




Gene flow, linked selection, and divergent sorting of ancient polymorphism shape genomic divergence landscape in a group of edaphic specialists

Fushi Ke¹  | Liette Vasseur² | Huiqin Yi^{1,3} | Lihua Yang¹ | Xiao Wei⁴ |
Baosheng Wang^{1,5}  | Ming Kang^{1,5} 

¹Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China

²Department of Biological Sciences, Brock University, St. Catharines, Ontario, Canada

³University of Chinese Academy of Sciences, Beijing, China

⁴Guangxi Institute of Botany, Guangxi Zhuang Autonomous Region and the Chinese Academy of Sciences, Guilin, China

⁵Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Guangzhou, China

Correspondence

Ming Kang, Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China.
Email: mingkang@scbg.ac.cn

Funding information

National Natural Science Foundation of China, Grant/Award Number: 32170237

Abstract

Interpreting the formation of genomic variation landscape, especially genomic regions with elevated differentiation (i.e. islands), is fundamental to a better understanding of the genomic consequences of adaptation and speciation. Edaphic islands provide excellent systems for understanding the interplay of gene flow and selection in driving population divergence and speciation. However, discerning the relative contribution of these factors that modify patterns of genomic variation remains difficult. We analysed 132 genomes from five recently divergent species in *Primulina* genus, with four species distributed in Karst limestone habitats and the fifth one growing in Danxia habitats. We demonstrated that both gene flow and linked selection have contributed to genome-wide variation landscape, where genomic regions with elevated differentiation (i.e., islands) were largely derived by divergent sorting of ancient polymorphism. Specifically, we identified several lineage-specific genomic islands that might have facilitated adaptation of *P. suichuanensis* to Danxia habitats. Our study is amongst the first cases disentangling evolutionary processes that shape genomic variation of plant specialists, and demonstrates the important role of ancient polymorphism in the formation of genomic islands that potentially mediate adaptation and speciation of endemic plants in special soil habitats.

KEYWORDS

ancient polymorphism, gene flow, genomic divergence landscape, linked selection, *Primulina*

1 | INTRODUCTION

Mounting evidence indicates that natural selection can drive population divergence and lead to speciation with gene flow (Coyne & Orr, 2004; Via, 2009). Peaks of differentiation (i.e., genomic regions with elevated differentiation [F_{ST}]), initially referred to as “islands of speciation” (Turner et al., 2005), are regions that serve as barriers to gene flow, and regions between islands are observed with reduced differentiation due to homogenizing process of gene flow (Feder et al., 2012; Nosil et al., 2009). However, genomic islands can also

emerge through processes unrelated to adaptation and speciation (Cruickshank & Hahn, 2014; Guerrero & Hahn, 2017; Ravinet et al., 2017; Wolf & Ellegren, 2017). Distinguishing the evolutionary processes that contribute to genomic divergence landscape thus is vital in understanding speciation and formation of biodiversity.

Modern sequencing technologies have facilitated our understanding of species divergence from a genomic perspective (Han et al., 2017; Vijay et al., 2016; Wang et al., 2016; Wolf & Ellegren, 2017). By investigating genome-wide variation and differentiation, a highly heterogenous genomic landscape governed by multiple

evolutionary processes can be discovered (Ravinet et al., 2017; Vijay et al., 2016; Wolf & Ellegren, 2017). Linked selection, including background selection and recurrent selective sweeps, can accelerate lineage sorting and increase genetic differentiation between extant species (Burri, 2017). Meanwhile, genomic regions with features such as low recombination and high gene density may affect the efficiency and extent of selection (Burri, 2017). Demographic history and gene flow further shape variation and differentiation by interplaying with genomic features and selection (Ravinet et al., 2017; Wolf & Ellegren, 2017). Without gene flow, selection can reduce genetic diversity and hence generate higher F_{ST} . It does not, however, impact nucleotide distance (d_{XY}) in these regions (Cruickshank & Hahn, 2014; Irwin et al., 2018). However, gene flow coupling with selection is predestined to evolve towards higher nucleotide divergence in genomic islands compared with other regions that can flow more freely (Irwin et al., 2018). The genomic islands with elevated nucleotide divergence (i.e., d_{XY}) can also be formed without the existence of gene flow by divergent lineage sorting of ancient polymorphism (Guerrero & Hahn, 2017). Highly diverged haplotypes are maintained in genomic regions with increased nucleotide diversity (π) and low relative divergence (F_{ST}) between nascent species by evolutionary process such as balancing selection, and form islands of divergence due to divergent lineage sorting of haplotypes in descendant species under selection (Guerrero & Hahn, 2017; Wang et al., 2019).

Disentangling the evolutionary processes that influence divergence landscape is thus difficult. A powerful way to make inferences about these processes is to analyse genomic variation of evolving lineages at different levels of divergence with detailed information of demographic history (Wolf & Ellegren, 2017). Employing multiple evolving lineages (i.e., lineage replications) is a potent way in understanding evolutionary forces that occur repeatedly (Wolf & Ellegren, 2017). In addition, correlation analysis between genetic variation and genomic features, and characterization of genomic islands are efficient to separate different evolutionary processes. For instance, correlations between genomic variation and factors affecting linked selection can prevail across the genome (Burri, 2017; Rettelbach et al., 2019). Meanwhile, characteristics of genomic islands across multiple lineages are informative in understanding the role of gene flow and type of selection that contribute to their formation (Burri, 2017; Cruickshank & Hahn, 2014; Han et al., 2017; Irwin et al., 2018; Ravinet et al., 2017; Wang et al., 2016, 2019; Wolf & Ellegren, 2017).

Removing the effects of processes that are not related to adaptation and speciation per se (e.g., background selection) is important when linking genomic islands to ecological adaptation. Booker et al. (2020) indicate genomic scan methods that do not take local recombination rate into consideration can increase the false positives. Recently, quantitative evolutionary controls based on lineage replications have been used to isolate lineage-specific genomic islands (Burri, 2017; Roesti et al., 2012). Berner and Salzburger (2015) suggest that delta differentiation (ΔF_{ST}) is useful in accounting for differentiation patterns resulting from background selection in genomic features by adjusting observed differentiation across multiple

comparisons. In addition, estimates such as population branch statistics (PBS, Yi et al., 2010), unlike the pairwise nature of F_{ST} , can be useful in isolating lineage-specific genomic islands.

Adaptation of plants to extreme edaphic physicochemical environments is an iconic example of strong natural selection (Brady et al., 2005; Rajakaruna, 2003). Karst and Danxia landscapes in southern China and Southeast Asia cover more than 800,000 km². They are characterized by diverse and extreme environmental conditions (Day & Urich, 2000). The typical soils of these habitats are surrounded by acid soils and highly heterogeneously distributed forming soil-islands (Hao et al., 2015). Limestone soils in Karst are shallow and contained high levels of Ca and Mg (Hao et al., 2015; Nie et al., 2011; Zhang et al., 2011). They differ in contents of soil chemicals and form locally contrasting mosaics of edaphic habitats (Hao et al., 2015). Danxia soils derive from red sediments, and are differentiated from Karst typical acid soils in contents of chemicals such as Ca, K, Fe, Mg and P (Hao et al., 2015). Both Karst and Danxia soils are highly porous with low water storage capacities and prone to chronic drought (Hao et al., 2015). These physical and chemical characteristics may have contributed to the remarkably high endemism and species richness in Karst and Danxia areas. The Karst and Danxia of Southeast Asia have been proposed as a world biodiversity hotspot (Clements et al., 2006). They have been recognized as a center of plant diversity and speciation (Davis et al. 1995), representing an important natural environment in which to study terrestrial plant speciation. However, the actual evolutionary mechanisms that drive plant diversification on Karst and Danxia landforms remain poorly understood.

The genus *Primulina* represents a group with about 200 plant species distributed across southern China and northern Vietnam (Xu et al., 2021). This plant taxon exhibits typical edaphic specialization to Karst and Danxia's soils. More than 90% of its species are distributed in one type of soil habitat, while a few other species grow on two types but tend to be found in a single habitat (Hao et al., 2015). Seeds of these species are small (>1 mm), lack of hooks or pappi, and can disperse through water. Most *Primulina* species are outcrossed and pollinated by insects (Gao et al., 2015). In this study, we focused on five closely related species (*P. hunanensis*, *P. polycephala*, *P. versicolor*, *P. alutacea* and *P. suichuanensis*) of this genus. These species can be distinguished from each other by flower characteristics (Cai et al., 2015; Pan et al., 2016; Zhou et al., 2016). The first four species grow only in Karst landforms and bloom from May to August. The last species, *P. suichuanensis*, is only found in Danxia landforms where it flowers from September to November. Each species has a narrow range, growing in specific microhabitats, with some of them being geographically closer to each other than others (Figure 1a). Their heterogeneous distribution thus may lead to varied intensity of gene flow among different species pairs. Previous population genetic analyses in *Primulina* have identified pronounced genetic structure as a result of genetic drift and adaptation (Wang et al., 2017). Based on nuclear polymorphism, Wang, Ai, et al. (2017) results further show multiple levels of divergence between *P. polycephala*, *P. alutacea*, and *P. suichuanensis*. Prevalent although weak historical gene flow has

