JIPB Journal of Integrative Plant Biology

Research Article
https://doi.org/10.1111/jipb.13653

Population genomics highlights structural variations in local adaptation to saline coastal environments in woolly grape

Tianhao Zhang[1,2,3,4†], Wenjing Peng[1,2,5†], Hua Xiao[1], Shuo Cao[1,6], Zhuyifu Chen[1], Xiangnian Su[1,2,3], Yuanyuan Luo[7], Zhongjie Liu[1], Yanling Peng[1], Xiping Yang[2,5], Guo-Feng Jiang[2,3*], Xiaodong Xu[1*], Zhiyao Ma[1*] and Yongfeng Zhou[1,8*]

1. National Key Laboratory of Tropical Crop Breeding, Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518000, China
2. State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Guangxi University, Nanning 530004, China
3. Guangxi Key Laboratory of Forest Ecology and Conservation, Guangxi Colleges and Universities Key Laboratory for Cultivation and Utilization of Subtropical Forest Plantation, College of Forestry, Guangxi University, Nanning 530004, China
4. College of Informatics, Huazhong Agricultural University, Wuhan 430070, China
5. Guangxi Key Laboratory of Sugarcane Biology, College of Agriculture, Guangxi University, Nanning 530004, China
6. Key Laboratory of Horticultural Plant Biology Ministry of Education, Huazhong Agricultural University, Wuhan 430070, China
7. Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou 450009, China
8. National Key Laboratory of Tropical Crop Breeding, Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China

†These authors contributed equally to this work.

*Correspondences: Guo-Feng Jiang (gfjiang@gxu.edu.cn); Xiaodong Xu (xuxiaodong@caas.cn); Zhiyao Ma (mazhiyao@caas.cn); Yongfeng Zhou (zhouyongfeng@caas.cn, Dr. Zhou is fully responsible for the distributions of all materials associated with this article)

Tianhao Zhang

Yongfeng Zhou

ABSTRACT

Structural variations (SVs) are a feature of plant genomes that has been largely unexplored despite their significant impact on plant phenotypic traits and local adaptation to abiotic and biotic stress. In this study, we employed woolly grape (*Vitis retordii*), a species native to the tropical and subtropical regions of East Asia with both coastal and inland habitats, as a valuable model for examining the impact of SVs on local adaptation. We assembled a haplotype-resolved chromosomal reference genome for woolly grape, and conducted population genetic analyses based on whole-genome sequencing (WGS) data from coastal and inland populations. The demographic analyses revealed recent bottlenecks in all populations and asymmetric gene flow from the inland to the coastal population. In total, 1,035 genes associated with plant adaptive regulation for salt stress, radiation, and environmental adaptation were detected underlying local selection by SVs and SNPs in the coastal population, of which 37.29% and 65.26% were detected by SVs and SNPs, respectively. Candidate genes such as *FSD2*, *RGA1*, and *AAP8* associated with salt tolerance were found to be highly differentiated and selected during the process of local adaptation to coastal habitats in SV regions. Our study highlights the importance of SVs in local adaptation; candidate genes related to salt stress and climatic adaptation to tropical and subtropical environments are important genomic resources for future breeding programs of grapevine and its rootstocks.

Keywords: climate change, grape breeding, local adaptation with gene flow, salt tolerance, viticulture, *Vitis*

Zhang, T., Peng, W., Xiao, H., Cao, S., Chen, Z., Su, X., Luo, Y., Liu, Z., Peng, Y., Yang, X., et al. (2024). Population genomics highlights structural variations in local adaptation to saline coastal environments in woolly grape. J. Integr. Plant Biol. **00:** 1–19.

INTRODUCTION

Structural variations (SVs) are alterations in the genomic sequence that exceed 50 base pairs (bp) in length. These variations encompass a variety of changes, such as insertions, deletions, duplications, inversions, translocations, and other complex variations (Alkan et al., 2011; Gaut et al., 2018). While SVs are not as common as other genetic variants like single nucleotide polymorphisms (SNPs) and small insertions/deletions (InDels), they have a much greater impact on the total number of base pairs and genomic landscapes (Mills et al., 2011; Sudmant et al., 2015; Zhou et al., 2019a). SVs not only play critical roles in humans and animals, such as phenotypic outcomes in human evolution and diseases (Collins et al., 2020) and local adaptation among different ecotypes of deer mice (Hager et al., 2022), but also influence plants regarding agronomic traits associated with crop domestication (Gaut et al., 2018). Although research has consistently shown the importance of SVs in plant evolution and agriculture over the past decade (Gaut et al., 2018; Alonge et al., 2020; Shang et al., 2022; Zhou et al., 2022), there has been a notable lack of investigation into SVs within plant genomes (Zhou et al., 2019a; Kou et al., 2020).

How evolutionary forces shape patterns of genetic diversity within species is a fundamental question in the field of evolutionary biology (Mitchell-Olds et al., 2007). These are based on the Hardy–Weinberg equilibrium and the neutral mutation-random drift theory (shortened to the neutral theory), which considers that the fixation of mutation alleles is random or undergoes neutral in selection under continuous mutation pressure by random drift (Kimura, 1968). While drift plays an important role in some portions of the allele's journey, the influence of natural selection could still have a dominant influence on the evolutionary trajectory at other points (Kern and Hahn, 2018). Characterizing the genetic architecture of different ecotypes and understanding the dynamics of migration, selection, and recombination are critical for comprehending the complex processes of local adaptation (Savolainen et al., 2013; Hamala and Savolainen, 2019; Liu et al., 2022). When different populations exchange migrants, local adaptation is the result of the balance between selection and migration. The evolution of local adaptation requires alleles that confer high fitness in one environment but confer lower fitness in the other environment (Kawecki and Ebert, 2004). Alternatively, alleles with the highest overall fitness would be expected to invade the other population, and the locus would become fixed or monomorphic (Hall et al., 2010). It has been reported in plants that the inversion type of SVs has been beneficial for species. For example, studies have shown that sunflowers (Todesco et al., 2020; Zhou and Gaut, 2020; Huang et al., 2022), maize (Fang et al., 2012) and monkeyflowers (Lowry and Willis, 2010), which have experienced adaptive divergence with gene flow, have largely avoided the accumulation of deleterious mutations. However, few studies have focused on other types of SVs and how the diversification of SVs contributes to local adaptation in plants (Hämälä et al., 2021).

Against the backdrop of the persistent and profound global climate change phenomenon, it is expected that temperatures will continue to rise (Malhi and Wright, 2004). The accompanying extreme climate such as heat waves and heavy rainfall is anticipated to have substantial consequences for the viability and geographical range of local species (Choat et al., 2012). In coastal habitats, plants are required to undergo adaptations to withstand and thrive in challenging environmental conditions such as elevated salt concentrations, extreme temperatures and radiation, intense winds, sandy soil, and various other detrimental factors that can impede plant growth and development. So, the significance of the fitness of specific species in coastal and inland regions cannot be overstated. Nevertheless, previous studies have primarily concentrated on the analysis of mangroves or semi-mangroves located in the intertidal zone of tropical coastal ecosystems (He et al., 2022; Wang et al., 2022). The mangrove species only inhabit coastal environments and have limited adaptability to inland habitats. Limited research has been conducted on the extent of local adaptation in specific widespread species across their coastal and inland environments. Investigating the significance of local adaptation in shaping the distribution and viability of plant species across inland and coastal habitats will enhance our understanding of ecotypic differentiation and offer valuable recommendations for crop breeding.

The grape genus (*Vitis* L.) is widely distributed in the Northern Hemisphere and is well known worldwide for its important fruit crop: domesticated grape (*Vitis vinifera* L.) (Chen and Manchester, 2007; Moore et al., 2016; Zhou et al., 2019a). East Asia harbors approximately 40 species of wild grapes, making it one of the genus's centers of diversity. (Chen et al., 2007; Ma et al., 2022). The evolutionary history of East Asian grapes indicates that this group migrated from Northeast Asia to Southeast Asia during the Miocene (Ma et al., 2018). The East Asian grapes have undergone adaptive radiation to occupy different niches (Ma et al., 2018; Ma et al., 2022). However, the molecular mechanism of local adaptation in wild grapes remains poorly understood, especially the woolly grape (*Vitis retordii* Roman. du Caill. ex Planch.), distributed in both coastal and inland habitats in East Asia. Our field survey confirmed that populations along the coast and inland exhibited high adaptability at each sampling point. Furthermore, we observed population-specific phenotypes with only subtle differences, suggesting potential local adaptation with gene flow (Kawecki and Ebert, 2004; Hamala and Savolainen, 2019). Previous studies of SVs in grapevine have concentrated on the role of domestication (Zhou et al., 2019a) rather than exploring the adaptability and genetic differences among wild grape varieties. The wild woolly grape, with both inland and coastal distributions, serves as an excellent model for exploring the effects of SVs that contribute to local adaptation to these contrasting environments (Figure S1). The investigation of wild grape

germplasms holds significant value for breeding programs aiming to improve high-quality traits in cultivated grapes and their rootstocks.

Our study aimed to elucidate the genetic mechanism of SVs that underlies the coastal and inland environments. To accomplish this, we initially constructed a haplotype-resolved reference genome for the woolly grape. Second, we genotyped SNPs, InDels < 50 bp, and SVs using whole-genome sequencing (WGS) data from accessions of natural inland and coastal populations. Third, we defined the demographic histories between inland and coastal populations through coalescent modeling. Primarily, we identified a set of candidate genes that exhibited significant differentiation and was associated with SVs in the context of local adaptation to coastal and inland habitats. Our study emphasizes the significance of SVs in facilitating local adaptation. Furthermore, the identified candidate genes, particularly those related to salt tolerance, hold promise for future grapevines and crop breeding applications.

