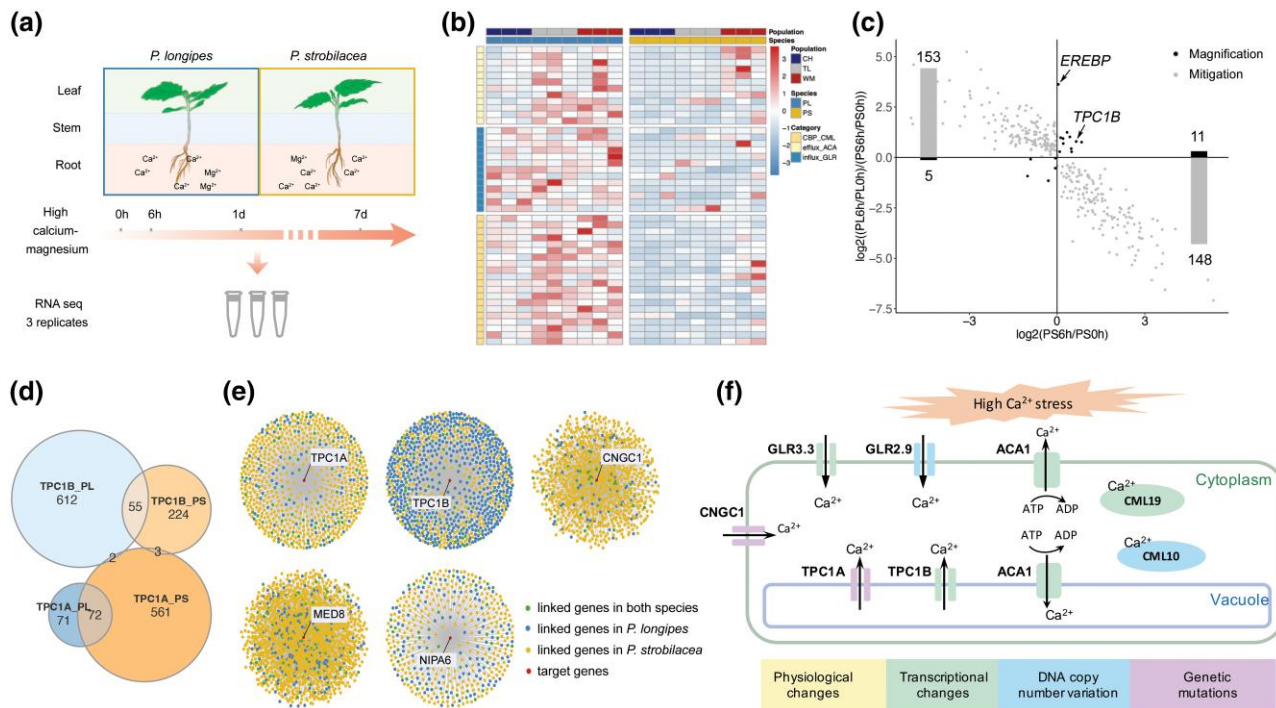
Supplementary Material online). Based on the function of this gene, we speculate that the *TCP12* gene may be involved in the morphological differentiation of *P. longipes* and *P. strobilacea*, where *P. strobilacea* exhibits more evident apical dominance with one main branch, whereas the branches of *P. longipes* radiate to the surroundings.

### Different Transcriptional Programs Indicate *P. longipes* Has a Greater Capacity to Respond to High Ca–Mg Stress than *P. strobilacea*

To systematically explore the transcriptional adaptations of *P. longipes* and compare the molecular responses to high Ca and Mg stress between the two species, we conducted RNA sequencing (RNA-seq) analysis upon 1) three versus three field-collected samples from both leaf and root tissues (fig. 1b) and 2) three versus three laboratory-collected samples from the root, stem, and leaf tissues with high Ca–Mg treatment for 0 h (baseline without treatment), 6 h, 1 day, and 7 days (fig. 6a).

First, to examine the transcriptional differences between two species in response to high Ca and Mg stress, we detected the differentially expressed genes (DEGs) between two species, *P. longipes* and *P. strobilacea*, in sympatric populations under similar climatic conditions but differing soil environments, using field-collected samples. Among the DEGs present in at least two populations, 4,296 genes were upregulated and 1,420 genes were downregulated in the leaf tissue of *P. longipes*, whereas 856 genes were upregulated and 913 genes were downregulated in the root tissue of *P. longipes*. Moreover, the upregulated genes in the wild leaf tissues of *P. longipes* were enriched in the three pathways associated with high Ca adaptation and ion transport across membranes ( $FDR < 0.05$ , fig. 6b). These pathways include *GLR 3.3*, putative Ca-binding protein *CML19*, and autoinhibited  $Ca^{2+}$ -ATPase 1 isoform 1, pointing to the systematic functional changes in Ca ion transport and metabolites in *P. longipes*.

Second, we evaluated the adaptive potential of the gene expression plasticity (He et al. 2021) of the two species under conditions of high Ca–Mg stress using laboratory-collected samples. We identified 913, 1,230, and 2,206 genes in each species that were differentially expressed in the leaf, root, and stem throughout time relative to 0 h, and 94%, 71%, and 87% of these genes were regulated in the same manner (supplementary fig. S7, Supplementary Material online). This indicated that under high Ca and Mg stress conditions, most DEGs are simultaneously regulated in two species. To assess the magnitude of plastic responses in *P. longipes* compared with *P. strobilacea*, we evaluated the slope of plastic response to high Ca–Mg stress at each time point  $t$ . Our analysis identified 78 genes in *P. longipes* with significantly enhanced response to stress ( $FDR < 0.1$ ). The KEGG enrichment analysis on these genes revealed terms such as autoinhibited  $Ca^{2+}$ -ATPase 1 isoform 1, Mg transporter *MRS2-2*, and dehydration-responsive element-binding protein (supplementary table S14, Supplementary Material