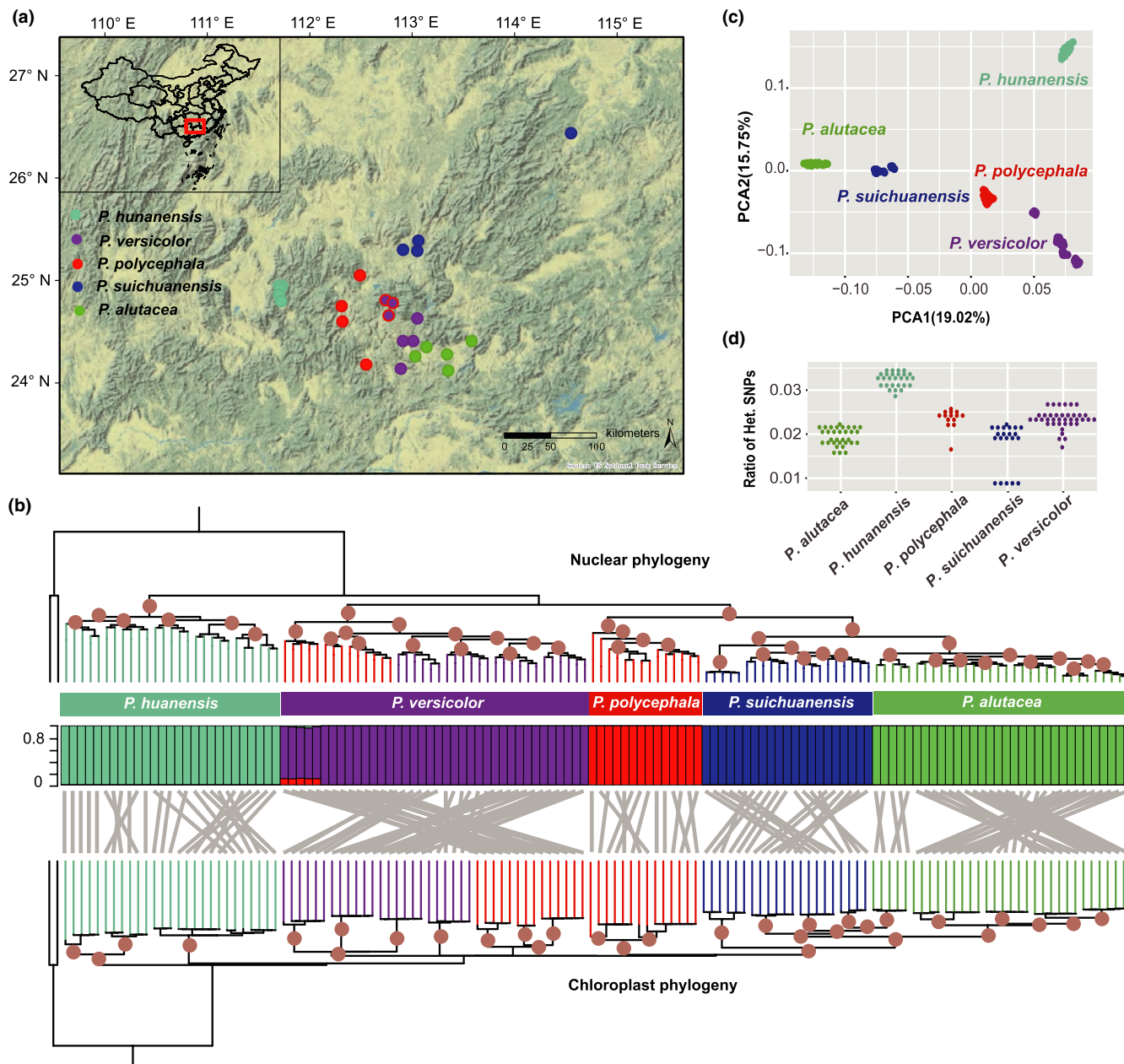


FIGURE 1 Phylogenetic relationships, population structure, and genetic diversity of five *Primulina* species. (a) Geographic locations of sampling sites. The purple points with red circles indicate samples identified as *P. polycephala* based on traditional taxonomy but show genetic similarity with *P. versicolor* based on nuclear polymorphisms. (b) Phylogenies and genetic structure. Maximum-likelihood (ML) nuclear and chloroplast phylogenies were inferred based on concatenated SNPs with 1000 bootstraps, and the discordance between phylogenies are shown with grey lines. The brown points at each node indicate bootstrap values of 100. Branches are coloured according to traditional taxonomy. Coloured columns inferred the ancestry proportion of each sample, according to fastStructure analysis at $K = 5$. See Figure S2 for $K = 2-5$ results of ADMIXTURE, fastStructure and sNMF. (c) Principal component analysis (PCA) based on the first two components. (d) Heterozygosity of all 132 individuals, by calculating the ratio of heterozygous to homozygous SNPs [Colour figure can be viewed at wileyonlinelibrary.com]

been identified among populations/species of *Primulina* (Gao et al., 2015; Wang, Ai, et al., 2017; Wang et al., 2017) and indicate the possibility of gene flow across the island-like habitats. These features make this group an ideal system to investigate different processes contributing to genomic differentiation landscape.

Here, we used resequencing data of 132 individuals to investigate genomic differentiation landscape of species at multiple levels of

divergence. First, we examined the genetic structure, phylogenetic relationship, and demographic history of this system, and characterized whether gene flow differed among lineages and contributed to genomic variation. In particular, we wanted to determine whether species pairs with high gene flow contributed to genome-wide variation and differentiation. We further examined processes that might shape genomic divergence landscape by correlation analysis and

characterization of genomic islands, and tested whether divergent selection was dominant in generating genomic islands. Finally, we took *P. suichuanensis* as an example, removed the effects of evolutionary forces unrelated to speciation to identify the genomic islands that could have facilitated adaptation to the Danxia habitat and phenological differentiation from its sibling species. Our aim was to disentangle processes that had shaped genomic differentiation of evolving species, and identify the genomic basis of adaptation and the formation of endemic terrestrial biodiversity.

2 | MATERIALS AND METHODS

2.1 | Population sampling, genome resequencing and variant calling

A total of 132 individuals including *P. hunanensis* (27), *P. versicolor* (24), *P. polycephala* (28), *P. alutacea* (32), and *P. suichuanensis* (21) were sampled across the Karst and Danxia landscapes in southern China (Figure 1a). In addition, two individuals of *P. huaijiensis* from South China were sampled as outgroup for phylogenetic analyses (Wang, Ai, et al., 2017). Total genomic DNA was extracted from dried leaves using a modified cetyl trimethylammonium bromide (CTAB) method (Doyle, 1987).

We prepared sequencing libraries with an insert size between 400–600 bp, and then sequenced on the Illumina HiSeq 2000 platform with paired-end 150 bp reads. Raw reads of each sequenced individual were filtered using in-house C scripts by removing reads with $\geq 10\%$ unidentified nucleotides, with $>50\%$ bases having phred quality < 5 , or >10 nucleotides aligned to the adapter, allowing $\leq 10\%$ mismatches. Duplicated reads were removed by using FastUniq (Xu et al., 2012). Clean reads were then aligned to the reference assembly of *P. eburnea* (accession number JAGGRQ000000000), a closely related species to *P. polycephala*, using the BWA-MEM algorithm in BWA v0.7.15 (Li, 2013). Picard (<http://broadinstitute.github.io/picard>) was used to mark PCR duplicates, and GATK v3.8 (DePristo et al., 2011) was employed to recalibrate sequences around indels. The final bam file of each individual was used to calculate coverage and sequencing depth using Bamdeal (<https://github.com/BGI-shenzhen/BamDeal/releases/tag/v0.22>). HaplotypeCaller was implemented to generate a gvcf file for each individual, the files were combined and genotyped by using CombineGVCFs and GenotypeGVCFs in GATK. Raw SNPs were filtered by VariantFiltration in GATK with default settings, and then filtered by VCFtools (Danecek et al., 2011) to remove loci with mean coverage depth lower than 3 or higher than 150.

We also generated variants of the chloroplast genome by mapping clean reads to the published chloroplast genome of *P. eburnea* (Feng et al., 2017). Bam files of every individual were generated using the same procedure of nuclear genome but mapping to chloroplast genome. Chloroplast SNPs of all individuals were generated following Scarcelli et al. (2016) by using a combination of SAMTOOLS 1.1 (Li et al., 2009) and VARSCAN v2.3.7 (Koboldt et al., 2012)

(SAMTOOLS-B/VARSCAN --min-var-freq 0.5 --min-freq-for-hom 0.5). Loci were filtered using --min-coverage 200 --min-reads 200 in VARSCAN to remove potential contamination of reads from the nuclear genome.

2.2 | Phylogenetic relationship and population genetic structure

We constructed phylogenetic trees using three different methods based on nuclear data. First, we generated a maximum likelihood (ML) tree using IQ-TREE (Nguyen et al., 2015) based on concatenated SNPs. Several nuclear SNP data sets thinned by intervals of 25, 100 and 200 bp were tested in ML tree construction. Second, we generated a phylogenetic tree using SVDquest (Vachaspati & Warnow, 2018) based on biallelic and unlinked SNPs. Bcftools (Danecek & McCarthy, 2017) was used to prune SNPs with a linkage disequilibrium (LD) value r^2 higher than 0.2 in an interval of 50 loci. Finally, we constructed a summary phylogeny of window-based gene trees using ASTRAL-III (Zhang et al., 2018). Windows with a length of 10 kb were isolated every 50 kb across the genome, of which those harbouring an SNP number lower than 200 or higher than 2,500 were removed from further analysis. A total of 4,436 windows pass the filtered thresholds and were successful in generating gene trees. A chloroplast phylogeny based on concatenated SNPs was also constructed using IQTREE. All of the phylogenies based on ML method were analysed in IQTREE with 1000 bootstraps (Hoang et al., 2018).

The genetic structure was explored using unlinked SNPs with a missing rate $\leq 20\%$ in the intergenic regions 20 kb away from the genic regions. Bcftools was used to prune SNPs based on LD value of 0.2 with an interval of 50 loci. We used three methods, ADMIXTURE (Alexander et al., 2009), FASTSTRUCTURE (Raj et al., 2014) and sNMF (Frichot et al., 2014) for population structure analyses. Each K from 1 to 20 was performed with 30 replicates in each software, and the best K value was identified based on the cross-entropy criterion. Furthermore, a principal component analysis (PCA) was conducted in plink (Chang et al., 2015) based on the same data set. For each cluster, we further estimated heterozygosity of each individual by calculating the ratio of heterozygous to homozygous SNPs.

2.3 | Gene flow and demographic history

Three individuals with high sequencing depth ($>25\times$) from each species were selected and used to infer historical effective population size using pairwise sequentially Markovian coalescence (PSMC) model (Li & Durbin, 2011). The generation time was set at 2 years, and the mutation rate (μ) was set as 7×10^{-9} per site per year following the estimates in *Arabidopsis thaliana* (Ossowski et al., 2010). Fluctuations in global temperature (Zachos et al., 2008) were displayed together with effective population sizes to investigate impacts of long-term climate change on population dynamics. We further employed SMC++ (Terhorst et al., 2017) to predict the divergence times of different species based on

multiple unphased individuals. The same generation time and mutation rate were used following PSMC analysis.

Genetic introgression among lineages were investigated by using ABBA-BABA test implemented in Dsuite (Malinsky et al., 2021). We further used partitioned D frequency spectrum (D_{FS}), in which D is split according to the frequencies of derived alleles, to verify particular gene flow event (Martin & Amos, 2020). We defined gene flow in species pair (i.e., P2 and P3 in ABBA-BABA analysis) with two significant D values and verified by D_{FS} as those with high level of gene flow, and categorized the others to the group with low level of gene flow. In addition, f_d statistics of species pairs with high and low levels of gene flow were calculated in nonoverlapping 100-kb windows using Python script downloaded from https://github.com/simonhmartin/genomics_general. Windows with positive D was retained for further comparison and correlation analysis. Note that D statistics were inflated in regions of low recombination or genomic inversion (Martin et al., 2015), and we employed this window-based parameter as a complementary evidence that support the relative lower differentiation (F_{ST}) in a lineage pair was with higher level of gene flow.