RESULTS

The haplotype-resolved reference genome for woolly grape

Since there were limited genomic resources for wild grapes, we assembled a haplotype-resolved telomere-to-telomere reference genome for woolly grapes using ~56.98 Gb (~106×) of PacBio HiFi CCS reads and ~71.02 Gb (~132×) of Hi-C reads. Before assembly, a genome survey estimated the genome size to be ~516.82 Mb and heterozygosity of 0.542% according to k-mer analysis (Figure S2). Using Hifiasm, contig N50 sizes of 21.09 Mb and 27.95 Mb were achieved for *V. retordii* haplotype 1 and haplotype 2, respectively. After anchoring, the scaffold N50 sizes reached 28.08 Mb and 28.37 Mb, respectively. With sufficient sequencing depth and optimization of the genome assembly strategy, the average number of gaps in each pseudo-chromosome was only 0.89, and more than half of the pseudo-chromosomes (21/38) consisted of one contig (Figure S3). After gap-closing, six missing genes of the core embryophyte gene set were found in unanchored contigs, and two fragmented genes were completely repaired (Table S2).

The woolly grape genome resulted in total lengths of 533.26 Mb for haplotype 1 and 541.28 Mb for haplotype 2, respectively. Benchmarking universal single-copy orthologs (BUSCOs) assessment indicated that 98.76% of the core embryophyte gene set was complete in the complete level assembly, combining two haplotypes (Table S2). The LTR Assembly Index (LAI) of haplotype 1 and haplotype 2 assessed 18.83 and 16.74, respectively, which characterized that both haplotypes of woolly genome reached reference level. Furthermore, we defined the positions of all centromeric regions (38/38) and almost all telomeres (74/76) except for one end of chromosome 17 (Figures 1A, S5). The woolly grape genome assembly boasts not only quality comparable with T2T published genomes (Chen et al., 2023; Shang et al., 2023) but also represents the highest-quality genome and largest genome size reported for grapevines (Table S14).

Transposable element (TE) identification using EDTA and a species-specific repetitive sequence database (Shi et al., 2023) revealed that ~50.79% of haplotype 1 and ~51.64% of haplotype 2 in the woolly grape genome consisted of repetitive sequences (Table S3). Interestingly, 47 TIR *Mavericks* (total 27,133 bp; also known as *Polintons*), which could encode virus capsid proteins (Krupovic and Koonin, 2015), were only identified in haplotype 1 but not in haplotype 2, which revealed heterozygosity of TEs in plant genomes. In total, 33,481 protein-coding genes in haplotype 1 and 29,845 in haplotype 2 were annotated with Maker pipeline using transcriptome sequences and homologous protein sequences. We also detected 3,215 (9.60%) hemizygous genes in the haplotype-resolved reference genome for woolly grape using two haplotypes.

Comparative genomics revealed SVs in the centromeric region

The k-mer-based estimate of heterozygosity was 0.542% in woolly grape, which ignored the effects of SVs. Inter-haplotype SVs were detected by mapping raw HiFi CCS reads to woolly grape haplotype 1 and haplotype 2. For PN_T2T, HiFi reads were mapped to the haploid PN_T2T reference (Figure 1B). The number of SVs in *V. retordii* was far more than that in PN_T2T originating from successive selfings (Shi et al., 2023). Among the five SV types, deletions (DELs) and insertions (INSs) were the most abundant, with 3,135 DELs and 2,981 INSs observed in PN_T2T and an average of 17,278 DELs and 15,187 INSs in *V. retordii* (Figure 1B).

The woolly grape and PN_T2T reference genome (Shi et al., 2023) exhibited synteny across most genomic regions. However, large SVs (>100 kb) between the two genomes, particularly inversions, occurred mainly in centromeric regions (Figure 1A). Taking the longest interspecific inversion in chromosome 18 as an example, sequences of centromeric region in chromosome 18 were highly repetitive and an inversion between PN_T2T and *V. retordii* was revealed by the collinearity map and (and the Hi-C heat map in chromosome 18 (Figures 1C, S6). The 107 bp repeats were most abundant and had the longest total length in centromeric regions of 19 chromosomes in the *V. retordii* reference genome, but not on chromosome 18 of PN_T2T (Figure 1D). We suspected that the centromeric regions might have changed during grapevine domestication. Genes related to carpel (GO: 0048440, q-value = 6.04E-07), gynoecium (GO: 0048467, q-value = 8.36E-07), and plant ovule development (GO: 0048481, q-value = 4.32E-06) were enriched in the centromeric region of chromosome 18 of woolly grape, but not in PN_T2T (Figure 1E; Table S13).

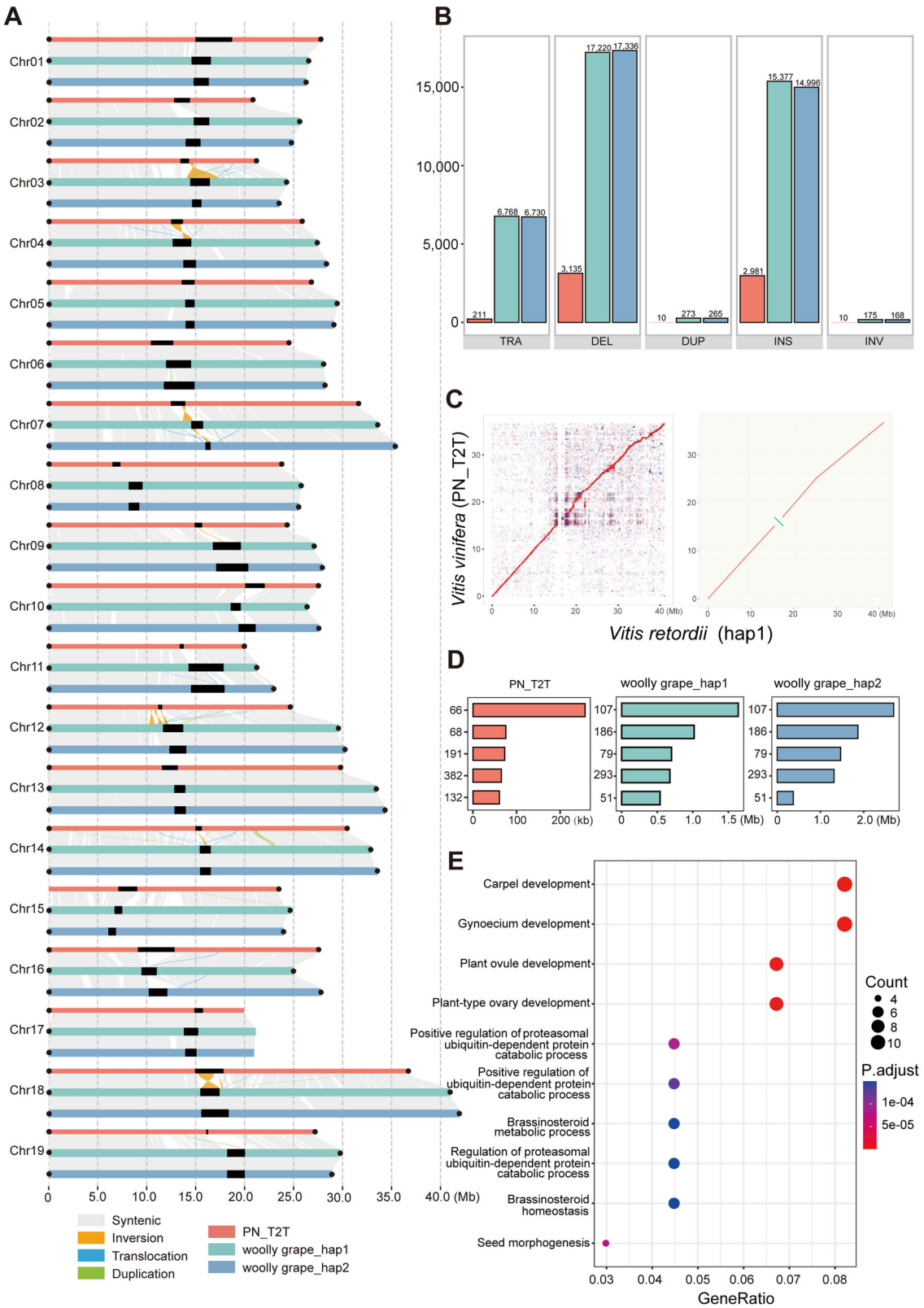

Figure 1. Telomere-to-telomere complete reference genomes of woolly grape *Vitis retordii*

(A) Synteny between PN_T2T reference genome (red) and two haplotypes of *V. retordii* reference genome (green and blue). Syntenic, Inversion, Translocation and Duplication consistent with legends. Black solid dots and areas covered by black rectangle indicated the telomere and centromeric regions, respectively. The *V. retordii* reference genome only missing the end telomere of chromosome 17. **(B)** Based on each HiFi CCS reads, structural variations of each type number statistics. **(C)** Chromosome 18 synteny between PN_T2T and *V. retordii* haplotype 1. **(D)** The total length of different repeat unit copies in chromosome 18. **(E)** The Gene Ontology (GO) enrichment of centromeric region in *V. retordii* chromosome 18.

Population structure and genetic diversity

Mapping whole-genome resequencing data to the woolly grape reference genome, we conducted population genetic analysis of 33 samples from three natural populations including one coastal (Shenzhen, SZ) and two inland populations (Luocheng and Longan, LC and LA, respectively; Figure 2A; Table S1). The rooted NJ tree showed three distinct branches corresponding to three natural populations with *Muscadinia* grape accessions as an outgroup, indicating that coastal population SZ had split earlier than inland populations. In admixture analyses, three populations were clustered according to their geographical location and individuals were clustered according to their respective populations. The PCA results were broadly consistent with the outcomes of the admixture analysis, with the first and second principal components (PC1 and PC2) accounting for 45.82% and 21.42% of the variance, respectively. Population structure analyses revealed a clear division among the three populations by admixture, phylogenetic tree, and PCA analyses (Figure 2C, D).