**FIG. 6.** Mechanism of high  $\text{Ca}^{2+}$  adaptation in *P. longipes*. (a) A schematic representation of the transcriptome experimental design. (b) Expression patterns of upregulated genes involved in the above three categories of  $\text{Ca}^{2+}$  concentration regulation in *P. longipes* compared with *P. strobilacea*. (c) Magnified and mitigated response to  $\text{Ca}^{2+}$  in the stem tissue of *P. longipes* compared with *P. strobilacea* after 6-h high Ca–Mg treatment.  $y$ -axis: the ratio of the response ratios;  $x$ -axis: the ratio of gene expression after 6 h in *P. strobilacea*. Bars represent the gene number in each quadrant of the plot. (d) The Euler diagram shows the intersections of the genes that show coexpression with *TPC1A* and *TPC1B*, in two species. *TPC1B\_PL* refers to the genes coexpressed with *TPC1B* in *P. longipes*, whereas *TPC1B\_PS* refers to the genes coexpressed with *TPC1B* in *P. strobilacea*. (e) The coexpression networks of genes under positive selection and *TPC1B*. (f) The relay mode of the high Ca adaptation mechanism in *P. longipes*. The color represents the adaptation stage of the gene. CMLs, CaM-like proteins; ACAs, Ca efflux channel autoinhibited  $\text{Ca}^{2+}$ -ATPases; CNGCs, Ca influx channel cyclic nucleotide-gated channels; GLRs, Ca influx channel glutamate receptor homologs; *TPC1s*, Ca influx channel two-pore channels.

online). The enhanced response of these genes to high Ca–Mg stress may contribute to the adaptation of *P. longipes* in karst field habitats.

Last, we compared the coexpression network of the two species, *P. longipes* and *P. strobilacea*. The coexpression network of *P. longipes* was clustered into 37 distinct modules, whereas the network of *P. strobilacea* was divided into 48 modules (supplementary fig. S7, Supplementary Material online). Out of the 37 modules in *P. longipes*, 28 (75.68%) were found to be conserved in *P. strobilacea*, whereas the remaining nine modules were unique to *P. longipes*. Further, we performed a functional enrichment analysis for each coexpression module. Interestingly, the nine unique coexpression modules from *P. longipes* are enriched in genes related to wounding response (supplementary table S15, Supplementary Material online), indicating that under the high Ca–Mg stress, the capacity of wounding response of *P. longipes* may differ from that of *P. strobilacea*.

In our findings, whereas no significant differences were observed in the expression of the genetically adapted *TPC1A* gene, the other *TPC1* gene, *TPC1B*, displayed significant upregulation in both species in the stem at 6 h and 1 day posttreatment, with a magnified response observed

in *P. longipes* (fig. 6c). Interestingly, our comprehensive evaluation of 578 coexpressed genes linked to *TPC1B* (fig. 6d and e) revealed that 190 of these genes exhibit differential expression between the two species, in at least one tissue. This result represents a significant enrichment of DEGs compared with the overall DEG pool, as determined by the chi-squared test ( $P = 0.0041$ ). This suggests that *TPC1B* and its interacting genes are more sensitive and active in regulating cytoplasmic  $\text{Ca}^{2+}$  influx in response to high Ca stress in *P. longipes*. On the other hand, the number of coexpressed genes linked to *TPC1A* was found to be significantly lower in *P. longipes* compared with *P. strobilacea*, suggesting that genetic changes have led to a substantial change in the molecular function of *TPC1A* in *P. longipes* (fig. 6d). These findings provided an additional line of the involvement of *TPC1* in the karst adaptation of *P. longipes*.

When focusing on the seven genes under positive selection in *P. longipes*, we found that although none of the seven genes was differentially expressed between two species across the measured tissues, three genes (*CNGC1*, *NIPA6*, and *MED8*) were significantly differentially coexpressed between two species ( $P_{\text{adjust}} < 0.0001$ ) (fig. 6e). These three genes belong to modules that were conserved in both

species, but the coexpression links between the candidate gene and other genes within the modules were different. Similar to *TPC1A*, the number of linked coexpression genes for these three genes is lower in *P. longipes* compared with in *P. strobilacea*. Regarding the expression patterns, the modules containing these three genes displayed higher expression levels in stem tissue compared with the other two tissues (as shown in [supplementary fig. S7, Supplementary Material](#) online, for module mediumpurple4, dark green, and indianred3). Further, the module containing *CNGC1* (mediumpurple4 in [supplementary fig. S7, Supplementary Material](#) online) was significantly downregulated under high Ca–Mg stress conditions, which is a common mechanism for regulating plant behavior in response to high Ca and Mg stress, to reduce cytoplasmic Ca influx.

Taken together, our results demonstrated distinct transcriptional programs of *TPC1* and other key genes/channels ([fig. 6f](#)) during the high Ca–Mg treatment in karst-adapted *P. longipes* and nonadapted *P. strobilacea*. This highlights the physiological adaptations and modifications in transcriptional plasticity of *P. longipes* and contributes to our understanding of the molecular mechanisms underlying karst adaptation in plants.

## Discussion

In this study, we investigated the evolutionary mechanisms that have shaped the adaptation of *P. longipes*, a woody plant, to the karst environment and observed incipient speciation for *Platycarya*. Our population structure analysis supports the “two species” model for the genus *Platycarya* (*P. strobilacea* and *P. longipes*) proposed by previous studies ([Luo 2015](#); [Wan et al. 2017](#)) and the original classification of the genus *Platycarya* ([Kuang and Lu 1979](#)). The fossil record indicates the widespread distribution of the genus *Platycarya* during the early Tertiary and its survival in East Asia during the Quaternary ice age ([Manchester 1999](#); [Zhekun and Momohara 2005](#)). Our population demography analysis estimates that the divergence between *P. longipes* and *P. strobilacea* occurred ~2.09 Mya during the Quaternary ([fig. 2d and e](#)), suggesting that the strong climatic oscillations of the Quaternary may have provided a possibility for a part of the ancestral populations to survive in karst habitats ([Bamba et al. 2019](#)).