2.4 | Calculation of population-genetic parameters in genomic windows

To examine how the genome-wide patterns of diversity and divergence varied among taxa, we calculated nucleotide diversity (π), relative differentiation (F_{ST}), and absolute divergence (d_{XY}), in non-overlapping 100-kb windows using custom Python scripts downloaded from https://github.com/simonhmartin/genomics_general. To provide unbiased estimates of diversity and divergence, the invariant sites were incorporated into the calculation by dividing the number of pairwise differences (within and between lineages, respectively) by the total number of valid sites (variant and invariant) within a window. Windows with less than a minimum number of 4000 genotyped sites (variant and invariant) were removed from the analysis. We further calculated a corrected estimate of sequence divergence (d_a) of each window by subtracting mean polymorphism of two lineages (π) from the total divergence (i.e., d_{XY}) (Cruickshank & Hahn, 2014). The d_a statistic represents the divergent time T (in generations) as they can be transformed by the equation: $T = d_a / (2\mu)$, where μ is the mutation rate.

We used FASTPRR (Gao et al., 2016) to generate the population scaled recombination rate (r) per 100 kb window. We used all individuals of each species in this analysis. The r value was estimated for every 100 kb window with default parameters, and transformed into ρ (per bp) in every window. To obtain a measure of recombination that is independent of local N_e , we further adjusted ρ by dividing it to the lineage-specific genetic variation (Wang et al., 2016). We also calculated linkage disequilibrium (LD) by using VCFtools in the nonoverlapping 100 kb-window. Window-LD value was the average of each biallelic site in each window. Coding sequence percentage (gene density) was estimated as the percentage of coding sequence within each of the 100-kb windows.

We calculated the Pearson's correlation coefficient for intrapopulation summary statistics between populations (e.g., π_i vs. π_j), within populations (e.g., π_i vs. ρ_i), in comparison with interpopulation statistics (e.g., π_i vs. F_{STij}) and for interpopulation statistics of lineage pairs (e.g., F_{STij} vs. d_{XYij}). We also summarized the variation of each statistic (π , F_{ST} , d_{XY} and ρ) across comparisons using a PCA, and used the first principal component of each parameter to explain the greatest covariance (Stankowski et al., 2019). Pearson's correlation coefficients were also calculated for PC1 π , PC1 F_{ST} , PC1 d_{XY} , PC1 ρ , and gene density.

2.5 | Outlier detection

To compare genomic landscapes across different species pairs and identify the genomic regions with elevated differentiation (i.e., genomic island), we standardized per-window F_{ST} in each group pair to a Z-score (Han et al., 2017) based on the formula: $Z-F_{ST} = (\text{window } F_{ST} - \text{mean } F_{ST}) / \text{standard deviation } F_{ST}$. The windows with $Z-F_{ST} \geq 2$ from each group pair were treated as genomic islands (F_{ST} -island).

We further combined two metrics including ΔF_{ST} and population branch statistic (PBS) (Yi et al., 2010) to isolate lineage specific outlier across the genome of *P. suichuanensis*. The ΔF_{ST} value was calculated in nonoverlapping 100 kb window by the formula: $\Delta F_{ST} = F_{ST}(P. suichuanensis \text{ vs. } P. alutacea) - F_{ST}(\text{mean})$, in which $F_{ST}(\text{mean})$, as a baseline, is the mean F_{ST} value over the non-*P. suichuanensis* lineage pairs of each window. We also implemented PBS of *P. suichuanensis* to isolate genomic outliers in adaptive analysis. Unlike pairwise nature of traditional metrics (such as F_{ST}), PBS can distinguish adaptive regions of specific lineages by combining information from two closely related populations and an outgroup (Burri, 2017), and has strong power to detect selective sweeps under neutral demographic models (Yi et al., 2010). By choosing *P. polycephala*, which was adapted to Karst, as the outgroup in PBS calculation of *P. suichuanensis*, we were able to identify lineage-specific selection in our focal lineage. Genome-wide ΔF_{ST} and PBS were standardized per-window to a Z-score, and a filtered criterion of $Z\text{-PBS} > 2$ and $Z\text{-}\Delta F_{ST} > 2$ was used in isolation of genomic regions related to edaphic adaptation of *P. suichuanensis*. To test lineage-specific windows under positive selection, we calculated Tajima's D value of each window based on VCFtools. We used Mann-Whitney U test to compare differences between the target and background windows.

2.6 | Analysis of putatively adaptive loci

We extracted genes that were within genomic islands and used the genome annotation of *P. eburnea* as a reference. We performed functional enrichment analysis of GO by TBtools (Chen et al., 2020) to identify functional classes of genes that were overrepresented in the target regions. p -values of Fisher's exact test were calculated, and were corrected for multiple testing with Benjamini-Hochberg (B-H) method (Benjamini & Hochberg, 1995). GO term with a false

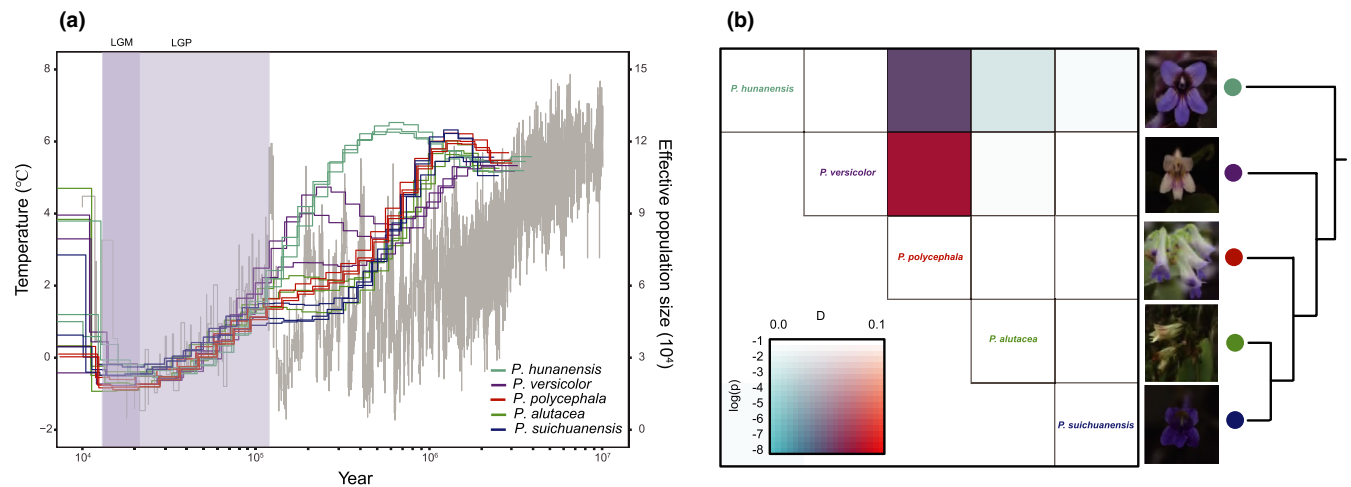


FIGURE 2 Demographic history of five *Primulina* species. (a) Chronological variation of effective population size in five species of *Primulina* inferred by pairwise sequentially Markovian coalescence (PSMC) model. The grey line represents past fluctuations of temperatures (data from Zachos et al., 2008). The shaded rectangles indicate the Last Glacial Maximum (LGM) and the Last Glacial Period (LGP). (b) Events of gene flow identified by Dsuite. Gene flow between *P. versicolor* and *P. polycephala*, and between *P. hunanensis* and *P. polycephala* identified by Dsuite were further supported by D_{F5} pattern (Figure S4) [Colour figure can be viewed at wileyonlinelibrary.com]

discovery rate (FDR) value < 0.05 after B-H correction was considered to be significantly enriched.

For loci located in genomic regions potentially related to adaptation of *P. suichuanensis* to Danxia soils, we generated haplotype network and gene tree individually using Popart (Leigh & Bryant, 2015) and IQTREE. Loci with conflicting topology in gene tree compared with species tree and extended branch length in target species were further used for protein modeling. Original peptide sequences (wild type) of each gene were used in protein modeling using Phyre2 (Kelley et al., 2015), and the effect of missense variants at a particular position was predicted using the SuSpect method (Yates et al., 2014). We also employed 13 transcriptomes (L. H. Yang et al., unpublished data) in Gesneriaceae and constructed gene tree of single loci under selection based on ML method in IQTREE.

3 | RESULTS

The whole genomes of 132 *P. polycephala* individuals were sequenced using Illumina Hiseq2000/2500 short-read technology. Raw reads were filtered and mapped onto the *Primulina eburnea* reference genome, which resulted in a mean depth of 27.5× (Table S1). SNPs were called using HaplotypeCaller in GATK and filtered by using stringent criteria (hard filter in GATK and VCFtools), which generated 54,354,111 high quality nuclear SNPs. In addition, a total of 3441 chloroplast SNPs were identified after filtering out potential contamination from nuclear genome.

3.1 | Population structure and demographic history

Nuclear SNPs used to infer individual ancestry of the sampled individuals identified five genetic clusters (Figure 1b; Figure S1). The

PCA supported this genetic pattern (Figure 1c; Figure S2). We further inferred phylogenetic relationships based on both nuclear and chloroplast SNPs (Figure 1b, Figure S3). Using *P. huaijensis* as the outgroup, we found a clear genetic clustering in nuclear phylogeny consistent with the genetic pattern identified by structure analysis (Figure 1b). However, the results based on nuclear polymorphisms were not consistent with traditional taxonomy. Chloroplast phylogeny showed a pattern of paraphyly for each lineage. For example, the *P. versicolor* lineage consisted of samples taxonomically identified as *P. versicolor* and *P. polycephala*, and exhibited a pattern of paraphyly in chloroplast phylogeny. However, a high genetic similarity was found within each cluster based on nuclear polymorphism (Figure 1b, c). We thus treated each identified lineage based on nuclear data as independent genetic units in our analyses. Further investigation of genome-wide variation based on the ratio of heterozygous SNPs showed mean values of each lineage at ~0.02–0.04 (Figure 1d).