The coastal population (SZ) had lower nucleotide genetic diversity ($\pi = 0.45 \times 10^{-3}$) than the inland populations (1.95×10^{-3} and 1.56×10^{-3} for LC and LA, respectively). Additionally, coastal populations had lower Tajima's *D* values (Tajima's $D = 0.40$) compared with inland populations (0.91 and 0.61 for LC and LA, respectively; Figures 2B, S7). It is suggested that both coastal and inland populations may suffer from bottleneck events and coastal populations face more severe selective pressure. In addition, the difference between the two inland populations ($F_{ST} = 0.05$) was smaller than that between them and the coastal population (F_{ST} is 0.16 and 0.11 for SZ-LC and SZ-LA, respectively), which aligns with the geographical distances of the populations. The genome-wide linkage disequilibrium (LD) decayed to half of the maximum value all within 100 kb in these three populations (Figure S8), but SZ decayed slowest, followed by LA and LC.

Demographic history showed asymmetric gene flow from the inland to the coastal population

We primarily utilized fastsimcoal27 to examine the population demographic history of these three populations. With the SMC++ results, which simulated one population (Figure S14) and population structure analysis above, we concluded all three populations had experienced bottleneck events in their history. To account for each situation and prevent overfitting, we evaluated a total of 32 models with different demographic scenarios. Each model ran 100 times independently (Table S5). Based on Akaike information criterion (AIC) scores (Figure S15), the best-fitting model was chosen. This model (Figure 2E) indicated that a recent bottleneck occurred approximately 1,100 years ago, two inland populations, LC and LA, diverged around 9,700 years ago, and the divergence between the coastal population SZ and the ancient inland population was about 2.52 Mya. The estimated effective population size of the coastal population SZ was lower than that of the inland populations LC and LA (Table S6). In addition, different migration rates existed in different periods between different populations, but migration rates between LA and SZ recently were higher than others. The recent gene flow between LA and SZ was heavily biased, with a much higher gene flow from the inland population LA toward the coastal population SZ gene flow than from SZ to LA. To ensure the preciseness of our gene flow results, we performed TreeMix analysis indicating that the optimal migration edge number was 1 and gene flow occurred from LA to SZ as well (Figure S17). Furthermore, F3 statistics (Figure 2F) also supported that LA and SZ significantly had a high gene flow. The *f* branch introgression (f_{dM}) statistics by Dsuite also suggested that LA and SZ had a higher number of introgression regions and a deeper introgression degree, which was consistent with the results above (Figure S18).

The deleterious and SV burdens in inland and coastal populations

In total, 21,148 SVs (Figure S19), including 10,973 translocations (TRAs), 8,839 deletions (DELs), 758 duplications (DUPs), 438 insertions (INSs), and 140 inversions (INVs) were detected from the three *V. retordii* populations using DELLY based on 33 resequenced genome data. In total, 4,520 genes (13.50% of all genes) overlapped with SVs or within the regions 5 kb of SVs, and were potentially affected by SVs. We further conducted statistics on the base lengths of SNPs, InDels, and TEs in those genes (Figure 3A), which found that at least one-third of SVs was most influenced by TEs. The base lengths of TEs were dominant in ~37.41% (1691/4520) of genes, while SNPs were dominant in 17.39% (786/4520) of genes and InDels were dominant in 3.10% (142/4520) of genes. We were interested in how SVs affected local adaptation between coastal and inland ecotypes so, based on the SVs data, we counted the unique SVs of coastal and inland populations (Figure 3C). The coastal population had a much smaller number of unique SVs than inland populations, while more proportions of deletion events were observed (Figure S20). This further indicated that there was no significant difference in the proportion of various classes of SVs between coastal and inland populations, and coastal populations had fewer rare SVs, which may be related to lower nucleotide genetic diversities.

Then, we quantified the burden of SVs and SNPs. For SVs, we hypothesized that they contributes to the deleterious load, potentially reflecting the cost of evolution. For SNPs, the detection of deleterious mutations used a sorting of intolerant from tolerant (SIFT) algorithm (Vaser et al., 2016) to detect the effect of single nucleotide non-synonymous mutations on protein function. We found that the burden of both SVs and SNPs of coastal populations was lower than that of inland populations, except that SZ had medium heterozygous SNP burdens (Table S10; Figure S24). In three populations, SV burdens of recessive, heterozygous, and additive LC were highest, followed by LA, and SZ as the lowest (Figure 3B). SNP burdens of recessive and additive LA were highest, followed by LC, SZ was lowest, while SNP burdens

A

Population	Habitat	Average longitude	Average latitude
Shenzhen (SZ)	Coastal	114°29′30.04″ (E)	22°32′44.46″ (N)
Longan (LA)	Inland	107°38′36.96″ (E)	22°56′58.42″ (N)
Luocheng (LC)	Inland	108°40′39.16″ (E)	24°45′58.58″ (N)

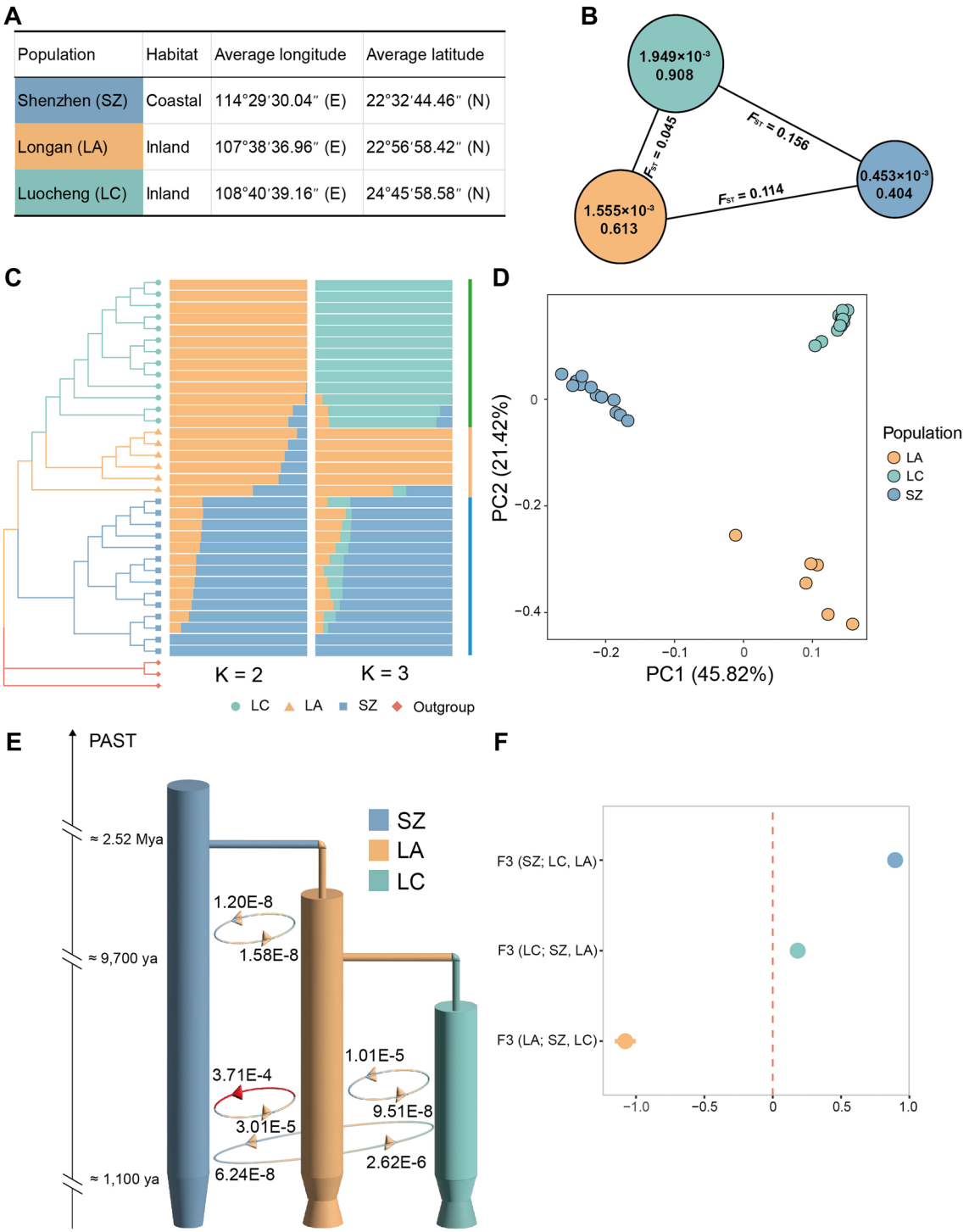

Figure 2. Population distribution, population structure and demographic history of *Vitis retordii*. Blue, yellow, and green are stand for population SZ (coastal), LA (inland) and LC (inland) respectively

(A) The distribution information of three *V. retordii* populations. **(B)** Estimated of nucleotide genetic diversities (π; above inside the circle), Tajima's *D* values (below inside the circle) and the Wright fixation index value (F_{ST}; solid line) summary statistics for these three populations, the size of the circle represents the size of the π value. **(C)** Phylogenetic tree and admixture analysis. The rooted tree toke *Muscadinia* (red) as outgroup and admixture analysis set $K = 2$ and $K = 3$. **(D)** Principal component analysis (PCA) of all 33 *V. retordii* individuals. **(E)** Schematic of the best demographic scenario modeled using fastsimcoal27. The direction of black arrow is ancient. The thinner place below represents the bottleneck and the direction of the arrow on the circle represents the direction of gene flow. **(F)** F3 statistics among these three populations. Used standard deviation (*SD*) as error bar.