Our study supports the relay race hypothesis ([Yona et al. 2015](#)) for woody plant adaptation to karst environments. By integrating genomic and transcriptomic data, we demonstrate that the adaptation of *P. longipes* to the karst environment involves both rapid nongenetic and long-lasting genetic mechanisms ([fig. 6f](#)). Previous studies have suggested that physiological adaptations are associated with changes in gene expression and alterations in transcriptional plasticity, which can lead to epigenetic adaptations ([Feinberg 2007](#); [Suzuki et al. 2020](#)). We identified the signals of genetic and physiological adaptations in *P. longipes*, especially in *TPC1A* and *TPC1B*. *TPC1A* displays

a strong signal of positive selection and contains a nonsynonymous mutation close to the Ca<sup>2+</sup>-binding EF-hands, which replaced the polar amino acid threonine with the nonpolar amino acid isoleucine, likely altering Ca transport regulation in *P. longipes*. According to [Johri et al. \(2022\)](#), sub- or neofunctionalization often occurs in one of the WGD paralogs, accompanied by the gradual fixation of nonsynonymous mutations and reduction in expression levels. Gene copies situated in recessive subgenomes are less constrained and display an increased likelihood of undergoing change. Our findings suggest that *TPC1A*, located in the recessive subgenome from WGD ([Ding et al. 2023](#)), has a lower activity and reduced number of coexpressed genes in response to high Ca stress in *P. longipes* ([fig. 6d](#)). In contrast, *TPC1B*, located in the dominant subgenome, is upregulated and coexpressed with a large number of DEGs during the high Ca–Mg treatment in *P. longipes* ([fig. 6d](#)), indicating alternative regulation to function in Ca influx. These results show that adaptations have occurred in both WGD paralogs of *TPC1* genes in *P. longipes*, but *TPC1A*, due to its location in the recessive subgenome, having been more prone to changes. We suspect that the molecular function of *TPC1A* may have changed, and *TPC1B* has been alternatively regulated to function in Ca influx. Although further functional validation is required, mutational changes and changes in transcriptome expression may have contributed to *P. longipes* maintaining cytoplasmic Ca ion homeostasis and preventing metal ion toxicity, enhancing its growth and development.

Our study offers insights into the convergent adaptive changes exhibited by *TPC1* in response to karst habitats, both in woody trees and herbs. Our results are in line with those of [Tao et al. \(2016\)](#), who detected multiple sites under positive selection in the *TPC1* gene during karst adaptation of *Primulina*. Although *P. longipes* and *Primulina* exhibit different mutations of the *TPC1* gene, the convergence on *TPC1* suggests the gene’s crucial role in mediating adaptability to high Ca stress in karst environments. Our phylogenetic analysis of the *TPC1* gene family in 26 core eudicots supports the notion of independent origins of mutations of *TPC1* ([fig. 5d](#)), implying that all these have been advantageous in enhancing fitness to cope with the shared high Ca stress in karst habitats. As a crucial component of Ca ion influx channels ([Taneja and Upadhyay 2021](#)), the ubiquitously present *TPC1* gene exhibits high conservation in sequence, copy number, and function ([Hedrich and Marten 2011](#); [Dadacz-Narloch et al. 2013](#)). It is noteworthy that *TPC1* is longer than a typical gene, making it more susceptible to mutations. Given the limited mechanisms for plants to develop Ca tolerance, it is plausible that most karst-endemic plants possess convergent adaptations in the *TPC1* gene. Further studies are imperative to validate this hypothesis.

Finally, our study suggests that *P. longipes* and *P. strobilacea* may be in the early stages of sympatric speciation, based on their concentrated distribution of highly differentiated regions on chromosomes, recent divergence, and abundance of intermediate morphological individuals



in hybrid populations. However, also here, further research is needed to fully understand the mechanisms contributing to reproductive isolation. Natural selection has been recognized as playing a crucial role in overcoming the homogenizing effect of gene flow (Schluter and Conte 2009; Feder et al. 2012; Nosil et al. 2021). Our findings highlight the profound impact of natural selection, particularly long-term linked selection, in generating genomic islands of divergence in the genus *Platycarya* (Sun et al. 2022). The observation of a negative relationship between population-scaled recombination rates and  $F_{ST}$ , but not with  $D_{XY}$ , suggests that natural selection has shaped patterns of genetic differentiation between the two species (Cruickshank and Hahn 2014; Wang et al. 2016). Prezygotic isolation, specifically earlier flowering time, may also contribute to reproductive isolation (Antonovics 2006). We identified the gene *MED8*, which is positively selected in *P. longipes* and linked to both plant defense and flowering time regulation. Nevertheless, to date, there have been no reports of the exact flowering time of *Platycarya* in sympatric regions of the two species. In conclusion, our study supports the “two species” model for the genus *Platycarya*, specifically for *P. strobilacea* and *P. longipes*. Our findings suggest that long-term linked selection in *P. longipes* played a dominant role in driving species divergence during the incipient stages of speciation of the two *Platycarya* lineages.

## Materials and Methods

### Sampling Collection and Sequencing

A total of 122 individuals from 36 populations of *P. strobilacea* and 85 individuals from 13 populations of *P. longipes* were obtained for the whole-genome resequencing (as depicted in fig. 1b and detailed in supplementary table S16, Supplementary Material online). For transcriptome sequencing, fresh tissues (roots and leaves) of *P. strobilacea* and *P. longipes* were procured from three populations (see supplementary table S17, Supplementary Material online, for details), where two species are found in sympatry. Simultaneously, seeds of *P. strobilacea* and *P. longipes* from a sympatric distributed site were subjected to RNA-seq under high Ca–Mg treatment.

### Genome Assembly and Annotation

A combination of three different sequencing strategies (short read from Illumina, subread from PacBio, and interaction read from Hi-C) was employed to obtain a high-quality chromosome-level assembly genome (supplementary table S16, Supplementary Material online). To determine the chromosome number of *Platycarya*, seeds of *P. strobilacea* and *P. longipes* were collected from Tianlin County, and a karyotype analysis of FISH was performed. The accuracy of the assembly was also assessed by mapping the clean Illumina reads to the assembly genome with BWA (Li and Durbin 2009).

For each genome, we identified repetitive sequences at the DNA and protein levels by a combination of

homology-based prediction and de novo identification and used TRF (<http://tandem.bu.edu/trf/trf.html>) to find tandem repeats. We predicted protein-coding gene structures combining de novo identification, homology-based prediction, and RNA-Seq-based prediction and then integrated this information into a nonredundant gene model set using EVIDENCEModeler (EVM) (Haas et al. 2008).