PSMC analyses revealed that effective population sizes (N_e) of the five *Primulina* species varied consistently with global temperature change during the series of glacial and interglacial periods (Figure 2a). In particular, from the beginning of the Last Glacial Period (LGP, ~0.115 millions years [Myr]), the N_e dropped continuously, and reached its lowest value at the Last Glacial Maximum (LGM, ~11.7 thousand years [kyr]) (Figure 2a). This result suggested N_e was qualitatively matched with global temperature, and indicated the role of long-term climate changes on N_e fluctuations of these species in Karst and Danxia habitats. Based on SMC++ analysis, the split times among species ranged from 0.227 Myr (95% confidence interval: 0.208–0.245 Myr) between *P. suichuanensis* and *P. alutacea* to 0.414 Myr (95% confidence interval: 0.385–0.442 Myr) between *P. hunanensis* and *P. polycephala* (Figure S4). Gene flow were identified among three closely distributed lineages (*P. hunanensis*, *P. versicolor* and *P. polycephala*), and between *P. hunanensis* and *P. alutacea*,

based on D statistic in Dsuite (Figure 2b; Table S2). The species pair (*P. versicolor* and *P. polycephala*) with the highest D value also showed discordance in nuclear and chloroplast phylogeny. We further divided D into the partitioned frequency of derived alleles (D_{FS}) to verify the gene flow events. The D_{FS} pattern further support gene flows between *P. hunanensis* and *P. polycephala*, and between *P. versicolor* and *P. polycephala* (Figure S5).

3.2 | Effects of gene flow on genomic variation and differentiation

As lineages are characterized by similar effective population sizes and divergent times but different levels of gene flow, we expected that lineage pairs with higher gene flows would generally have reduced F_{ST} compared with those of lower gene flows. We calculated D statistics, a convenient measure of excess shared variation consistent with introgression, based on ABBA-BABA method implemented in Dsuite (Table S2), and categorized gene flow as either low or high (Table S3). Lineage pairs with two significant D values and verified by D_{FS} pattern (i.e., *P. versicolor* vs. *P. polycephala* and *P. hunanensis* vs. *P. polycephala*; Figure S5) were treated as "high", and the others were in "low". Note that we categorized *P. hunanensis* versus *P. alutacea* as the "low" group because this pair possessed only one significant D value and was not supported by the D_{FS} pattern. Consistent with evidence of gene flow based on D -related statistic (Figure 2b, Figure S5, Table S2), genome wide F_{ST} of *P. hunanensis* versus *P. versicolor* and *P. versicolor* versus *P. polycephala* (with high gene flows) were generally lower than other lineage pairs with low gene flow (Figure S6). In addition, each of the 18 chromosomes independently showed the same trend of lower F_{ST} in lineage pairs with higher levels of gene flow than in lineage pairs with lower levels of gene flow (Figure S6). The lineage pair of *P. suichuanensis* versus *P. alutacea* was in the category of low gene flow and exhibited low F_{ST} (Figure S6), of which the smaller differentiation is probably due to a recent divergence in this lineage pair. We could not, however, exclude the possibility of gene flow between *P. suichuanensis* versus *P. alutacea*, because the ABBA-BABA method could not estimate gene flow between these sister species.

The influence of gene flow on genomic variation and differentiation was then further investigated by calculating window f_d values across the genome. If gene flow left a footprint on genomic variation, we expected that windows with higher gene flow would have higher genetic variation (π) and lower relative population differentiation (F_{ST}). Comparisons were conducted for lineage pairs at two levels of gene flow, with lineage pairs of high (*P. versicolor* to *P. polycephala*) and low (*P. hunanensis* to *P. alutacea*) gene flow. We divided genomic windows of 100 kb into low, medium, and high introgression levels, and found a general pattern that higher π and low F_{ST} (*P. versicolor* vs. *P. polycephala*) in windows with higher introgression from *P. versicolor* (Figure S7a, b). In *P. hunanensis* versus *P. alutacea* lineage pair, no similar effects of introgression on genome-wide π and F_{ST} were observed (Figure S7c, d).

3.3 | Correlation analysis of genome-wide variation and genomic features

We investigated genome-wide variation structuring based on π , F_{ST} , absolute sequence divergence (d_{XY}), gene density (CodingP), linkage disequilibrium (LD), and population scaled recombination rate (ρ) calculated in nonoverlapping 100-kb windows by considering LD values across the genome. First, positive correlations between all possible lineage pairs were found for π , ρ , and LD (Figure S8). In addition, π was found to be negatively correlated with CodingP, but positively correlated with ρ of each population (Figure S8). We further identified that F_{ST} was negatively correlated with π and ρ , and positively correlated with d_{XY} . As the pattern of correlations among each statistic pair was similar across all comparisons, we then extracted principal component one (PC1) of each summary statistic to summarize and visualize the common pattern across multiple lineages. Indeed, the genome-wide pattern of these summary statistics were distributed nonrandomly across the genome and in each individual chromosome (Figure 3a). Correlation analysis among PC1 of each summary statistic showed qualitatively similar pattern with those based on single lineage pair (Figure 3b).

As we found positive correlations between genome-wide F_{ST} and d_{XY} (Figure 3b; Figure S8), and significant higher d_{XY} in genomic islands (i.e., higher F_{ST} , see below), but a trend of negative relationship between these two parameters (i.e., Z-PC1 F_{ST} and Z-PC1 d_{XY} ; Figure 3a), we suspected that this conflict might reflect differentiation of evolutionary forces working on genomic islands and genomic backgrounds ($Z-F_{ST} < 2$). We then removed the genomic islands and redid the correlation analysis. All of the correlations were consistent with previous analysis based on genome-wide windows, except those between F_{ST} and d_{XY} . We found six out of 10 species pairs that had significantly negative correlations. When windows were restricted to those with $Z-F_{ST} < 1$, eight out of 10 species pairs showed significantly negative correlations, with the other two presenting significant but weak positive correlations.

3.4 | Characterization of genomic islands

We further isolated genomic islands ($Z-F_{ST} \geq 2$) from each lineage pair, and investigated the characteristics of these genomic islands. Overall, we found a pattern of heterogenous distribution of genomic islands (100 kb nonoverlapping window, $Z-F_{ST} > 2$) among different lineage pairs (Figure S9), with only nine genomic islands (0.9 M) found in all of the 10 lineage pairs. These nine common genomic islands were characterized with lower recombination rate compared with the genomic background ($\rho_{\text{common}} = 0.00215 \pm 0.00117$ vs. $\rho_{\text{background}} = 0.00320 \pm 0.00112$). The total size of genomic islands varied from 26.9 M in lineage pair *P. polycephala* and *P. hunanensis* (genome-wide $F_{ST} = 0.197$) to 10.5 M in *P. versicolor* and *P. alutacea* (genome-wide $F_{ST} = 0.230$) (Table S3). The size and number of genomic islands were similar between species pairs with high and low levels of gene flow

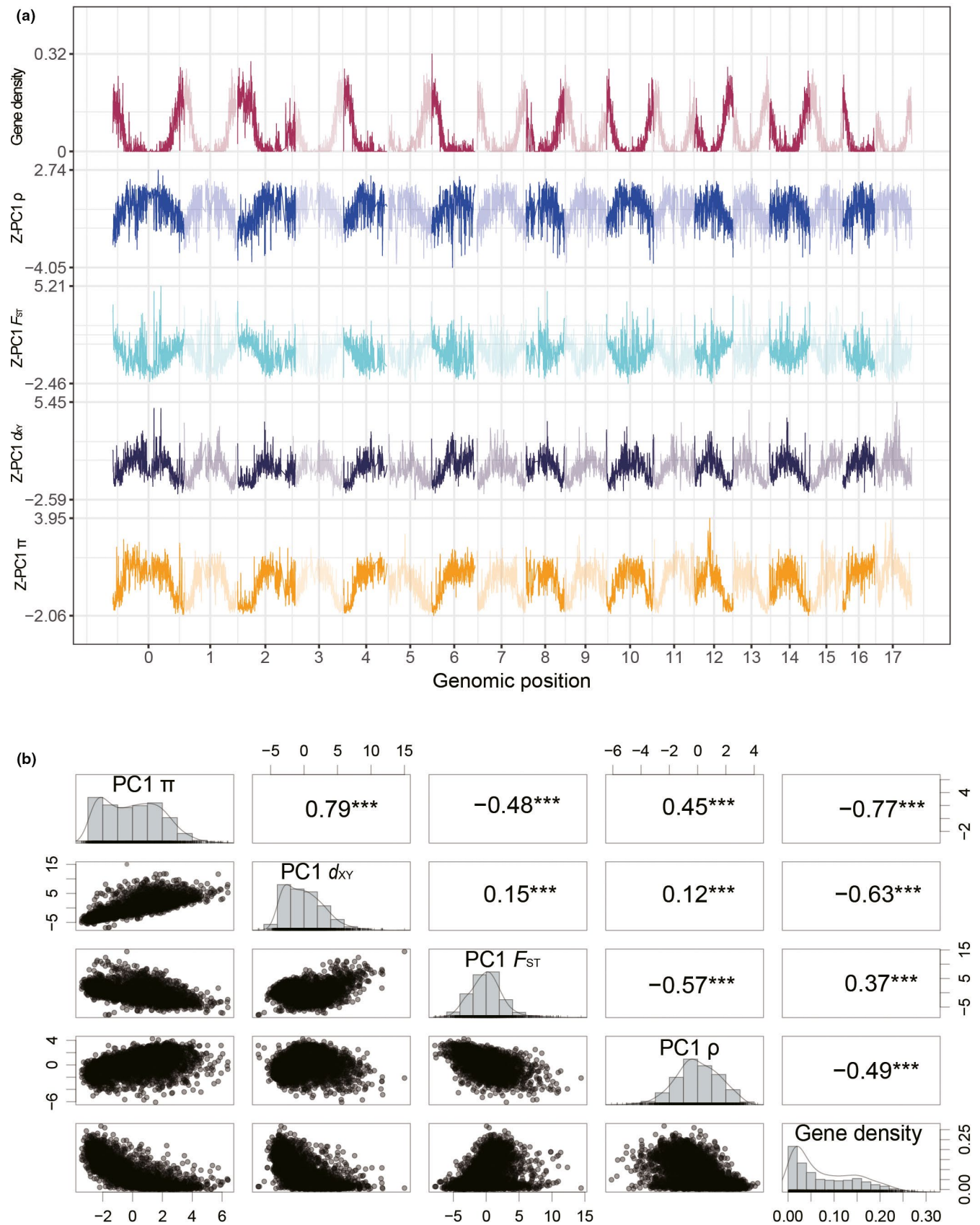


FIGURE 3 Landscape of genomic variation and structural properties. (a) Common genomic landscapes mirror variation in the local properties of the genome. Line charts of gene density, Z-transformed PC1 for recombination rate (ρ), relative differentiation (F_{ST}), absolute divergence (d_{XY}) and genetic variation (π) in nonoverlapping 100-kb genomic windows. (b) Correlations between genetic diversity and intrinsic features across the genome. Matrix of pairwise correlations between PC1 π , PC1 d_{XY} , PC1 F_{ST} , PC1 ρ , and gene density were calculated in nonoverlapping 100 kb genomic windows. *** indicates a p -value $< .001$ [Colour figure can be viewed at wileyonlinelibrary.com]