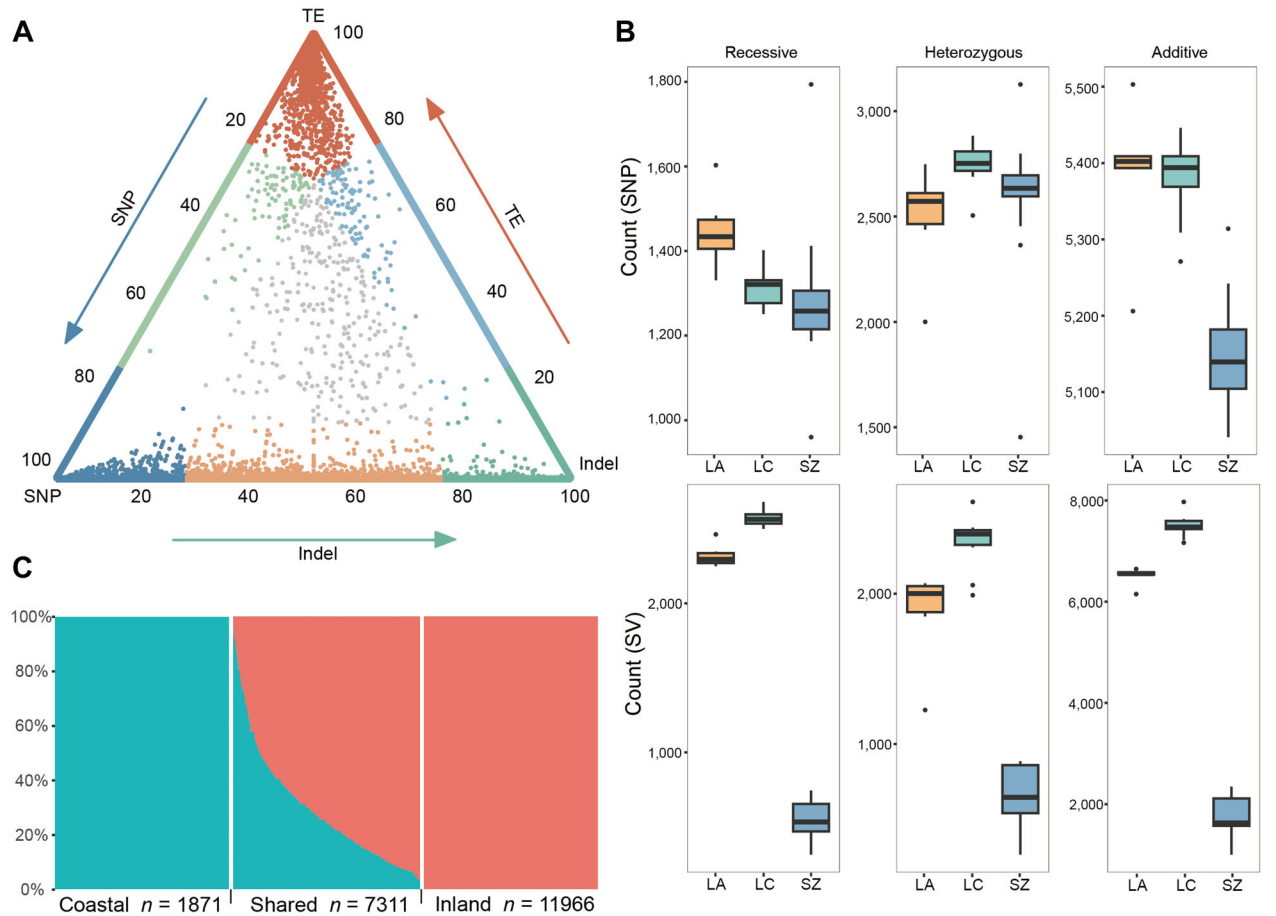

Figure 3. Population SVs pattern and burden

(A) The proportion of SNP, InDel and TE occupied base length in genes, every point stands for a gene related to SVs in the genome. **(B)** SVs and SNPs burden of these three populations show coastal population have fewer burdens. Blue, yellow, and green are stand for population SZ (coastal), LA (inland) and LC (inland), respectively. **(C)** Distribution of SVs between coastal and inland populations.

of heterozygous LC was highest, followed by SZ, and LA as the lowest (Figure 3B). The additive SNP burdens of two inland populations, LA and LC, were not significantly different ($P > 0.92$, two-sided KS test), but the LC population had more heterozygous burdens and less recessive burdens than the LA population. In total, this may reflect that populations with larger effective populations are more likely to conceal heterozygous SNP burdens under similar conditions, while heterozygous SV burdens were not sheltered. These contrasting patterns potentially support the notion that SVs exert a greater impact on the genome compared with SNPs. Once introduced, SVs may impose a more substantial burden, leading to their rapid elimination through selection or purging via hybridization between individuals.

Local adaptation to coastal and inland habitats

To investigate the population-specific differentiation between coastal and inland populations, we detected outlier selection events between coastal and inland populations and found some outlier loci stood out, including 386 genes selected in the coastal population by SVs, such as *FSD2* (GO: 0006801, *q*-value = 3.39E-09; Figure S19). Annotated as *Vitis16g00281* in woolly grape haplotype 1, *FSD2* (Fe superoxide dismutase 2) is a protein-coding gene that produces superoxide dismutase in *Arabidopsis thaliana*, and was an orthologous protein in grapevine. It had been confirmed to regulate various abiotic and biotic stress conditions, including nutrient depletion or salt stress via both transcriptional and post-translational mechanisms (Garcia-Molina et al., 2017). We also used population branch statistic (PBS) analysis (Yi et al., 2010) to detect recent natural selection by PBScan based on 237,534 filtered SNPs with a window size of 50 kb (Figure S22), and found 725 coastal and 1048 inland (810 in LC and 238 in LA) population-specific selective genes (Figure 4B). Coastal population-specific selective genes were found to be enriched in several biological processes (BP). The most significant processes included regulation of inward rectifier potassium channel activity, negative regulation of response to salt stress (GO:1901001, *q*-value = 1.66E-05), regulation of potassium ion transmembrane transporter activity, and regulation of cation/potassium ion transmembrane transport (Figure 4D). Response to photo-oxidative stress (GO: 0080183, *q*-value = 1.85E-04) and

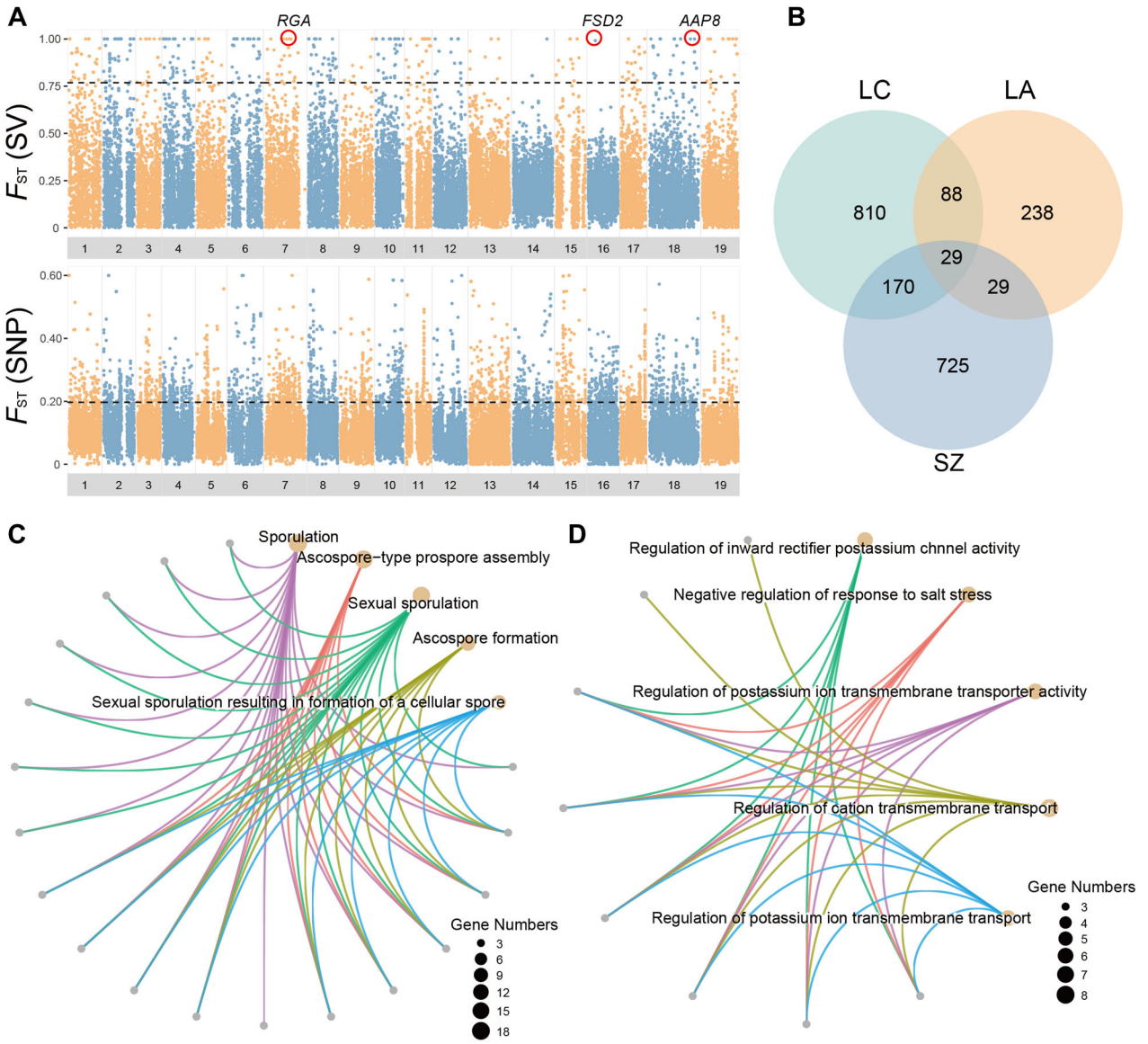

Figure 4. Population-specific differentiations in local adaptation of *Vitis retordii*

(A) SVs and SNPs F_{ST} values distribution in the whole genome between coastal and inland populations. The black dashed line is the boundary line for the highest 5% F_{ST}. **(B)** Number of annotated genes within 5 kb of the significant ($q < 0.05$) PBS outlier window. **(C, D)** GO enrichment of genes under inland habitat population-specific selection and coastal population-specific selection, respectively. The gray dots represent a gene, the yellow dots represent enriched biological processes, and the same-colored lines represent enriched genes in that biological process.

regulation of response to salt stress (GO:1901000, *q*-value = 6.92E-04) were under positive selection as well (Table S8). In total, 1,035 genes associated with plant adaptive regulation for salt stress, radiation, and environmental adaptation were detected underlying local selection in coastal populations, of which 37.29% and 65.26% were detected by SVs and SNPs, respectively.