### SNP Calling

We mapped all reads to *P. strobilacea* reference genome with default settings implemented in BWA v0.7.12 using the BWA-MEM algorithm (Li 2013). Approximately 95% of *P. strobilacea* sequence reads and 88% of *P. longipes* sequence reads were accurately mapped to the reference genome. SNPs from each individual were called and joined to create a multisample SNP data set using SENTIEON DNaseq software packages v. 201808.08 (Weber et al. 2016). After quality control based on mapping depth and individual kinship (Manichaikul et al. 2010) (supplementary note and fig. S8, Supplementary Material online), a total of 11,282,483 high-quality SNPs from 130 unrelated individuals were used to detect genomic islands of divergence and positive selection. After masking the SNPs located in or near ~25 kb of the coding sequence, 3,264,018 SNPs were remained. To obtain independent SNPs, we thinned the SNPs using a distance filter of an interval of 25 kb based on LD results calculated by PopLDdecay v3.40 (Zhang et al. 2019) (supplementary fig. S8, Supplementary Material online). After LD filtering, 2,182 independent and neutral SNPs remained and were used for STRUCTURE analysis.

### Genetic Diversity and Population Structure

Population genetic structure was inferred using STRUCTURE v. 2.3.4 (Pritchard et al. 2000). The optimal value of  $K$  was determined by  $\Delta K$  and the rate of change in  $\ln(D|K)$  between successive  $K$  values (Evanno et al. 2005). All 130 unrelated individuals were assigned to a cluster with an ancestry index greater than 0.85. PCA was run using the R package SNPRelate v. 1.6.2 (Zheng et al. 2012) with default settings. Additionally, a phylogenetic tree of 130 unrelated samples was constructed using SVDquartets (Chifman and Kubatko 2014) with 1,000 replicates for bootstrap confidence analysis, based on 2,182 SNPs presumed to be neutral and independent, using *J. regia* as the outgroup.

Nucleotide diversity ( $\pi$ ) and fixation index were calculated for the 11,282,483 SNPs in 25 kb nonoverlapping windows using VCFtools v0.1.17 (Danecek et al. 2011) and homozygosity (ROH) were identified with plink v1.9 (Chang et al. 2015) with a sliding window of 25 kb. Population-scaled recombination rates ( $\rho = 4N_e c$ ) were estimated using LDhat v.2.2 (McVean et al. 2004; Auton and McVean 2007), with 10,000,000 MCMC iterations sampling every 2,000 iterations and a block penalty parameter of five. The observed genome heterozygosity ( $H_{e,c}$ ) and absolute pairwise nucleotide divergence between species

( $D_{xy}$ ) were calculated for the consensus sequence using python scripts (<https://github.com/yongshuai-sun/hhs-omei/>).  $H_{et}$  was estimated for each individual as the proportion of heterozygous variants in the genome.

### Demography History

We employed PSMC (Li and Durbin 2011) to estimate the changes in the population sizes ( $N_e$ ) over historical time in both species. Four individuals of *P. strobilacea* and *P. longipes* were mapped to the *P. strobilacea* and *P. longipes* reference genomes, respectively. The recommendations of using sequencing data with a mean genome coverage of  $\geq 18$ , a per-site filter of  $\geq 10$  reads, and no more than 25% of missing data were followed (Nadachowska-Brzyska et al. 2016). For all conversations of demography parameters, a generation time of 30 years and a mutation rate of  $2.06 \times 10^{-9}$  per site per year were used (Bai et al. 2018).

To conduct the coalescent simulations, we extracted 458 neutral and independent 300–1,000 bp noncoding loci, which was at least 25 kb apart from each. The haplotypes of each locus were reconstructed using PHASE 2.1 (Stephens and Donnelly 2003). Based on 200 loci randomly selected from the 458 loci and 47 individuals (20 samples of *P. longipes* and 27 samples of *P. strobilacea*), gene flow and divergence of *P. strobilacea* and *P. longipes* were assessed under the “Isolation with Migration” Bayesian framework of IMA3 (Hey et al. 2018).

### Identification of Genomic Islands of Divergence

To identify the genomic islands of divergence, we compared the observed  $F_{ST}$  distribution to the neutral  $F_{ST}$  distribution based on the best-fitting population demography model (Malinsky et al. 2015; Wang et al. 2016; Sun et al. 2022) using software ms (Hudson 2002). To evaluate the compatibility between the simulated and observed data, nucleotide diversity ( $\pi$ ) and Tajima’s  $D$  were calculated as summary statistics. To assess whether the observed  $F_{ST}$  values deviated significantly from neutral expectations, the conditional probability ( $P$ ) of observing more extreme interspecific  $F_{ST}$  values among simulated data sets than among the observed data was estimated. A correction for multiple testing was performed using FDR adjustment, and windows with FDR  $< 1\%$  were considered the highly (lower) divergent regions.

### Genome Synteny Analysis

Protein sequences within and between genomes were searched against one another to detect putative homologous genes ( $E < 1e^{-5}$ ) by BLASTP. MCScanX (Wang et al. 2012) was implemented to infer homologous blocks involving collinear genes within genomes. The maximal gap length between collinear genes along a chromosome region was set to 50 genes (Wang et al. 2017, 2018). Tbttools (Chen et al. 2020) was used to visualize the collinear blocks of *P. longipes* and *P. strobilacea*. JCVI (Tang et al.

2008) was used to infer and plot homologous blocks between genomes.

### Gene Family Expansion and Contraction Analysis

The protein sequences of 11 plant species were used to cluster gene families, including six from the walnut family, *P. longipes*, *P. strobilacea*, *Alfaropsis roxburghiana* (Ding et al. 2023), *Rhoiptelea chiliantha* (Ding et al. 2023), *Carya illinoensis* (Lovell et al. 2021), and *J. regia* (Zhu et al. 2019), and five well-annotated outgroup plants, *Oryza sativa* (Goff et al. 2002), *Vitis vinifera* (Jaillon et al. 2007), *Arabidopsis thaliana* (Arabidopsis Genome Initiative 2000), *Ostryopsis davidiana* (Wang et al. 2021), and *Populus trichocarpa* (Tuskan et al. 2006). For genes with alternative splicing variants, the longest transcript was selected. An all-against-all comparison was performed using BLASTP (Altschul et al. 1990) with an  $E$  value cutoff of  $1e^{-5}$ , and OrthoFinder (Emms and Kelly 2019) was used to cluster gene families. The 792 gene families with only one copy from each of the 11 species were regarded as single-copy genes and used for subsequent analysis. Codons 1, 2, and 3 were joined into three “supergenes” for each species.