(Table S3). Comparison of population genetic parameters revealed significantly higher absolute divergence (d_{XY}) and linkage disequilibrium (LD) in the genomic islands than background windows in all lineage comparisons (Figure S10a, b, c). For each lineage pair, not all genomic islands showed significantly lower nucleotide diversity (π) compared with genomic background in each of the two species (Figure S10d). We also found not all genomic islands of each lineage pair enriched in functional elements (Coding sequence percentage) or genomic regions with a low recombination rate (ρ) (Figure S10e, f). In addition, a strong positive correlation between F_{ST} and d_{XY} across these islands was observed, with larger and smaller values indicating deeper and shallower cross-species coalescence, respectively (Figure S10g), indicating accumulation of genetic differentiation over time in these genomic islands. The average d_a values of these genomic islands were higher than all of the background windows (Figure S10h). Using a generation time of two years and a mutation rate of 7×10^{-9} per site per year, the average coalescent time of genomic islands in *P. suichuanensis* versus *P. alutacea*, which was the minimum value across ten species pairs (Figure S10h), is around 5.4 Myr.

3.5 | Adaptation of *P. suichuanensis* to Danxia habitats

To identify genomic regions that facilitated adaptation of *P. suichuanensis* to Danxia habitats as well as phenological differentiation with sibling species, we used ΔF_{ST} to identify genomic outliers in the target population (i.e., *P. suichuanensis* that adapted to Danxia). ΔF_{ST} can adjust observed differentiation across multiple comparisons (Burri, 2017; Vijay et al., 2016) and separate genomic islands from those resulting from selection pressures shared across multiple species (e.g., regions with low recombination rate, Booker et al., 2020). In addition, population branch statistic (PBS) of *P. suichuanensis* was used to identify lineage-specific selective sweeps in this species. Window-based analysis of these two parameters showed a significant correlation (Pearson's $r = 0.44$) (Figure 4a). Under the criterion of $Z\text{-PBS} > 2$ and $Z\text{-}\Delta F_{ST} > 2$, we identified 29 islands across the genome. These target regions had higher recombination rate than the background windows ($\rho_{\text{target}} = 0.0319 \pm 0.0696$ vs. $\rho_{\text{background}} = 0.00954 \pm 0.0338$), which indicated that they were unlikely connected with genomic features and background selection. In addition, the lineage-specific islands had significantly higher absolute divergence ($p = 2.2 \times 10^{-16}$) and linkage disequilibrium ($p = .0014$) when compared with the background windows (Figure 4b), and could be generated by selective sweeps (Jacobs et al., 2016). Species-specific windows with significantly more negative Tajima's D values than other windows ($p = .0016$) further supported positive selection in these windows (Figure 4b). Functional analysis of these genomic regions identified one significantly enriched GO term, which included four flavin-containing monooxygenases (FMOs) and one cytochrome P450 (CYP86A1) that related to monooxygenase activity ($p = .000627$, B-H method, Table S4). We also isolated five genes in

E3 ubiquitin ligase gene family and several transmembrane proteins in these regions (Table S4).

For the region with the highest values of PBS and ΔF_{ST} across the genome (Figure 4a), we investigated the potential effects of nonsynonymous mutation on functions of each locus in this region by using SuSpect method (Yates et al., 2014). However, we did not identify any missense mutation that significantly altered molecular function based on the simulation analysis. We further calculated ΔF_{ST} and PBS in nonoverlapping 10 kb windows at this region, and identified a 10 kb window containing the gene *evm.model.ctg621.5* (translation initiation factor IF-3, *eIF3*) with the highest value. Apart from discordant tree topology and extended branch length, haplotypes of this gene in *P. suichuanensis* differed from the other lineages (Figure 4c). *eIF3* is important in start codon recognition in both prokaryotes and eukaryotes and influences translational control in eukaryotic cells (Kearse & Wilusz, 2017; Valášek et al., 2017). A close investigation of mutations at this gene identified a mutation turning initiation codon from AUG to CUG. As well as *P. suichuanensis*, this mutation (allele) was not present in the other species. The gene tree based on transcriptomic data demonstrated that all species had the AUG initiation codon at this gene, showing a de novo mutation in *P. suichuanensis* (Figure 4d).

4 | DISCUSSION

Our study represents a comprehensive empirical analysis aimed at detangling the evolutionary processes that shaped the patterns of genomic variation in a group of edaphic plant specialists. We found that gene flow and long-term linked selection contributed to genome wide variation and thus the formation of differentiation in these species. In spite of gene flow and widespread linked selection, majority of genomic regions with elevated differentiation was caused by divergent sorting of ancestral polymorphism. Teasing apart processes unrelated to lineage specific adaptation (i.e., background selection) by incorporating different genetic parameters, we were able to identify genomic outliers that could have facilitated the ecological adaptation of *P. suichuanensis* to Danxia habitats.

4.1 | Gene flow contributed to genomic variation and differentiation

Despite current geographical isolation between species, our ABBA-BABA analysis demonstrated high levels of gene flow among three geographically close lineages (i.e., *P. versicolor* vs. *P. polycephala* and *P. hunanensis* vs. *P. polycephala*) and limited gene flow among other geographically isolated lineage pairs. Similarly, previous analyses based on chloroplast DNA and microsatellite markers on *Primulina* have supported widespread historical gene flow among these species despite their distribution in soil-island outcrops (Hao et al., 2015; Wang, Ai, et al., 2017; Wang, Feng, et al., 2017). Given the strong interspecies genetic differentiation and limited seed dispersal

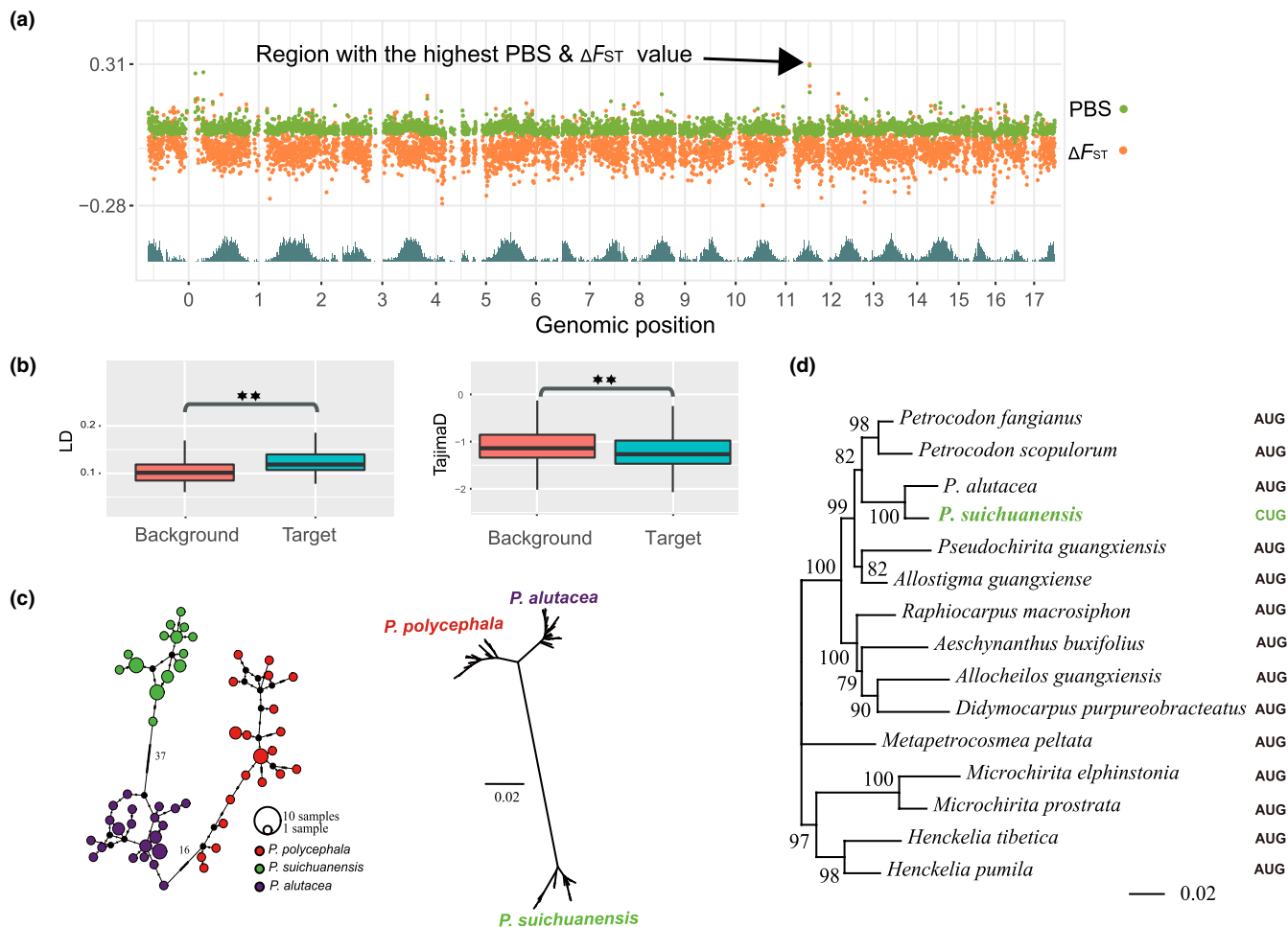


FIGURE 4 Genome-wide differentiation mirror ecological adaptation of *P. suichuanensis* to Danxia landform. (a) Genome-wide net genetic differentiation (ΔF_{ST}) and population branch statistic (PBS) values calculated in 100-kb nonoverlapping windows. *P. polycephala* and *P. alutacea* were used as the outgroup and ingroup in PBS calculation of *P. suichuanensis*. The lower panel is the bar plot of gene density in 100-kb nonoverlapping windows. (b) Linkage disequilibrium (LD) and Tajima's D values between genomic islands of divergence ($Z\text{-}\Delta F_{ST} > 2$ & $Z\text{-PBS} > 2$, named target) and the genomic background (other windows, named background). Significant differences are indicated with $* < .05$, $** < .01$, and $*** < .001$. (c) Haplotype network and gene tree of translation initiation factor IF-3 (*eIF3*). (d) Phylogeny shows the de novo mutation in initiation codon of *eIF3* (AUG to CUG) in *P. suichuanensis*. Coding sequence of *eIF3* isolated from 13 unpublished transcriptome were used [Colour figure can be viewed at wileyonlinelibrary.com]

capability, it seems unlikely that there is contemporary gene flow between species. One explanation for historical gene flow among species may be that their ancestors had much wider range than their current distributions and allowed gene flow at the beginning of species differentiation. Then the climatic fluctuations caused by extensive glaciation drove the ancestral populations into different Karst habitats as refuges. Indeed, the fluctuation in effective population size during alternating series of glacial and interglacial periods could have been associated with range shift that might have led to gene flows among three geographically adjacent lineages.