Furthermore, we found some genes in both the highest F_{ST} peak for SVs and SNPs (Figure 4A), such as *RGA1* and *AAP8*. We found that inland populations had a 15,228 bp deletion downstream of the *RGA1* gene (*Vitis07g01301*) on chromosome 7, which resulted in a 10 bp deletion in its coding sequence (CDS; Figure S26). *RGA1* is an important negative regulator of the gibberellin (GA) signal transduction pathway. It could suppress the defects in stem elongation, flowering time, and leaf abaxial trichome initiation (Silverstone et al., 1998) and it is also related to the translation of Ca^{2+} signals. The *AAP8* gene (*Vitis18g01992*) in chromosome 18 intron segment had a 3,731 bp deletion in inland populations (Figure 5G), which may influence the transport of nitrogen in the xylem, as reported in *Arabidopsis thaliana* (Santiago and Tegeder, 2016). We also found a highly differentiated inversion in chromosome 18, with length 95,967 bp from 34,660,228 to 34,756,195 bp. Both from SVs

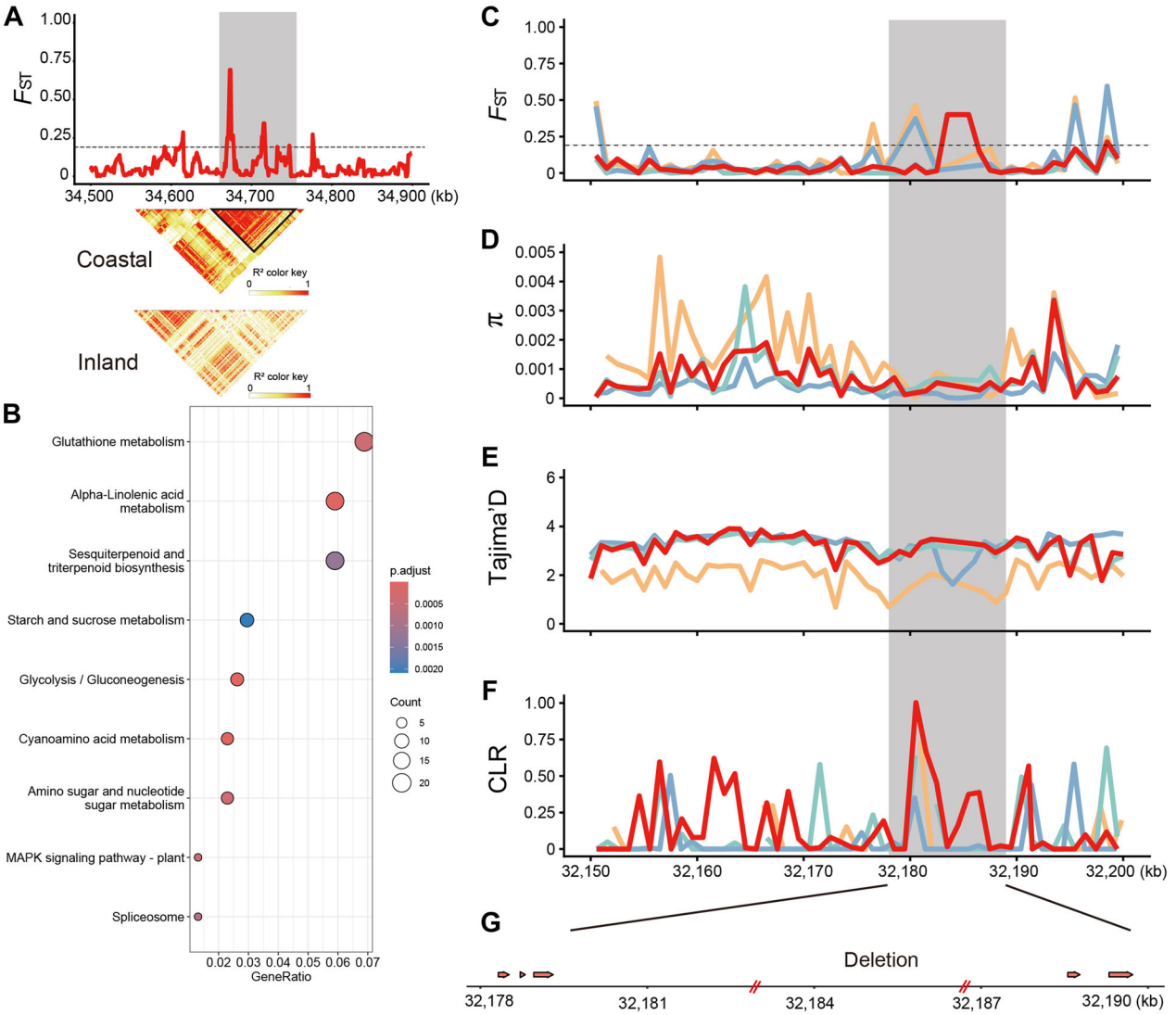

Figure 5. The detailed information of loci and evidences of positive selective sweeps among three populations and SVs burden
(A) F_{ST} in chromosome 18 (34,500–34,900 kb). The gray part represents inversion region (34,600–34,756 kb). The above figure and the following figure respectively represent the linkage disequilibrium in the inversion section of coastal and inland populations. The black dashed line is the boundary line for the highest 5% F_{ST}. **(B)** Enrichment analysis of KEGG pathway related to SVs. **(C)** For F_{ST}, Blue, yellow, green, and red were stand for population SZ versus LA, LA versus LC, SZ versus LC, and population coastal versus two inland population. The black dashed line is the boundary line for the highest 5% F_{ST}. **(D–F)** For π, Tajima's *D* and CLR, blue, yellow, green, and red were stand for population SZ (coastal); LA; LC; LA and LC (all inland) respectively. **(G)** The *AAP8* locus detailed information, within the red slash stands for a 3,731 bp deletion in the intron region from 32,182,897 to 32,186,628 bp.

and SNPs, these BPs focus on the regulation of salinity and light stress in coastal habitats, reflecting the local adaptation of coastal ecotypes to natural selection in high-radiation and high-salt environments.

Furthermore, under inland habitat population-specific selection, Gene Ontology (GO) enrichment indicated several genes related to the BP, including for sporulation (GO:0043934, *q*-value = 1.88E-08), ascospore-type pros pore assembly, sexual sporulation, ascospore formation, sexual sporulation resulting in formation of a cellular spore, and so on (Figure 4C; Table S7). In plants, the meiosis of microspore mother cells (microsporocytes) produces four daughter cells, and furthermore forms four haploid microspores, namely mononuclear pollen grains. This revealed that the formation of plant pollen in inland populations was under the strongest natural selection. This may be a response to the pollination process in more arid inland habitats that are more vulnerable to high-temperature stress against the background of global warming. In addition, we also conducted GO enrichment analysis to summarize the BP with the shared outlier genes from all three populations (Figure S21). These were vascular and phloem transport (GO:0010232/0010233, *q*-value = 4.38E-10), detection of biotic stimulus and bacterium (GO:0009595/0098543, *q*-value = 1.88E-09), and many other BPs under selection in both coastal and inland populations (Table S9). This not only showed that water is the basis for

plant survival, but also revealed the stress resistance necessary for wild plant survival and evolution, which is also consistent with the subtropical habitats of *V. retordii*.

Evidence of SVs in coastal adaptation

Chromosomal inversion plays an important role in local adaptation, although no inter-population inversions greater than 1 Mb have been detected by population genomics. However, we identified a highly differentiated inversion (~96 kb) between coastal and inland populations on chromosome 18. Resequencing reads mapping back to the genome and viewing breakpoint locations in Integrative Genomics Viewer (IGV) proved the authenticity of inversion (Figure S25). Notably, 13 out of 14 individuals in the coastal population harbored this inversion, exhibiting complete homozygosity, while individuals from inland populations lacked it entirely (Figure 5A). We also found that F_{ST} was consistently elevated in this inversion region, and in high LD in coastal populations but not in inland populations (Figure 5A). There were six short genes in this region, including a gene found only in grapevines that has not yet been verified for its function (*Vitis07g02049*), an *RPM1* gene (*Vitis07g02050*), two (Reduced Chloroplast Coverage) *REC* genes (*Vitis07g02051*; *Vitis07g02052*) and two (transporter A family member) *ABC* genes (*Vitis07g02053; Vitis07g02054*). *RPM1* is a famous resistance (R) gene that affects plant disease resistance and morphology, which could act on *ABC* to resist cellular oxidative stress and prevent the invasion of pathogens. Recent research has suggested that this gene cluster may play a role in regulating light responses in grapes (Zhang et al., 2023). Many selected loci related to SVs are enriched in different KEGG pathways, which could be related to coastal adaptation (Figure 5B). Many significantly enriched pathways including alpha-linolenic acid metabolism, glycolysis/gluconeogenesis, amino sugar and nucleotide sugar metabolism, glutathione metabolism, and the MAPK signaling pathway; plants are related to salt tolerance to varying degrees.