The divergence times of 11 species were calculated using the MCMCTREE program implemented in the Phylogenetic Analysis by Maximum Likelihood (PAML) package (Yang 2007). Three calibration nodes, the node of the core eudicots (Aptian, 117 Ma; Jiao et al. 2012), the node of the walnut family (Late Turonian to Santonian, 89.8–83.6 Ma; Heřmanová et al. 2011), and the node of Juglandoideae (Danian, 64 Ma; Zhang et al. 2013), were applied to calibrate the divergence times.

Finally, to analyze the expansion and contraction of gene families, CAFÉ v4.2 (De Bie et al. 2006; Han et al. 2013) was run under a random birth-and-death model. The clustering results and the information from the estimated divergence times were used. Using conditional likelihood as the test statistic, the corresponding  $P$  values of each lineage were calculated, and a  $P$  value of 0.01 was regarded as significant. The expanded and contracted gene families were performed KEGG and GO enrichment to determine their functions.

### Detection of Positive Selection in *P. longipes*

To detect regions under selection across the genome of *P. longipes*, we scanned the genome for multiple patterns of molecular variation: 1) local losses of heterozygosity; 2) locally elevated levels of genetic differentiation; 3) distortions in the allele frequency spectrum; and 4) long-range haplotypes of high population frequency. These types of signatures were conducted using a combination of statistics calculated in nonoverlapping 25-kb windows: reduction of Tajima’s  $D$  (Tajima 1989), composite likelihood ratio (CLR, Nielsen et al. 2005), nucleotide diversity cross-population extended haplotype homozygosity (XP-EHH, Sabeti et al. 2007), and absolute pairwise nucleotide divergence between species ( $D_{xy}$ ). Windows with

fewer than ten SNPs were eliminated. For each window, we further computed  $\Delta\pi$  as  $\pi_{P. strobilacea}/\pi_{P. longipes}$  and  $\Delta CLR$  as  $CLR_{P. longipes} - CLR_{P. strobilacea}$ . The window size was chosen based on patterns of LD decay (supplementary fig. S2, Supplementary Material online).

All the above five statistics were combined using the DCMS method (Ma et al. 2015). The correlation of the  $P$  value of each statistic was calculated and used to calculate their respective weight factors (supplementary table S18, Supplementary Material online). Next, the DCMS was estimated for each 25-kb window, and a  $P$  value was derived for each window. Regions under putative positive selection were defined as the windows with  $P < 0.05$ . Windows showing more diversity in *P. longipes* than in *P. strobilacea* ( $\Delta\pi > 1$ ) were excluded if there was no significant negative Tajima's  $D$  in *P. longipes* (Tajima's  $D > -1.8$ ). Outlier windows with interval lengths less than 25 kb were merged for each detection. Only genes with potential impacts (MODIFIER, MODERATE, and HIGH in the SnpEff annotation results) were considered candidate genes under selection during the karst adaptation process of *P. longipes*.

GO and KEGG enrichment analyses were conducted using the R package clusterProfiler with all annotated genes in the *P. strobilacea* genome as background and 232 genes encoded in the regions under putative positive selection as foreground. The Benjamini–Hochberg method was used for the correction of multiple tests.

### Transcriptome Analysis

Reads containing adapter or poly-N and with low quality were filtered out by Trimmomatic v0.39 (Bolger et al. 2014) before mapping to the reference genome of *P. strobilacea*. The matrixes of transcript abundance both in TPM and raw counts were obtained in each sample using Salmon v1.8.0 (Patro et al. 2017) as implemented in the pipeline of the R package bears (Almeida-Silva and Venancio 2021). DESeq2 R package v1.10.1 was used to perform differential expression analysis and gene expression variation over time. Gene expression plasticity was evaluated via the methodology outlined by He et al. (2021). We inferred a signed gene coexpression network with the R package BioNERO (Almeida-Silva and Venancio 2022) and performed differential coexpression analysis using the R package diffcoexp (Wei et al. 2022). We further compared the expression profile of genes encoded in the genomic regions under putative positive selection in *P. longipes* and *P. strobilacea*.

### Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

### Acknowledgments

D.-Y.Z. acknowledges funding from the National Natural Science Foundation of China (32170223 and 31421063),

the “111” Program of Introducing Talents of Discipline to Universities (B13008), Beijing Advanced Innovation Program for Land Surface Processes, and the National Key R&D Program of China (2017YFA0605104). Y.V.d.P. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (No. 833522) and from Ghent University (Methusalem funding, BOF.MET.2021.0005.01). Y.C. acknowledges funding from China Scholarship Council (No. 202106040114). We thank Jun Chen for help with running Ima3 software, Yang Yang for help with using the PAML software, Wen-Juan Lan for collection of laboratory transcriptome samples, Yu Liang for his help with R scripts to visualize figures, and Xiao-Xu Pang, Xin-Rui Lin, Rui-Min Yu, and Lin-Lin Xu for their help with Perl or Python scripts. We thank Zhen Li, Heng-Chi Chen, and Xiao Ma for their help with the comparative analysis.

### Author Contributions

D.-Y.Z. conceived the research, Y.V.d.P. polished it. All analyses were performed by Y.C., with contributions to gene coexpression network analysis and gene differential expression analysis from F.A.-S. and D.B., respectively. W.-P.Z. contributed plant material collection and guided on the calculation of recombination rate, Y.-M.D. guided on the analysis of population genomics, Y.C., B.-W.Z., W.-N.B., Y.V.d.P. and D.-Y.Z. wrote the manuscript. D.-Y.Z. and Y.V.d.P. supervised the project.

### Data Availability

All sequencing data used in this study have been deposited at GenBank under the accession PRJNA356989. The reference genome and gene annotations have also been deposited on the website (<http://cmb.bnu.edu.cn/juglans>). All codes used for the main analyses in this paper are available for download from <https://github.com/Caoyu819/protocols-for-karst-adaptation-of-Platycarya.git>.

### References

- Aguilar-Martínez JA, Poza-Carrón C, Cubas P. 2007. *Arabidopsis* BRANCHED1 acts as an integrator of branching signals within axillary buds. *Plant Cell* **19**:458–472.
- Almeida-Silva F, Venancio T. 2021. bears: building expression Atlas from RNA-Seq data. R package version 0.99.0.
- Almeida-Silva F, Venancio TM. 2022. BioNERO: an all-in-one R/Bioconductor package for comprehensive and easy biological network reconstruction. *Funct Integr Genomic*. **22**:131–136.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*. **215**:403–410.
- Antonovics J. 2006. Evolution in closely adjacent plant populations X: long-term persistence of prereproductive isolation at a mine boundary. *Heredity (Edinb)*. **97**:33–37.
- Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**:796–815.