We further explored the effects of gene flow on genomic variation and differentiation by dividing lineage pairs into two groups based on introgression levels identified in ABBA-BABA analysis. Lineage pairs with high gene flow generally showed lower genome-wide differentiation (F_{ST}) compared with lineage pairs in the low-level group. This pattern was consistent across the genome and within each individual chromosome, indicating effects of gene flow on genome-wide

variation and differentiation (Martin et al., 2013; Via, 2009). In addition, the window-based analysis of f_d showed that the genomic windows with low and high introgression have high and low F_{ST} , respectively. This pattern, however, could not be found in the species pairs with low level of gene flow (e.g., *P. hunanensis* vs. *P. alutacea*). We note that the window-based D statistic may generate false positives in regions with low recombination rate (Martin et al., 2015). Nevertheless, combining genome-wide effects of gene flow on genomic differentiation suggested that our hypothesis on the influence of gene flow on genomic variation and differentiation was not violated.

4.2 | Divergent sorting of ancient polymorphism dominating the formation of genomic islands

For plants that adapted to special habitats (e.g., Karst landforms), an intriguing open question concerns the formation of genomic islands in

different species (Konečná et al., 2020). Our analysis of the genomic islands across 10 lineage pairs (of five species) showed significantly higher d_{xy} compared with other genomic windows and a global pattern of positive correlation between F_{ST} and d_{xy} in these genomic islands. Evolutionary models including both divergent selection with gene flow and divergent sorting of ancient polymorphism can generate genomic regions with both elevated F_{ST} and d_{xy} (e.g., Han et al., 2017; Ravinet et al., 2017). However, the divergent selection with gene flow may not be common in these lineages because we found gene flow only among three geographically adjacent species while no gene flow in the other species pairs based on D statistic. Meanwhile, if gene flow had contributed to the formation of genomic islands, the size and number of genomic islands with elevated d_{xy} in lineage pairs with gene flow should be higher than those without gene flow (Han et al., 2017). However, similar number and size of genomic islands between these two categories were identified (Table S4). Our results thus did not support the major role of gene flow in generating genomic islands.

An alternative scenario may be that most of the genomic islands are generated by divergent sorting of ancient polymorphism (Guerrero & Hahn, 2017). In this model, highly diverged haplotypes are maintained in genomic regions in ancestral species by evolutionary forces such as balancing selection and development of genomic islands due to lineage sorting of divergent haplotypes in the nascent lineages under geographically varied selection or drift (Guerrero & Hahn, 2017; Han et al., 2017; Wang et al., 2019). In this study, the five *Primulina* species inhabited spatially and temporally varying microhabitats and had high genetic diversity. These evolutionary processes (e.g., balancing selection) that harboured standing genetic variation might have prevailed across the genome. When species colonized a new habitat, local selection might have facilitated the formation of genomic islands by sorting of preadapted genetic variation (e.g., ancient divergent haplotypes). Unlike the model of divergent selection with gene flow where the genomic islands clustered together, the genomic islands heterogeneously distributed across the genome might have favoured formation of islands by ancient polymorphism. In addition, the genomic islands in all of the lineage pairs were with higher d_a value than that in background windows (Figure S10) and the average coalescent time between the divergent haplotypes of these genomic islands (>5.4 Myr) was higher than the split of these lineages (0.227–0.414 Myr) (Figure S4), suggesting most of the observed differentiation was due to ancestral polymorphism. We thus proposed divergent sorting of ancient polymorphism as the major mode in generating genomic islands. Nevertheless, we could not exclude the contribution of gene flow in generating genomic islands between geographically adjacent lineages.

4.3 | Linked selection affects the genomic variation and differentiation

Multiple evolutionary processes can generate islands across the genome (Cruickshank & Hahn, 2014; Guerrero & Hahn, 2017; Ravinet

et al., 2017; Vijay et al., 2016; Wolf & Ellegren, 2017). Processes including background selection cannot be related to adaptation or speciation but may have impacts on isolation of genomic islands in genome scans (Wolf & Ellegren, 2017). Accumulated effects of long-term background selection that repeatedly occur among lineages would interplay with genomic features (e.g., regions with low-recombination rate and high gene density) and contributed to the formation of genomic islands (Burri, 2017; Burri et al., 2015; Rettelbach et al., 2019). Analysis of population-scaled recombination rate in the genomic islands showed that around half of these lineage pairs were with reduced recombination rate in genomic islands. This feature indicated some of the genomic islands might be generated by linked selection in regions with low recombination rate and intrinsic genomic features (Burri, 2017; Cortés et al., 2018). Indeed, we found nine genomic islands with lower recombination rate than background windows commonly presented across 10 lineage pairs, which is probably generated by background selection that repeatedly occurred across these lineages.

Meanwhile, if the rate of mutations were high enough across the genome, linked selection would be a common evolutionary force in driving genetic variation and differentiation (Burri, 2017; Payseur & Nachman, 2002; Torres et al., 2020). The positive correlations between population recombination rates (and LD) among all species pairs did support a broad scale conserved population recombination rate and shared long-term linked selection across populations. In addition, the positive correlation between π and ρ , a negative correlation between π and functional elements, and a negative correlation between ρ and F_{ST} , suggest the pervasive effects of background selection (Burri, 2017; Payseur & Nachman, 2002; Vijay et al., 2016). When correlation analyses were done with only background windows, we further found that most species pairs showed negative correlations between F_{ST} and d_{xy} . Our results thus demonstrated widespread effects of linked selection on genomic variation and differentiation, especially on the genomic backgrounds. Moreover, a subset of genomic islands contributed through background selection (i.e., nine common genomic islands) even though the majority of the genomic islands were generated by divergent sorting of ancient polymorphism.

4.4 | Edaphic adaptation of *P. suichuanensis* to Danxia habitat

Colonization of novel habitats, including edaphic islands, can be achieved via ecological adaptation (Hereford, 2009; Palacio-López et al., 2015). Edaphic adaptation of ecotypes is a critical stage that can lead to speciation (Rajakaruna, 2018). Compared with its sibling species, *P. suichuanensis* is restricted to Danxia landform, a habitat different from the Karst soils in terms of chemical and physical properties (Hao et al., 2015). Its flowering time (from September to November) is later than the other closely related lineages (from May to August) (Cai et al., 2015; Zhou et al., 2016), possibly linking ecological adaptation to phenological adaptation (Osborne et al.,

2019). The differentiation in flowering time may be related to soil type pushing the development of late-flowering phenotypes of *P. suichuanensis* in Danxia while earlier flowering time in Karst. Similar patterns of divergence in phenology across edaphic habitats have been identified in several studies such as sympatric speciation of *Howea* palms (Osborne et al., 2019; Savolainen et al., 2006) and *Leptosiphon parviflorus* (Dittmar & Schemske, 2018).

When the effects of background selection were removed, we identified lineage-specific genomic regions that might have involved local adaptation of *P. suichuanensis* Danxia soil habitat. It should be noted that lineage-specific differentiation could also have resulted from demographic events, such as population expansion, and should be verified by more powerful methods such as supervised machine learning (e.g., Libbrecht & Noble, 2015; Schrider & Kern, 2016). We thus only focused on the region with the most distinguished signal of selective sweep, and on GO terms and genes that were overrepresented, instead of investigating every single gene in these regions.

In the region with the strongest genome-wide selection signal, we identified a fixed mutation turning initiation codon of translation factor IF-3 (*eIF3*) from AUG to CUG. *eIF3* plays an important role in start codon recognition in both prokaryotes and eukaryotes (Kearse & Wilusz, 2017), and coordinates the progress in most of the initiation steps (Gutu et al., 2013; Valášek et al., 2017). The change of AUG to CUG maintains *eIF3* generation under stress conditions (Kearse & Wilusz, 2017) which in turn controls the initiation of transcription and regulation of metabolism across the plant cells. The highly enrichment GO term of monooxygenase activity, which has been show to exhibit specific functions in auxin biosynthesis, metabolism of glucosinolates, and immunity-related activities in plants (Hartmann et al., 2018; Schlaich, 2007), may control flowering time under varying environmental conditions (Campos-Rivero et al., 2017; Cho et al., 2017). Meanwhile, CYP86A1, one in GO term of monooxygenase activity, is a key enzyme for biosynthesis of aliphatic root suberin that restricts water and nutrient loss, and prevents the invasion of pathogens in *Arabidopsis* (Höfer et al., 2008). In addition, five genes in E3 ubiquitin ligase gene family involved in the regulation of a number of biological processes, including plant innate immunity, plant reproduction, hormonal control of vegetative growth, light response, and stress tolerance (Craig et al., 2009; Mazzucotelli et al., 2006), were identified. Overall, these identified genomic regions might have facilitated the ecological adaptation of *P. suichuanensis* during its colonization of Danxia habitats. Further molecular evidence based on functional analysis of target genes should improve our understanding of speciation through ecological adaptation *P. suichuanensis*.

5 | CONCLUSIONS

Our analysis of genome-wide polymorphism supports differential gene flow in a group of *Primulina* species distributed in island-like

landforms of Karst and Danxia in China. We demonstrated that both gene flow and long-term linked selection have contributed to genome-wide variation in this system. Our analysis based on species pairs at multiple divergent levels support the important role of divergent sorting of ancient polymorphism in the formation of genomic islands. We further isolated lineage-specific islands in *P. suichuanensis*, some of which might contain loci facilitating local adaptation of this species to Danxia habitats. Our comprehensive analysis disentangles multiple evolutionary processes that shape genomic variation and differentiation of plant specialists, and clarifies the evolution and speciation of such endemic plants in South China Karst and Danxia.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (32170237) and the Strategic Priority Research Program of Chinese Academy of Sciences (XDB31000000). We also thank three anonymous reviewers for their insightful suggestions.