We characterized the different summary statistics that were sensitive to the effects of a selective sweep around *AAP8* in coastal and inland populations. Quantified population genetic differentiation (F_{ST}) was significantly higher in the *AAP8* locus (Figure 5C), and F_{ST} in *AAP8* was higher than the whole-genome mean F_{ST}. Through the F_{ST} value of coastal versus inland the most obvious difference can be seen. The nucleotide diversity (π) at the *AAP8* locus was significantly reduced, which showed strong natural selection following a decrease in heterozygosity (Figure 5D). Tajima's *D* values were lower in this region, suggesting a reduction in nucleotide polymorphism within alleles, while positive values indicated a persistence of rare alleles, albeit at a lower frequency (Figure 5E). We performed a SWeeD analysis (CLR) and separately evaluated the evidence of positive selection among different populations, and found that the inland population had positive selection as expected for positive selection (Figure 5F). All these findings provided strong evidence for the occurrence of a selective sweep in the *AAP8* locus.

DISCUSSION

Coastal areas, where land meets the ocean, are important zones in natural ecosystems, often providing a habitat for a diverse range of plant species. Due to the transportation of sea salt aerosols from the ocean region by wind, the concentration of salt in coastal land regions differs significantly from that of regions further inland (Suzuki et al., 2002). Surprisingly, some species not only thrive in inland regions but are also fully adapted to coastal environments. How species could adapt to such divergent environments remains poorly understood, particularly in terms of the population genomic mechanisms.

In this study, we assembled the haplotype-resolved diploid reference genome for the wild grape species *V. retordii*. This genome is also the largest among all grapevine genomes assembled by far, measuring 541.28 Mb. Leveraging the woolly grape reference genome, we conducted comprehensive population genomic analyses comparing coastal with inland populations. Our analyses revealed extensive structural variants, gene flow, and some selective signals between the coastal and inland populations, which could be a model to explore the genetic mechanism of local adaptation. Candidate genes, including *FSD2*, *RGA1*, and *AAP8*, are associated with salt tolerance, plant growth, and adaptive regulation. These genes showed significant differentiation in the selected and SV regions and exhibited divergent selective pressures between coastal and inland populations. The insights derived from our findings will offer valuable resources for population genomics, evolution, functional genomics, and genetic improvement for grapevine and other crops.

Unbalanced gene flow between coastal and inland populations

Adaptive introgression, the transfer of adaptive genetic variation between species or populations, has contributed to the altitude adaptation of Tibetans (Huerta-Sánchez et al., 2014), the genetic rescue of Gulf killifish (Oziolor et al., 2019), the adaptive radiation of *Heliconius* butterflies (Edelman et al., 2019); and global spread and adaptation in domesticated grapevine (Xiao et al., 2023). In this study, we detected both ancient and contemporary two-way unbalanced gene flow between inland and coastal populations of *V. retordii* (Figure 2E). In introgressed regions, several genes associated with pollen tube guidance (GO:0010183, *q*-value = 3.64E-08), pollen maturation (GO:0010152, *q*-value = 2.47E-05), and drought recovery (GO:0009819, *q*-value = 1.31E-06) had been enriched. The gene flow from the inland population to the coastal population was 10 times stronger than the reverse direction, indicating that these introgressed genes most likely originated from the inland population. The inland population may have faced more severe reproductive stress. Thus some of the genes related to reproduction, such as sporulation and ascospore formation, were detected to be under selection in PBS analyses. When these genes are introgressed into the coastal population, they will presumably

enhance their adaptability. The inland and coastal populations were several hundred kilometers away from each other, which made direct contact unlikely. However, we hypothesize that seed dispersal by birds may have facilitated gene flow between the inland and coastal populations. Wild grape fruits are known to be appealing to birds, and researchers have reported that many bird species consume berries and disperse seeds during the migration transition (Zhou et al., 2019b; Gu et al., 2021).

SVs are constructive for local adaptation

Investigating the role of SVs in adaptation and divergence, we sought to identify the corresponding genes and BP underlying the distinct environmental responses of coastal and inland woolly grape populations. Despite the overall detrimental effects of SVs, we identified individual SVs that exhibited signatures of local adaptation. Some SVs that were specific to the coastal population were likely to be associated with the adaptation of *V. retordii* to the coastal environment. Genes involved in pathogen resistance, nutrient depletion, and salt stress are strongly enriched near these coastal-specific SVs, highlighting the contribution of SVs to coastal adaptation. Meanwhile, we utilized the PBS to process SNPs, via truly quantitative traits, which could detect large-effect loci more easily than small-effect ones (Yi et al., 2010). The same genes may not be identified through SNP-based analyses, demonstrating that SVs need to be considered to gain a full understanding of how selection shapes genomic variation (Gaut et al., 2018). SVs could influence a larger proportion of the genome than SNPs, and have been suggested to promote the adaptation of various species (Gaut et al., 2018; Zhou et al., 2019a; Collins et al., 2020; Hager et al., 2022).

We detected interesting selection patterns in the inland populations; the most selected genomic regions were found to be related to the formation of plant pollen grains. This may be related to the pressure of pollination on inland wild grapes; most of the cultivated grapes that have been domesticated today are cloned crops (Zhou et al., 2017), so there is no need to face the pressure of plant pollination. Whether grapes are mainly pollinated by wind or insects has not been decided (Branties, 1978), but the stamens had cracked and released pollen before the stigma matured, and the segregation phenomenon of sexual maturity may reveal that the wild grapes are mostly self-pollinated (Shackelford et al., 2013). Pollen has a short survival time in the outside world, the inland population-specific signal might have been be caused by the intensification of environmental changes (increased drought stress) or the difficulty of pollinators in reducing traditional pollination methods in recent years. In contrast, adaptive characteristics of coastal population related to coastal environments were distributed in positive and negative regulation of salt stress and response to photo-oxidative stress. It was the same as the adaptation of seashore plants to increased solar radiation and salt stress in seashore environments in many studies. Similar GO terms were also enriched in some other coastal plants, such as mangrove and *Hibiscus hamabo* (He et al., 2022; Wang et al., 2022), which indicated the possibility of convergent selections. Furthermore, some of these genes dominated by SVs were detected to be under selection, which probably promoted the adaptation of *V. retordii* to the coastal environment. For example, the selected *FSD2* gene (Figure 4A) has been proven to be related to salt stress and nutrient depletion, which could promote *V. retordii* to adapt to the high-salt concentration soil on the coast. Studies have reported that gene with the same pathway were involved in response to salt stress in wild grapes (Carrasco et al., 2022). SVs could also directly affect the gene function by specifically inserting into the CDS region of negative regulator gene *RGA1* of the coastal population, which eliminates the suppression of the gibberellin (GA) signal transduction pathway, and will probably promote the reproductive and vegetative growth of coastal *V. retordii*. The 90 kb inversion in chromosome 18, by suppressing the recombination of some functionally related R genes, probably also promoted the disease resistance of the coastal *V. retordii* population. Taking all the evidence into consideration, it is evident that SVs have played a significant role in various aspects of coastal adaptation in *V. retordii*.

The cost of local adaptation to the marginal coast

The coastal and inland populations of *V. retordii*, after diverging from a common ancestor, have evolved independently for approximately 1.85 million years (Figure 2E). This lengthy period allowed them to adapt extensively to their respective environments and likely resulted in the accumulation of divergent genetic variations. We have detected a significant number of genetic variations, including both SVs and SNPs, that were specific to either the coastal or inland populations of *V. retordii*. This provided an ideal system for studying the genetic implications of coastal adaptations.

Why did the coastal population have lower diversity and a smaller effective population size compared with the inland population? The most evident cost of coastal adaptations is probably more severe genetic diversity losses. Ancient coastal populations diverged earlier than the two inland populations, and should have had relatively higher genetic diversities, but conversely it exhibited the lowest genetic diversities among the three populations. On the one hand, this was consistent with a positive Tajima's *D* value, which suggested that a recent bottleneck occurred in all woolly grape populations. LD decayed slowest in the coastal population, which indicated that the coastal population experienced more serious bottlenecks so that the population diversity was not fully restored. On the other hand, this was consistent with the results of population demographic history by both SMC++ and fastsimcoal27. These all pointed out that the bottleneck occurred approximately 1,100 years ago and likely coincided with increasing human activity. Inland populations of *V. retordii* could migrate to more primitive areas but once coastal population lost their habitats it was difficult to recover. There was more extensive human disturbance in

coastal populations at Shenzhen, one of the largest cities in China.

The genetic burdens of both SNPs and SVs in the coastal population were lower than the inland populations, which indicated that the coastal population might be under even higher selection pressures, probably to purge deleterious burden and SVs burdens induced by strong bottlenecks and a relatively harsh and frequently changing environment. This is consistent with the depletion of rare deleterious variants in the range-edge population in previous studies (Takou et al., 2021). PBS analyses have indeed detected (15% percentage) more selected loci in the coastal population, which makes genetic hitchhiking more likely to happen (Stephan, 2010). The extensive diversity losses and ongoing bottleneck have probably also promoted the purging of deleterious mutations in coastal populations. This reduction in genetic burden, particularly regarding SVs, reflects the cost of coastal adaptation under linkage-neutral conditions, where deleterious mutations decrease allelic diversity. Compared with SNPs, the SVs were probably under more severe selections (Zhou et al., 2019a; Kou et al., 2020). Genome-scale analyses have revealed that SVs usually segregate at low frequencies and are depleted from functional regions of the genome, indicating a strong purifying selection of SVs. Despite the overall detrimental effects of SVs (Zhou et al., 2019a), we identified some individual SVs bearing signatures of local adaptation, several of which were associated with genes differentially selected between populations (Figure 4A). Genes involved in pathogen resistance, nutrient depletion and salt stress are strongly enriched near these coastal-specific SVs, highlighting the contribution of SVs to coastal adaptation. The large amount of diversity losses and ongoing bottleneck have probably also promoted the purging of deleterious mutations in coastal populations.