- Auton A, McVean G. 2007. Recombination rate estimation in the presence of hotspots. *Genome Res.* **17**:1219–1227.
- Bai WN, Yan PC, Zhang BW, Woeste KE, Lin K, Zhang DY. 2018. Demographically idiosyncratic responses to climate change and rapid Pleistocene diversification of the walnut genus *Juglans* (Juglandaceae) revealed by whole-genome sequences. *New Phytol.* **217**:1726–1736.
- Bamba M, Kawaguchi YW, Tsuchimatsu T. 2019. Plant adaptation and speciation studied by population genomic approaches. *Dev Growth Differ.* **61**:12–24.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**:7.
- Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant.* **13**:1194–1202.
- Chen S-C, Zhang L, Zeng J, Shi F, Yang H, Mao Y-R, Fu C-X. 2012. Geographic variation of chloroplast DNA in *Platycarya strobilacea* (Juglandaceae). *J Syst Evol.* **50**:374–385.
- Chifman J, Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* **30**:3317–3324.
- Clements R, Sodhi NS, Schilthuizen M, Ng PKL. 2006. Limestone karsts of Southeast Asia: imperiled arks of biodiversity. *BioScience* **56**:733–742.
- Cruikshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol.* **23**:3133–3157.
- Dadacz-Narloch B, Kimura S, Kurusu T, Farmer EE, Becker D, Kuchitsu K, Hedrich R. 2013. On the cellular site of two-pore channel *TPC1* action in the Poaceae. *New Phytol.* **200**:663–674.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**:2156–2158.
- Darwin C. 2004. *On the origin of species*. 1859. London: Routledge.
- De Bie T, Cristianini N, Demuth JP, Hahn MW. 2006. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**:1269–1271.
- Ding YM, Pang XX, Cao Y, Zhang WP, Renner SS, Zhang DY, Bai WN. 2023. Genome structure-based Juglandaceae phylogenies contradict alignment-based phylogenies and substitution rates vary with DNA repair genes. *Nat Commun.* **14**:617.
- Emms DM, Kelly S. 2019. Orthofinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**:238.
- Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* **14**:2611–2620.
- Fang J, Wang Z, Tang Z. 2011. *Atlas of woody plants in China: distribution and climate*. Beijing: Springer Science & Business Media.
- Feder JL, Egan SP, Nosil P. 2012. The genomics of speciation-with-gene-flow. *Trends Genet.* **28**:342–350.
- Feinberg AP. 2007. Phenotypic plasticity and the epigenetics of human disease. *Nature* **447**:433–440.
- Feng C, Wang J, Wu L, Kong H, Yang L, Feng C, Wang K, Rausher M, Kang M. 2020. The genome of a cave plant, *Primulina huaijiensis*, provides insights into adaptation to limestone karst habitats. *New Phytol.* **227**:1249–1263.
- Fukuhara T, Tokumaru S. 2014. Inflorescence dimorphism, heterodichogamy and thrips pollination in *Platycarya strobilacea* (Juglandaceae). *Ann Bot.* **113**:467–476.
- Fukushima K, Fang X, Alvarez-Ponce D, Cai H, Carretero-Paulet L, Chen C, Chang TH, Farr KM, Fujita T, Hiwatashi Y, et al. 2017. Genome of the pitcher plant *Cephalotus* reveals genetic changes associated with carnivory. *Nat Ecol Evol.* **1**:59.
- Geekiyange N, Goodale UM, Cao K, Kitajima K. 2019. Plant ecology of tropical and subtropical karst ecosystems. *Biotropica* **51**:626–640.
- Goff SA, Ricke D, Lan T-H, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**:92–100.
- Gou X, Wang Z, Li N, Qiu F, Xu Z, Yan D, Yang S, Jia J, Kong X, Wei Z, et al. 2014. Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. *Genome Res.* **24**:1308–1315.
- Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**:R7.
- Han MV, Thomas GW, Lugo-Martinez J, Hahn MW. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol.* **30**:1987–1997.
- Hao Z, Kuang Y, Kang M, Niu S. 2014. Untangling the influence of phylogeny, soil and climate on leaf element concentrations in a biodiversity hotspot. *Funct Ecol.* **29**:165–176.
- He F, Steige KA, Kovacova V, Gobel U, Bouzid M, Keightley PD, Beyer A, de Meaux J. 2021. Cis-regulatory evolution spotlights species differences in the adaptive potential of gene expression plasticity. *Nat Commun.* **12**:3376.
- He Z, Xu S, Zhang Z, Guo W, Lyu H, Zhong C, Boufford DE, Duke NC, International Mangrove C, Shi S. 2020. Convergent adaptation of the genomes of woody plants at the land-sea interface. *Natl Sci Rev.* **7**:978–993.
- Hedrich R, Marten I. 2011. TPC1–SV channels gain shape. *Mol Plant.* **4**:428–441.
- Heřmanová Z, Kvaček J, Friis EM. 2011. *Budvaricarpus serialis* Knobloch & Mai, an unusual new member of the Normapolles complex from the Late Cretaceous of the Czech Republic. *Int J Plant Sci.* **172**:285–293.
- Hey J, Chung Y, Sethuraman A, Lachance J, Tishkoff S, Sousa VC, Wang Y. 2018. Phylogeny estimation by integration over isolation with migration models. *Mol Biol Evol.* **35**:2805–2818.
- Huang Y, Xiao L, Zhang Z, Zhang R, Wang Z, Huang C, Huang R, Luan Y, Fan T, Wang J, et al. 2019. The genomes of pecan and Chinese hickory provide insights into *Carya* evolution and nut nutrition. *Gigascience* **8**:giz036.
- Hudson RR. 2002. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* **18**:337–338.
- Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Cassagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, et al. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**:463–467.
- Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR, McNeal J, Rolf M, Ruzicka DR, Wafula E, Wickett NJ, et al. 2012. A genome triplication associated with early diversification of the core eudicots. *Genome Biol.* **13**:1–14.
- Jin W, Long Y, Fu C, Zhang L, Xiang J, Wang B, Li M. 2018. Ca<sup>2+</sup> imaging and gene expression profiling of *Lonicera Confusa* in response to calcium-rich environment. *Sci Rep.* **8**:7068.
- Johri P, Gout JF, Doak TG, Lynch M. 2022. A population-genetic lens into the process of gene loss following whole-genome duplication. *Mol Biol Evol.* **39**:msac118.
- Katju V, Bergthorsson U. 2013. Copy-number changes in evolution: rates, fitness effects and adaptive significance. *Front Genet.* **4**:273.
- Kautt AF, Kratochwil CF, Nater A, Machado-Schiaffino G, Olave M, Henning F, Torres-Dowdall J, Harer A, Hulsey CD, Franchini P, et al. 2020. Contrasting signatures of genomic divergence during sympatric speciation. *Nature* **588**:106–111.
- Kidd BN, Edgar CI, Kumar KK, Aitken EA, Schenk PM, Manners JM, Kazan K. 2009. The mediator complex subunit PFT1 is a key regulator of jasmonate-dependent defense in *Arabidopsis*. *Plant Cell* **21**:2237–2252.
- Kozłowski G, Bétrisey SB, Song YG. 2018. *Wingnuts (Pterocarya) and walnut family. Relict trees: linking the past, present and future*. Switzerland: Natural History Museum Fribourg.
- Kuang KR, Lu AM. 1979. *Juglandaceae*. In: *Flora Reipublicae Popularis Sinica*. Beijing: Science Press. p. 8–9.