AUTHOR CONTRIBUTIONS

Ming Kang conceived the project and designed the study. Fushi Ke performed the data analysis, and generated figures and tables. Fushi Ke, Liette Vasseur, Baosheng Wang and Ming Kang wrote the manuscript. Huigin Yi, Lihua Yang and Xiao Wei contributed to sample collection, taxonomy, and DNA extraction. All authors read and approved the final manuscript.

DATA AVAILABILITY STATEMENT

This Whole Genome Shotgun project has been deposited at GenBank under the accession JAGGRQ000000000. All short reads have been deposited on NCBI's Sequence Read Archive (SRA) under accession number PRJNA717907.

ORCID

Fushi Ke  <https://orcid.org/0000-0002-5837-7260>

Baosheng Wang  <https://orcid.org/0000-0002-0934-1659>

Ming Kang  <https://orcid.org/0000-0002-4326-7210>

REFERENCES

- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19, 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological*, 57, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Berner, D., & Salzburger, W. (2015). The genomics of organismal diversification illuminated by adaptive radiations. *Trends of Genetics*, 31, 491–499. <https://doi.org/10.1016/j.tig.2015.07.002>
- Booker, T. R., Yeaman, S., & Whitlock, M. C. (2020). Variation in recombination rate affects detection of outliers in genome scans under neutrality. *Molecular Ecology*, 30, 1–25.
- Brady, K. U., Kruckeberg, A. R., & Bradshaw, H. D. (2005). Evolutionary ecology of plant adaptation to serpentine soils. *Annual Review of Ecology, Evolution, and Systematics*, 36, 243–266. <https://doi.org/10.1146/annurev.ecolsys.35.021103.105730>

- Burri, R. (2017). Interpreting differentiation landscapes in the light of long-term linked selection. *Evolution Letters*, 1, 118–131. <https://doi.org/10.1002/evl3.14>
- Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., Suh, A., Dutoit, L., Bures, S., Garamszegi, L. Z., Hogner, S., Moreno, J., Qvarnstrom, A., Ružič, M., Sæther, S.-A., Sætre, G.-P., Török, J., & Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula flycatchers*. *Genome Research*, 25, 1656–1665. <https://doi.org/10.1101/gr.196485.115>
- Cai, X. Z., Tian, J., Xiao, S. Y., Peng, L., & Liu, K. M. (2015). *Primulina hunanensis* sp. nov. (Gesneriaceae) from a limestone area in southern Hunan, China. *Nordic Journal of Botany*, 33, 576–581. <https://doi.org/10.1111/njb.00757>
- Campos-Rivero, G., Osorio-Montalvo, P., Sánchez-Borges, R., Us-Camas, R., Duarte-Aké, F., & De-la-Peña, C. (2017). Plant hormone signaling in flowering: An epigenetic point of view. *Journal of Plant Physiology*, 214, 16–27. <https://doi.org/10.1016/j.jplph.2017.03.018>
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, 4, s13715–s13742. <https://doi.org/10.1186/s13742-015-0047-8>
- Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., & Xia, R. (2020). *TPTools*—an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant*, 13, 1194–1202. <https://doi.org/10.1016/j.molp.2020.06.009>
- Cho, L., Yoon, J., & An, G. (2017). The control of flowering time by environmental factors. *The Plant Journal*, 90, 708–719. <https://doi.org/10.1111/tpj.13461>
- Clements, R., Sodhi, N. S., Schilthuizen, M., & Ng, P. K. L. (2006). Limestone karsts of southeast Asia: Imperiled arks of biodiversity. *BioScience*, 56, 733–742. [https://doi.org/10.1641/0006-3568\(2006\)56\[733:LKOSAI\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2006)56[733:LKOSAI]2.0.CO;2)
- Cortés, A. J., Skeen, P., Blair, M. W., & Chacón-Sánchez, M. I. (2018). Does the genomic landscape of species divergence in *Phaseolus* beans co-occur parallel signatures of adaptation and domestication? *Frontiers in Plant Science*, 9, 1816. <https://doi.org/10.3389/fpls.2018.01816>
- Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer Associates.
- Craig, A., Ewan, R., Mesmar, J., Gudipati, V., & Sadanandom, A. (2009). E3 ubiquitin ligases and plant innate immunity. *Journal of Experimental Botany*, 60, 1123–1132. <https://doi.org/10.1093/jxb/erp059>
- Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, 23, 3133–3157. <https://doi.org/10.1111/mec.12796>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Danecek, P., & McCarthy, S. A. (2017). BCFtools/csq: Haplotype-aware variant consequences. *Bioinformatics*, 33, 2037–2039. <https://doi.org/10.1093/bioinformatics/btx100>
- Davis, S. D., Heywood, V. H., & Hamilton, A. C. (1995). *Centres of plant diversity Vol. 2: Asia, Australasia and the Pacific*. WWF/IUCN, Gland.
- Day, M., & Ulrich, P. (2000). An assessment of protected karst landscapes in Southeast Asia. *Cave and Karst Science*, 27, 61–70.
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., Del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernysky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43, 491–498. <https://doi.org/10.1038/ng.806>
- Dittmar, E. L., & Schemske, D. W. (2018). The edaphic environment mediates flowering-time differentiation between adjacent populations of *Leptosiphon parviflorus*. *Journal of Heredity*, 109, 90–99. <https://doi.org/10.1093/jhered/esx090>
- Doyle, J. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, 19, 11–15.
- Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, 28, 342–350. <https://doi.org/10.1016/j.tig.2012.03.009>
- Feng, C., Xu, M., Feng, C., von Wettberg, E. J., & Kang, M. (2017). The complete chloroplast genome of *Primulina* and two novel strategies for development of high polymorphic loci for population genetic and phylogenetic studies. *BMC Evolutionary Biology*, 17, 224. <https://doi.org/10.1186/s12862-017-1067-z>
- Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, 196, 973–983. <https://doi.org/10.1534/genetics.113.160572>
- Gao, F., Ming, C., Hu, W., & Li, H. (2016). New software for the fast estimation of population recombination rates (FastEPRR) in the genomic era. *G3: Genes|genomes|genetics*, 6, 1563–1571. <https://doi.org/10.1534/g3.116.028233>
- Gao, Y., Ai, B., Kong, H., Kang, M., & Huang, H. (2015). Geographical pattern of isolation and diversification in Karst habitat islands: A case study in the *Primulina eburnea* complex. *Journal of Biogeography*, 11, 2131–2144. <https://doi.org/10.1111/jbi.12576>
- Guerrero, R. F., & Hahn, M. W. (2017). Speciation as a sieve for ancestral polymorphism. *Molecular Ecology*, 26, 5362–5368. <https://doi.org/10.1111/mec.14290>
- Gutu, A., Nesbit, A. D., Alverson, A. J., Palmer, J. D., & Kehoe, D. M. (2013). Unique role for translation initiation factor 3 in the light color regulation of photosynthetic gene expression. *Proceedings of the National Academy of Sciences*, 110, 16253–16258. <https://doi.org/10.1073/pnas.1306332110>
- Han, F., Lamichhaney, S., Grant, B. R., Grant, P. R., Andersson, L., & Webster, M. T. (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Research*, 27, 1004–1015. <https://doi.org/10.1101/gr.212522.116>
- Hao, Z., Kuang, Y., & Kang, M. (2015). Untangling the influence of phylogeny, soil and climate on leaf element concentrations in a biodiversity hotspot. *Functional Ecology*, 29, 165–176. <https://doi.org/10.1111/1365-2435.12344>
- Hartmann, M., Zeier, T., Bernsdorff, F., Reichel-Deland, V., Kim, D., Hohmann, M., Scholten, N., Schuck, S., Bräutigam, A., Hölzel, T., Ganter, C., & Zeier, J. (2018). Flavin monooxygenase-generated N-Hydroxy-pipecolic acid is a critical element of plant systemic immunity. *Cell*, 173, 456–469. <https://doi.org/10.1016/j.cell.2018.02.049>
- Hereford, J. (2009). A quantitative survey of local adaptation and fitness trade-offs. *The American Naturalist*, 173, 579–588. <https://doi.org/10.1086/597611>
- Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). *UFBoot2*: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35, 518–522. <https://doi.org/10.1093/molbev/msx281>
- Höfer, R., Briesen, I., Beck, M., Pinot, F., Schreiber, L., & Franke, R. (2008). The *Arabidopsis* cytochrome P450 CYP86A1 encodes a fatty acid ω -hydroxylase involved in suberin monomer biosynthesis. *Journal of Experimental Botany*, 59, 2347–2360. <https://doi.org/10.1093/jxb/ern101>
- Irwin, D. E., Milá, B., Toews, D. P. L., Brelsford, A., Kenyon, H. L., Porter, A. N., Gossen, C., Delmore, K. E., Alcaide, M., & Irwin, J. G. (2018). A comparison of genomic islands of differentiation across three young avian species pairs. *Molecular Ecology*, 27, 4839–4855. <https://doi.org/10.1111/mec.14858>