CONCLUSIONS

In this study, our population genomic analyses of the inland and coastal population of *V. retordii* provided new insight into the genomic mechanism that has contributed to the adaptation of plants to the high-salt coastal environment. The high-quality genome of the woolly grape also served as a treasurable genetic resource for screening salt-resistant grape genotypes. Some of the selection signals could be detected by both SNPs and SVs, while some could only have been discovered by SVs, which proved it was essential to take SVs into consideration in population genomic analyses. These results motivate incorporating SVs in population research to better elucidate the genomic mechanism of plant adaptation. We also detected a series of population-specific differentiation gene loci that was highly correlated with coastal local adaptation and could be utilized in genetic improvements of grapevine and other crops.

MATERIALS AND METHODS

Sample collection and genome sequencing

The woolly grape, *Vitis retordii* Rom.Caill. ex Planch. (2n = 2× = 38), is a wild grape distributed in the Northern Hemisphere subtropical zone, including Guangdong, Guangxi, and other provinces in southern China (Wen-Tsai, 1979), and parts of Laos and Vietnam (Bân, 2003; Newman et al., 2007). The sample for genome assembly was collected on a cliff on the coast (Figure S1A), which is located in Shenzhen City, Guangdong Province (latitude 22°32′49.66″N, longitude 114° 28′45.07″E, altitude 21.05 m). The *V. retordii* is distributed in subtropical areas, and the northernmost distribution record is 25°55′N. We sampled three natural populations at three sites, including one coastal population in Shenzhen City: SZ, 22°32′44.29″N, 114°29′33.52″E, and two inland populations: Luocheng: LC, 24°45′58.58″N, 108°40′39.16″E, and Longan: LA, 22°56′58.42″N, 107°38′36.96″E, respectively (Figure S1C). All 33 *V. retordii* samples were collected in the wilds in China. Whole-genome resequencing was derived from three sites where *V. retordii* grew well, including six individuals in Longan, 11 individuals in Luocheng and 14 individuals in Shenzhen; each of them had exact longitude and latitude information (BioProject: PRJNA953920; Table S1); moreover, we quoted two individuals for resequencing, SRR5891945 (Liang et al., 2019) and SRR7819168 (Ma et al., 2018), both from Luocheng.

Genome survey and genome assembly

GenomeScope 2.0 (Vurture et al., 2017) was used from the k-mer 51 histogram computed by Jellyfish v. 2.3.0 (Marcais and Kingsford, 2011); the estimated genome size was 516,816,094 bp and heterozygosity was 0.542% (Figure S2).

Genome assembly first used Hifiasm-0.15.5 "--hom-cov 101 -l 3" parameters (https://github.com/chhylp123/hifiasm) to complete the contigs level assembly, with two SMRT bell libraries constructed and sequenced on a PacBio Sequel II that generated ~56.98 Gb of HiFi CCS reads in total; coverage calculated with haploid was ~106×. Second, to complete the scaffold-level assembly, on the one hand Juicer v. 1.6 was used (https://github.com/aidenlab/juicer) to anchor all contigs (Durand et al., 2016b), with ~71.02 Gb of Hi-C reads generated in total, coverage calculated with the haploid was ~132×. On the other hand RagTag v2.1.0 (https://github.com/malonge/RagTag) "scaffold" function (Alonge et al., 2022), acquired.APG file with Cabernet Sauvignon was used as the reference genome (Chin et al., 2016), then agp2assembly.py was used (https://github.com/phasegenomics/juicebox_scripts) to convert the APG file to an assembly file, which contained the location information on contigs in the scaffold. Then, we used the 3d-dna (https://github.com/aidenlab/3d-dna) "visualize" function with parameter "-q 0" (Dudchenko et al., 2017), Hi-C map was checked and refined with Juicebox v.1.11.08 (Durand et al., 2016a). Only a few gaps on the Hi-C pseudo-chromosome, and even some of them a contig formed an integral scaffold

(Figure S3). After manual examination in Juicebox (Durand et al., 2016a), scaffold-level assembly was generated with 3d-dna. Finally, to patch gaps in the pseudo-chromosome and complete the complete level assembly, we adopted three strategies even though there were only 34 gaps in 38 pseudo-chromosomes.

Genome assembly assessment and identification of telomeres/centromeres

To assess genome assembly completeness, BUSCO v.5.2.2 was used with the "eudicots_odb10" and "embryophyta_odb10" databases (Manni et al., 2021). The LAI was utilized by the dataset of LTR in TE annotations below (Ou et al., 2018). QUAST v5.2.0 was used calculate the basic information of N50 and other parameters (Gurevich et al., 2013).

To find telomeric repeats, used tidk 0.2.0 to look for (TTTAGGG)n in the genome (https://github.com/tolkit/telomeric-identifier). To identify centromere localization, we used TRF (Benson, 1999) to identify the length and repeat of whole-genome tandem repeats monomers, we found that the 107 bp repeat monomers appeared continuously and with high frequency in centromeric regions. We referred to the results of TRF, TE annotation, and gene annotation to jointly determine the centromere region.

Comparative genomics between haplotypes and individuals

Whole-genome alignments used the genome PN_T2T (Shi et al., 2023), *V. retordii* haplotype 1 and haplotype 2 to obtain genome synteny. We used minimap2 with the parameters "-ax asm5 --eqx" and nucmer (Kurtz et al., 2004) from MUMmer v4.00rc1 with the parameters "--mum -p", respectively. Then SyRI 1.6 (Goel et al., 2019) was used to detect structural variants, cuteSV 2.0.2 (Jiang et al., 2020) was also used to call SV from HiFi reads. Finally, a genome comparison was plotted using plotsr (Goel and Schneeberger, 2022).

Genome annotation

Following https://github.com/zhouyflab/Genome-Wide-Annotation-Pipeline, the annotation pipeline, TEs were annotated using the Extensive de novo TE Annotator (EDTA) v2.0.0 including long terminal repeats (LTR), terminal inverted repeats (TIR), long interspersed nuclear element (LINE) repeats, Helitron-like DNA transposons, and so on (Ou et al., 2019). Then, repeat elements sequences were hard and soft masked using RepeatMasker version open-4.1.2 (Tarailo-Graovac and Chen, 2009).

RNA-seq data were also collected using the sample to sequence the genome; different plant tissues including *V. retordii* plant stems, leaves, and buds were utilized to generate more than ~10.31 Gb of Illumina raw data. First, for raw data we used fastp version 0.23.2 (Chen et al., 2018) to remove adapter sequences and low-quality sequence reads so that those clean reads could be mapped to the *V. retordii* reference genome using HISAT2 v.2.2.1 (Kim et al., 2019). Second, Stringtie v.1.3.3 was used for transcriptome assembly (Pertea et al., 2015), in addition de novo prediction and homologous gene prediction were performed using Trinity v.2.4.0 (Haas et al., 2013). Third, GffRead v0.12.7 (Pertea and Pertea, 2020) was used to generate the transcript sequence and BUSCO was used to confirm that the assembly result was fine. Program to Assemble Spliced Alignments (PASA) Release v2.5.2 (https://github.com/PASApipeline) was used throughout the gene annotation pipeline to correct gene predictions by combining them with the transcriptome (Haas et al., 2003).

Based on the above process and generated data, the Maker2 v.3.01.03 pipeline was applied to combine those predicted gene methods (Holt and Yandell, 2011). With EST and protein evidence, on the one hand SNAP v.2006-07-28 was used (Korf, 2004), which utilized the HMM algorithm, to obtain a preliminary gene model. On the other hand, AUGUSTUS v.3.4.0 (https://github.com/Gaius-Augustus/Augustus) was used to generate predicted gene models as well (Stanke et al., 2006). Maker was run at least two rounds. Finally, the final annotation results removed duplicates, verified, and corrected by PASA, and renamed, and sorted by a series of scripts.

Functional annotation was first performed using eggNOG-mapper version emapper-2.1.9 (Cantalapiedra et al., 2021), based on eggNOG orthology data (Huerta-Cepas et al., 2019). Second, we used parse_eggNOG.py to parse KEGG and GO results (https://github.com/Hua-CM/HuaSmallTools) from the eggNOG 5.0 web server results to a table format, and used the R package *AnnotationForge* version 1.38.1 build SQLite-based annotation data (https://bioconductor.org/packages/AnnotationForge). Finally, the subsequent enrichment was performed using R package *clusterProfiler* version 4.0 and plotted with it (Wu et al., 2021). Here, 28,714 (85.75%) protein-coding genes were functionally annotated using eggNOG with more genes and genomes than existing databases.

SNP and SV calling and filtering

The quality of the whole-genome resequencing generated raw reads was controlled using FastQC version 1.0.0 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) with default parameters. After quality control, the total bases of every *V. retordii* whole-genome resequencing were greater than 20 Gb. We used haplotype 1 of the woolly grape high-quality genome as the reference as its quality was slightly higher. Clean reads were mapped to the reference genome by BWA-MEM version 0.7.17-r1188 (Li and Durbin, 2009), the output.bam files were sorted and duplicates marked using SAMtools version 1.13 (Li and Handsaker, Wysoker, et al., 2009). Finally the GTX "vc" function was used with default parameters to generate SNPs and short InDels in the in.gvcf format (http://www.genetalks.com/gtxlab.html). Then the BCFTools v.1.13 "merge" function was used to change merge.gvcf files to.vcf files (Li, 2011). VCFTools v.0.1.16 was used filter.vcf files to generate high quality and confidence, with parameters "--max-missing 0.8 --maf 0.01 --minQ 200"

(Danecek et al., 2011). Besides the 33 *V. retordii* samples mentioned above, three *Muscadinia* grapevine resequencing samples had been previously published in NCBI SRR094945, SRR6729330, and SRR6729331 were used as the outgroup.