- Lalanne E, Michaelidis C, Moore JM, Gagliano W, Johnson A, Patel R, Howden R, Vielle-Calzada J-P, Grossniklaus U, Twell D. 2004. Analysis of transposon insertion mutants highlights the diversity of mechanisms underlying male progamic development in *Arabidopsis*. *Genetics* **167**:1975–1986.
- Leng Q, Mercier RW, Hua B-G, Fromm H, Berkowitz GA. 2002. Electrophysiological analysis of cloned cyclic nucleotide-gated ion channels. *Plant Physiol.* **128**:400–410.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:303.3997.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760.
- Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* **475**:493–496.
- Li Y, Liu Z, Shi P, Zhang J. 2010. The hearing gene Prestin unites echolocating bats and whales. *Curr Biol.* **20**:R55–R56.
- Liu Y, Cotton JA, Shen B, Han X, Rossiter SJ, Zhang S. 2010. Convergent sequence evolution between echolocating bats and dolphins. *Curr Biol.* **20**:R53–R54.
- Lovell JT, Bentley NB, Bhattarai G, Jenkins JW, Sreedasyam A, Alarcon Y, Bock C, Boston LB, Carlson J, Cervantes K, et al. 2021. Four chromosome scale genomes and a pan-genome annotation to accelerate pecan tree breeding. *Nat Commun.* **12**:4125.
- Lu A-M, Stone D, Grauke L. 1999. Juglandaceae. Flora of China. Beijing: Science Press.
- Luo R-S. 2015. *Systematics studies on endemic plants of Platycarya (Juglandaceae) in Eastern Asia*. Guangxi: Guangxi Normal University.
- Lynch M, Conery JS. 2003. The evolutionary demography of duplicate genes. *Genome Evol.* **3**:35–44.
- Lynch M, Walsh B. 2007. *The origins of genome architecture*: Sinauer Associates Sunderland, MA.
- Lyu H, He Z, Wu Cl, Shi S. 2018. Convergent adaptive evolution in marginal environments: unloading transposable elements as a common strategy among mangrove genomes. *New Phytol.* **217**:428–438.
- Ma Y, Ding X, Qanbari S, Weigend S, Zhang Q, Simianer H. 2015. Properties of different selection signature statistics and a new strategy for combining them. *Heredit (Edinb)*. **115**:426–436.
- Malinsky M, Challis RJ, Tyers AM, Schiffels S, Terai Y, Ngatunga BP, Miska EA, Durbin R, Genner MJ, Turner GF. 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science* **350**:1493–1498.
- Manchester SR. 1999. Biogeographical relationships of North American tertiary floras. *Ann Mo Bot Gard.* **86**:472–522.
- Mandakova T, Lysak MA. 2018. Post-polyploid diploidization and diversification through dysploid changes. *Curr Opin Plant Biol.* **42**:55–65.
- Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**:2867–2873.
- Manos PS, Stone DE. 2001. Evolution, phylogeny, and systematics of the Juglandaceae. *Ann Mo Bot Gard.* **88**:231–269.
- McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. 2004. The fine-scale structure of recombination rate variation in the human genome. *Science* **304**:581–584.
- Nadachowska-Brzyska K, Burri R, Smeds L, Ellegren H. 2016. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. *Mol Ecol.* **25**:1058–1072.
- Nie YP, Chen HS, Wang KL, Tan W, Deng PY, Yang J. 2010. Seasonal water use patterns of woody species growing on the continuous dolostone outcrops and nearby thin soils in subtropical China. *Plant Soil* **341**:399–412.
- Nielsen R, Wakeley J. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* **158**:885–896.
- Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. 2005. Genomic scans for selective sweeps using SNP data. *Genome Res.* **15**:1566–1575.
- Nosil P, Feder JL, Gompert Z. 2021. How many genetic changes create new species? *Science* **371**:777–779.
- Ohno S. 2013. *Evolution by gene duplication*. New York: Springer Science & Business Media.
- Oliver PM, Laver RJ, De Mello Martins F, Pratt RC, Hunjan S, Moritz CC. 2017. A novel hotspot of vertebrate endemism and an evolutionary refugium in tropical Australia. *Divers Distrib.* **23**:53–66.
- Parker J, Tsakogea G, Cotton JA, Liu Y, Provero P, Stupka E, Rossiter SJ. 2013. Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* **502**:228–231.
- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods.* **14**:417–419.
- Preite V, Sailer C, Syllwasschy L, Bray S, Ahmadi H, Kramer U, Yant L. 2019. Convergent evolution in *Arabidopsis halleri* and *Arabidopsis arenosa* on calamine metalliferous soils. *Philos Trans R Soc Lond B Biol Sci.* **374**:20180243.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**:945–959.
- Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, et al. 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**:913–918.
- Sackton TB, Clark N. 2019. Convergent evolution in the genomics era: new insights and directions. *Philos Trans R Soc Lond B Biol Sci.* **374**:20190102.
- Schluter D, Conte GL. 2009. Genetics and ecological speciation. *Proc Natl Acad Sci USA.* **106**(Suppl 1):9955–9962.
- Stephens M, Donnelly P. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet.* **73**:1162–1169.
- Sun Q, Wang R, Chen X. 2022. Genomic island of divergence during speciation and its underlying mechanisms. *Biodiversity Sci.* **30**:21383.
- Suzuki Y, McKenna KZ, Nijhout HF. 2020. Regulation of phenotypic plasticity from the perspective of evolutionary developmental biology. In: *Phenotypic switching*. p.403–442.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585–595.
- Taneja M, Upadhyay SK. 2021. An introduction to the calcium transport elements in plants, Santosh Kumar Upadhyay, editors. *Calcium transport elements in plants*. Cambridge, MA, USA: Academic Press. p.1–18.
- Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. 2008. Synteny and collinearity in plant genomes. *Science* **320**:486–488.
- Tao J, Feng C, Ai B, Kang M. 2016. Adaptive molecular evolution of the two-pore channel 1 gene *TPC1* in the karst-adapted genus *Primulina* (Gesneriaceae). *Ann Bot.* **118**:1257–1268.
- Todesco M, Owens GL, Bercovich N, Legare JS, Soudi S, Burge DO, Huang K, Ostevik KL, Drummond EBM, Imerovski I, et al. 2020. Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature* **584**:602–607.
- Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**:1596–1604.
- Wan Q, Zheng Z, Huang K, Guichoux E, Petit RJ. 2017. Genetic divergence within the monotypic tree genus *Platycarya* (Juglandaceae) and its implications for species' past dynamics in subtropical China. *Tree Genet Genomes.* **13**:1–11.
- Wang Z, Jiang Y, Bi H, Lu Z, Ma Y, Yang X, Chen N, Tian B, Liu B, Mao X, et al. 2021. Hybrid speciation via inheritance of alternate alleles of parental isolating genes. *Mol Plant.* **14**:208–222.
- Wang J, Street NR, Scofield DG, Ingvarsson PK. 2016. Variation in linked selection and recombination drive genomic divergence during allopatric speciation of European and American aspens. *Mol Biol Evol.* **33**:1754–1767.
- Wang J, Sun P, Li Y, Liu Y, Yu J, Ma X, Sun S, Yang N, Xia R, Lei T, et al. 2017. Hierarchically aligning 10 legume genomes