- Jacobs, G. S., Sluckin, T. J., & Kivisild, T. (2016). Refining the use of linkage disequilibrium as a robust signature of selective sweeps. *Genetics*, 203, 1807–1825. <https://doi.org/10.1534/genetics.115.185900>
- Kearse, M. G., & Wilusz, J. E. (2017). Non-AUG translation: A new start for protein synthesis in eukaryotes. *Genes & Development*, 31, 1717–1731. <https://doi.org/10.1101/gad.305250.117>
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10, 845–858. <https://doi.org/10.1038/nprot.2015.053>
- Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., Miller, C. A., Mardis, E. R., Ding, L., & Wilson, R. K. (2012). VARSCAN 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*, 22, 568–576. <https://doi.org/10.1101/gr.129684.111>
- Konečná, V., Yant, L., & Kolář, F. (2020). The evolutionary genomics of serpentine adaptation. *Frontiers in Plant Science*, 11, 574616. <https://doi.org/10.3389/fpls.2020.574616>
- Leigh, J. W., & Bryant, D. (2015). POPART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6, 1110–1116. <https://doi.org/10.1111/2041-210X.12410>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R., & 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, 1303.3997.
- Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475, 493–496. <https://doi.org/10.1038/nature10231>
- Libbrecht, M. W., & Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nature Review Genetics*, 16, 321–332. <https://doi.org/10.1038/nrg3920>
- Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite-fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*, 21, 584–595.
- Martin, S. H., & Amos, W. (2020). Signatures of introgression across the allele frequency spectrum. *Molecular Biology and Evolution*, 38, 716–726. <https://doi.org/10.1093/molbev/msaa239>
- Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., & Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, 23, 1817–1828.
- Martin, S. H., Davey, J. W., & Jiggins, C. D. (2015). Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Molecular Biology and Evolution*, 32, 244–257. <https://doi.org/10.1093/molbev/msu269>
- Mazzucotelli, E., Belloni, S., Marone, D., De Leonardis, A., Guerra, D., Di Fonzo, N., Cattivelli, L., & Mastrangelo, A. (2006). The e3 ubiquitin ligase gene family in plants: Regulation by degradation. *Current Genomics*, 7, 509–522.
- Nguyen, L., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32, 268–274. <https://doi.org/10.1093/molbev/msu300>
- Nie, Y., Chen, H., Wang, K., Tan, W., Deng, P., & Yang, J. (2011). Seasonal water use patterns of woody species growing on the continuous dolostone outcrops and nearby thin soils in subtropical China. *Plant and Soil*, 341, 399–412. <https://doi.org/10.1007/s11104-010-0653-2>
- Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, 18, 375–402. <https://doi.org/10.1111/j.1365-294X.2008.03946.x>
- Osborne, O. G., Ciezarek, A., Wilson, T., Crayn, D., Hutton, I., Baker, W. J., Turnbull, C. G. N., & Savolainen, V. (2019). Speciation in howea palms occurred in sympatry, was preceded by ancestral admixture, and was associated with edaphic and phenological adaptation. *Molecular Biology and Evolution*, 36, 2682–2697. <https://doi.org/10.1093/molbev/msz166>
- Ossowski, S., Schneeberger, K., Lucas-Lledó, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., Weigel, D., & Lynch, M. (2010). The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science*, 327, 92–94.
- Palacio-López, K., Beckage, B., Scheiner, S., & Molofsky, J. (2015). The ubiquity of phenotypic plasticity in plants: A synthesis. *Ecology and Evolution*, 5, 3389–3400. <https://doi.org/10.1002/ece3.1603>
- Pan, B., Wang, B., He, J., & Wen, F. (2016). *Primulina versicolor* and *P. alutacea* sp. nov. (Gesneriaceae), two new species with yellow flowers from northern guangdong, china. *Edinburgh Journal of Botany*, 73, 25–37. <https://doi.org/10.1017/S09600428615000268>
- Payseur, B. A., & Nachman, M. W. (2002). Gene density and human nucleotide polymorphism. *Molecular Biology and Evolution*, 19, 336–340. <https://doi.org/10.1093/oxfordjournals.molbev.a004086>
- Raj, A., Stephens, M., & Pritchard, J. K. (2014). FASTSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, 197, 573–589. <https://doi.org/10.1534/genetics.114.164350>
- Rajakaruna, N. (2003). Edaphic differentiation in *Lasthenia*: A model for studies in evolutionary ecology. *Madrono*, 34–40.
- Rajakaruna, N. (2018). Lessons on evolution from the study of edaphic specialization. *The Botanical Review*, 84, 39–78. <https://doi.org/10.1007/s12229-017-9193-2>
- Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlovi, M., Noor, M. A. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, 30, 1450–1477. <https://doi.org/10.1111/jeb.13047>
- Rettelbach, A., Nater, A., & Ellegren, H. (2019). How linked selection shapes the diversity landscape in *Ficedula flycatchers*. *Genetics*, 212, 277–285.
- Roesti, M., Hendry, A. P., Salzburger, W., & Berner, D. (2012). Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology*, 21, 2852–2862. <https://doi.org/10.1111/j.1365-294X.2012.05509.x>
- Savolainen, V., Anstett, M., Lexer, C., Hutton, I., Clarkson, J. J., Norup, M. V., Powell, M. P., Springate, D., Salamin, N., & Baker, W. J. (2006). Sympatric speciation in palms on an oceanic island. *Nature*, 441, 210–213. <https://doi.org/10.1038/nature04566>
- Scarcelli, N., Mariac, C., Couvreur, T., Faye, A., Richard, D., Sabot, F., Berthouly Salazar, C., & Vigouroux, Y. (2016). Intra-individual polymorphism in chloroplasts from NGS data: Where does it come from and how to handle it? *Molecular Ecology Resources*, 16, 434–445.
- Schlauch, N. L. (2007). Flavin-containing monooxygenases in plants: Looking beyond detox. *Trends in Plant Science*, 12, 412–418. <https://doi.org/10.1016/j.tplants.2007.08.009>
- Schrider, D. R., & Kern, A. D. (2016). S/HIC: Robust identification of soft and hard sweeps using machine learning. *PLoS Genetics*, 12, e1005928. <https://doi.org/10.1371/journal.pgen.1005928>
- Stankowski, S., Chase, M. A., Fuiten, A. M., Rodrigues, M. F., Ralph, P. L., & Streisfeld, M. A. (2019). Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers. *PLoS Biology*, 17, e3000391. <https://doi.org/10.1371/journal.pbio.3000391>
- Terhorst, J., Kamm, J. A., & Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics*, 49, 303–309. <https://doi.org/10.1038/ng.3748>
- Torres, R., Stetter, M. G., Hernandez, R. D., & Ross-Ibarra, J. (2020). The temporal dynamics of background selection in nonequilibrium populations. *Genetics*, 214, 1019–1030. <https://doi.org/10.1534/genetics.119.302892>

- Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, 3, e285. <https://doi.org/10.1371/journal.pbio.0030285>
- Vachaspati, P., & Warnow, T. (2018). SVDquest: Improving SVDquartets species tree estimation using exact optimization within a constrained search space. *Molecular Phylogenetics and Evolution*, 124, 122–136. <https://doi.org/10.1016/j.ympev.2018.03.006>
- Valášek, L. S., Zeman, J., Wagner, S., Beznosková, P., Pavlíková, Z., Mohammad, M. P., Hronová, V., Herrmannová, A., Hashem, Y., & Gunišová, S. (2017). Embraced by *elF3*: Structural and functional insights into the roles of *elF3* across the translation cycle. *Nucleic Acids Research*, 45, 10948–10968. <https://doi.org/10.1093/nar/gkx805>
- Via, S. (2009). Natural selection in action during speciation. *Proceedings of the National Academy of Sciences USA*, 106, 9939–9946. <https://doi.org/10.1073/pnas.0901397106>
- Vijay, N., Bossu, C. M., Poelstra, J. W., Weissensteiner, M. H., Suh, A., Kryukov, A. P., & Wolf, J. B. W. (2016). Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature Communications*, 7, 13195. <https://doi.org/10.1038/ncomms13195>
- Wang, B., Mojica, J. P., Perera, N., Lee, C. R., & Mitchell-Olds, T. (2019). Ancient polymorphisms contribute to genome-wide variation by long-term balancing selection and divergent sorting in *Boechera stricta*. *Genome Biology*, 20. <https://doi.org/10.1186/s13059-019-1729-9>
- Wang, J., Ai, B., Kong, H., & Kang, M. (2017). Speciation history of a species complex of *Primulina eburnea* (Gesneriaceae) from limestone Karsts of southern China, a biodiversity hot spot. *Evolutionary Applications*, 10, 919–934.
- Wang, J., Feng, C., Jiao, T., Von Wettberg, E. B., & Kang, M. (2017). Genomic signature of adaptive divergence despite strong non-adaptive forces on edaphic islands: A case study of *Primulina juliae*. *Genome Biology & Evolution*, 9, 3495–3508. <https://doi.org/10.1093/gbe/evx263>
- Wang, J., Street, N. R., Scofield, D. G., & Ingvarsson, P. K. (2016). Variation in linked selection and recombination drive genomic divergence during allopatric speciation of European and American aspens. *Molecular Biology and Evolution*, 33, 1754–1767. <https://doi.org/10.1093/molbev/msw051>
- Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, 18, 87–100. <https://doi.org/10.1038/nrg.2016.133>
- Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., Chen, J., & Chen, S. (2012). FASTUNIQU: A fast *de novo* duplicates removal tool for paired short reads. *PLoS One*, 7, e52249. <https://doi.org/10.1371/journal.pone.0052249>
- Xu, M. Z., Yang, L. H., Kong, H. H., Wen, F., & Kang, M. (2021). Congruent spatial patterns of species richness and phylogenetic diversity in karst flora: Case study of *Primulina* (Gesneriaceae). *Journal of Systematics and Evolution*, 59, 251–261. <https://doi.org/10.1111/jse.12558>
- Yates, C. M., Filippis, I., Kelley, L. A., & Sternberg, M. J. (2014). SUSPECT: Enhanced prediction of single amino acid variant (SAV) phenotype using network features. *Journal of Molecular Biology*, 426, 2692–2701. <https://doi.org/10.1016/j.jmb.2014.04.026>
- Yi, X., Liang, Y. U., Huerta-Sanchez, E., Jin, X., Cuo, Z. X. P., Pool, J. E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T. S., Zheng, H., Liu, T., He, W., Li, K., Luo, R., Nie, X., Wu, H., Zhao, M., Cao, H., ... Wang, J. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, 329, 75–78. <https://doi.org/10.1126/science.1190371>
- Zachos, J. C., Dickens, G., & Zeebe, R. E. (2008). An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature*, 451, 279–283. <https://doi.org/10.1038/nature06588>
- Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, 19, 153. <https://doi.org/10.1186/s12859-018-2129-y>
- Zhang, Z., Hu, G., Zhu, J., & Ni, J. (2011). Spatial heterogeneity of soil nutrients and its impact on tree species distribution in a Karst forest of Southwest China. *Chinese Journal of Plant Ecology*, 35, 1038–1049. <https://doi.org/10.3724/SP.J.1258.2011.01038>
- Zhou, D. S., Zhou, J. J., Li, M., & Yu, X. L. (2016). *Primulina suichuanensis* sp. nov. (Gesneriaceae) from Danxia landform in Jiangxi. *China. Nordic Journal of Botany*, 34, 148–151. <https://doi.org/10.1111/njb.00956>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Ke, F., Vasseur, L., Yi, H., Yang, L., Wei, X., Wang, B., & Kang, M. (2022). Gene flow, linked selection, and divergent sorting of ancient polymorphism shape genomic divergence landscape in a group of edaphic specialists. *Molecular Ecology*, 31, 104–118. <https://doi.org/10.1111/mec.16226>