- establishes a family-level genomics platform. *Plant Physiol.* **174**:284–300.
- Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, 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.
- Wang N, Yang Y, Moore MJ, Brockington SF, Walker JF, Brown JW, Liang B, Feng T, Edwards C, Mikenas J, *et al.* 2019. Evolution of Portulacineae marked by gene tree conflict and gene family expansion associated with adaptation to harsh environments. *Mol Biol Evol.* **36**:112–126.
- Wang JP, Yu JG, Li J, Sun PC, Wang L, Yuan JQ, Meng FB, Sun SR, Li YX, Lei TY, *et al.* 2018. Two likely auto-tetraploidization events shaped kiwifruit genome and contributed to establishment of the Actinidiaceae family. *iScience* **7**:230–240.
- Weber JA, Aldana R, Gallagher BD, Edwards JS. 2016. Sentieon DNA pipeline for variant detection—software-only solution, over 20× faster than GATK 3.3 with identical results. *PeerJ PrePrints.* **4**: e1672v1672.
- Wei W, Amberkar S, Hide W. 2022. Diffcoexp: differential co-expression analysis. R package version 1.14.0.
- Wu DD, Yang CP, Wang MS, Dong KZ, Yan DW, Hao ZQ, Fan SQ, Chu SZ, Shen QS, Jiang LP, *et al.* 2020. Convergent genomic signatures of high-altitude adaptation among domestic mammals. *Natal Sci Rev.* **7**:952–963.
- Xie D-F, Cheng R-Y, Fu X, Zhang X-Y, Price M, Lan Y-L, Wang C-B, He X-J. 2021. A combined morphological and molecular evolutionary analysis of karst-environment adaptation for the genus *Urophysa* (Ranunculaceae). *Front Plant Sci.* **12**:667988.
- Xu S, He Z, Guo Z, Zhang Z, Wyckoff GJ, Greenberg A, Wu CI, Shi S. 2017. Genome-wide convergence during evolution of mangroves from woody plants. *Mol Biol Evol.* **34**:1008–1015.
- Xu S, Wang J, Guo Z, He Z, Shi S. 2020. Genomic convergence in the adaptation to extreme environments. *Plant Commun.* **1**:100117.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* **24**:1586–1591.
- Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, Degner JC, Nurkowski KA, Smets P, Wang T, Gray LK. 2016. Convergent local adaptation to climate in distantly related conifers. *Science* **353**: 1431–1433.
- Yona AH, Frumkin I, Pilpel Y. 2015. A relay race on the evolutionary adaptation spectrum. *Cell* **163**:549–559.
- Zhang C, Dong SS, Xu JY, He WM, Yang TL. 2019. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**:1786–1788.
- Zhang Z-H, Hu G, Zhu J-D, Luo D-H, Ni J. 2010. Spatial patterns and interspecific associations of dominant tree species in two old-growth karst forests, SW China. *Ecol Res.* **25**:1151–1160.
- Zhang JB, Li RQ, Xiang XG, Manchester SR, Lin L, Wang W, Wen J, Chen ZD. 2013. Integrated fossil and molecular data reveal the biogeographic diversification of the eastern Asian-eastern North American disjunct hickory genus (*Carya* Nutt.). *PLoS One* **8**:e70449.
- Zhekun Z, Momohara A. 2005. Fossil history of some endemic seed plants of East Asia and its phytogeographical significance. *Acta Bot Yunnanica.* **27**:449–470.
- Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**: 3326–3328.
- Zhou Y, Fan W, Zhang H, Zhang J, Zhang G, Wang D, Xiang G, Zhao C, Li L, He S, *et al.* 2023. The genome of *Marsdenia tenacissima* provides insights into calcium adaptation and tenacissoside biosynthesis. *Plant J.* **113**:1146–1159.
- Zhu T, Wang L, You FM, Rodriguez JC, Deal KR, Chen L, Li J, Chakraborty S, Balan B, Jiang CZ, *et al.* 2019. Sequencing a *Juglans regia* × *J. microcarpa* hybrid yields high-quality genome assemblies of parental species. *Hortic Res.* **6**:55.