SV calling of the whole-genome resequencing used DELLY v1.1.6 (Rausch et al., 2012), with the *V. retordii* genome as the reference genome. The "call" and "merge" functions were used first to generate the.bcf files, then the DELLY "call" function was used again with the "-v" parameter and the merged. bcf files. Subsequently the BCFTools v.1.13 "merge" function was used for the merged. bcf files. Finally the DELLY "filter" function with "-f germline" parameter was used to retain the SVs that passed the quality filters, following the methods described in a previous grapevine study (Zhou et al., 2019a).

Population structure

After sample filtering, VCFTools was used to convert the vcf format to a. bed format, PLINK v1.90b6.21 parameter "--indep-pairwise 100 10 0.2" was used to generate LD reports and then parameter "--extract" was used to extract LD $r^2 < 0.2$ and generate the SNP-filtered LD. bed files (Chang et al., 2015). Principal component analysis (PCA) based on the SNP-filtered LD data was performed using the PLINK with "--make-rel --pca" parameter. ADMIXTURE version 1.3.0 was used to estimate individual admixture proportions from ancestors (Alexander et al., 2009). A phylogenetic tree at the individual level was also based on the SNP-filtered LD data, and was performed using PLINK with the "--distance 1-ibs flat-missing" parameter, generating a genetic distance matrix, and was constructed using the Neighbor-Joining Algorithm with MEGA7 (Kumar et al., 2016). We also used IQ-TREE v.2.1.4 to construct a phylogenetic tree using the maximum likelihood method (Minh et al., 2020); it was found that the phylogenetic tree formed by the two methods was consistent.

Population nucleotide diversity (π), quantified population genetic differentiation (F_{ST}) and Tajima's *D* values were calculated using VCFTools, with 50 kb sliding windows and a 10 kb step size. Identity by descent (IBD) was calculated by PLINK. The LD of each population was calculated using PopLDdecay v3.42 (Zhang et al., 2019).

Population demographic history

To analyze population migration and introgression events, PLINK was used to calculate each population allele frequency with parameter "--freq --within", then converted by plink2treemix.py (https://github.com/thomnelson/tools) to the TreeMix input format. Then TreeMix v.1.12 was used to construct a maximum likelihood population tree with parameter "-k 500 -noss" (Pickrell and Pritchard, 2012), which meant setting the block size to 1000 SNPs to estimate the covariance matrix, closed correcting the number of samples of individuals among different populations. In addition, we set three *Muscadinia* grapevine samples as the outgroup and the migration edges (migration events) from 1 to 5, each repeated five times when TreeMixwas run. The R package *OptM* was used to estimate the optimal number of migrations edges to add to the tree (Fitak, 2021). The best result from TreeMix was plotted using the plotting_funcs.R (https://github.com/joepickrell) script. The "threepop" and "fourpop" functions of TreeMix were used to calculate the F3 and F4 statistics. Dsuite v0.5 r44 was used to analyze introgression events, the "Dinvestigate" function was used in the whole genome to calculate f_d and D values, with 50 SNP sliding windows and five SNP step sizes (Malinsky et al., 2021).

To further analyse divergence and population fluctuations during the demographic history. SMC++ (https://github.com/popgenmethods/smcpp) is a program for estimating the size history of populations from whole-genome sequence data (Terhorst et al., 2017). The mutation rate and generation time was set as 7e-9 and 2.5 years (Xiao et al., 2023), respectively. SMC++ v1.15.2 was used first to estimate the histories of a single population using the "vcf2smc", "estimate" and "plot" functions; the population split time was estimated simply using the SMC++ "split" function. Demographic models were established and selected by fastsimcoal27 (http://cmpg.unibe.ch/software/fastsimcoal27), with parameters "-d -n 100000 -L 50 -s 0 -M -c 16 -C 10 -B 40" (Excofffier et al., 2021). The fastsimcoal27 was based on the genome-wide 2D site frequency spectra (2D-SFS), therefore to calculate the 2DSFS the superSFS script was used (https://github.com/xhchauvet/superSFS) first with the set of three *Muscadinia* grapevine samples as the outgroup. The darker color near the $y = x$ axis regions in these figures suggested that there were significant gene flows between these three populations. Simulations were repeated 100 times to acquire the closest maximum likelihood estimates compared with the observed 2D-SFS. The best model selection was based on the AIC scores, which were calculated using calculateAIC.sh (https://github.com/speciationgenomics).

The histories of the two populations were estimated at first using fastsimcoal27 with four different models, which were defined as no migration, unidirectional migration (from population 1 to population 2 and from population 2 to population 1), and bidirectional migration. However, the AIC scores of simple models in each pair of the three populations were too close to distinguish which model was better (Figures S11, S12, S13). Then conservative estimates of every two-population results and SMC++ results were considered to avoid overfitting. It could be concluded whether or not there were gene flows among those three populations, and there was a high probability that bottleneck events had occurred recently (Figure S14). To verify whether a bottleneck had actually occurred, the occurrence of a bottleneck in the two-population model without gene flow was further simulated, the result showed that the fitting degree was significantly improved (Model 5 in Table S5). Finally, the three population joint demographic history was estimated using fastsimcoal27. Observed and excepted joint 2D-SFS for each

pair of three populations of *V. retordii* in this study was also plotted using sfstool.R script (https://github.com/marqueda/SFS-scripts).

Population selection detection

Our study detected outlier selection events between coastal and inland populations by population nucleotide diversity (π) and quantified population genetic differentiation (F_{ST}) calculated by SVs. Based on the top 5% of F_{ST} and the top π ratio between coastal and inland populations, we investigated the population-specific differentiation between coastal population and inland populations. Population branch statistics (PBS) is a new method that considers both the population genetic differentiation index (F_{ST}) and the mean pairwise nucleotide differences in intergroup comparisons (d_{XY}), which could detect selective sweeps among two closely related populations and an outgroup (Yi et al., 2010). PBS analysis combined F_{ST} and d_{XY}, which represented the relative and the absolute allele frequency differentiation, respectively. This made measurements more reliable than conducted by F_{ST} alone, as this might be inflated by a reduction within population nucleotide diversities. In this study, PBScan was used (https://github.com/thamala/PBScan) to perform PBS analysis (Hamala and Savolainen, 2019). There were three *V. retordii* populations in LC, LA, and SZ, as well as a *Muscadinia* outgroup. PBS values were calculated for each of the two populations of *V. retordii* and *Muscadinia* during outlier detection. The program ms was used to generate neutral sample data under neutral models (Hudson, 2002), which considered the genome-wide recombination rates and the demographic history of these populations in this study. With more than 50,000 neutral sample data, each population obtained the *p*-values, which were defined as the proportion of neutral estimates that had the same or higher PBS value than the observed one by PBScan. The *p*-values < 0.05 were considered significant. Finally, the smallest 1% *p*-values locations were regarded as selective signals. We also used SWeeD version 4.0.0 to detect genomic signatures of selection among these three populations with a sliding window size of 1 kb (Pavlidis et al., 2013).

Population deleterious mutations

Detection of deleterious mutations used the SIFT 2.0.0 algorithm (Vaser et al., 2016) to detect the effect of single nucleotide non-synonymous mutation on protein function. We used the *V. retordii* genome and annotation to make a SIFT database with SIFT (https://github.com/pauline-ng/SIFT4G_Create_Genomic_DB) predictions, then applied the database to annotate the SNP files, which reversed according to the ancestor state inferred from the outgroup *Muscadinia* population, with SIFT4G_Annotator.jar.

Data availability statement

All data used in this study are publicly available as described in Materials and Methods. All PacBio sequence data have been deposited in the NCBI Sequence Read Archive under project number PRJNA953920 (genome) and PRJNA1082915 (WGS). Genome assembly is also available from the National Genomics Data Center (NGDC) under project number PRJCA024068. The genome assembly and annotation is also available on the website (https://zenodo.org) under number 10842451.

ACKNOWLEDGEMENTS

This work was supported by the Science Fund Program for Distinguished Young Scholars of the National Natural Science Foundation of China (Overseas) to Yongfeng Zhou, National Natural Science Foundation of China (Nos. 32300191; 32372662), Guangxi University, Bama Institute of Integration of Industry and Education, postgraduate joint training project (Project Nos. 20210020; 20210039), the National Key Research and Development Program of China (grants 2023YFF1000100 and 2023YFD2200700). We are also particularly grateful for the services of the High-Performance Computing Cluster and experimental sites at the Agricultural Genomics Institute in Shenzhen, Chinese Academy of Agricultural Sciences.

CONFLICT OF INTEREST

The authors have no conflict of interest.

AUTHOR CONTRIBUTIONS

Y.Z., Z.M., and X.X. designed the project, and G.-F.J. also provided assistance for the project. T.Z. and W.P. conducted all bioinformatics analyses described in the article and other main work. S.C. performed genome annotation. Z.C., X.S., Y.L., and Y.P. made contributions to visualization and formal analysis. T.Z., X.W., X.S., Z.M., and X.X. carried out the woolly grape material collection. Z.L. and X.S., T.Z., Z.M., X.X., and H.X. helped with writing the original draft. Y.Z., G.J., X.X., and Z.M. revised the manuscript. All the authors read and approved the final manuscript.

Edited by: Dr. Jianquan Liu.

Received Sep. 27, 2023; **Accepted** Mar. 4, 2024

REFERENCES

Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. **19:** 1655–1664.

Alkan, C., Coe, B.P., and Eichler, E.E. (2011). Genome structural variation discovery and genotyping. Nat. Rev. Genet. **12:** 363–376.

Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., Wang, X., Lippman, Z.B., Schatz, M.C., and Soyk, S. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biol. **23:** 258.

Alonge, M., Wang, X., Benoit, M., Soyk, S., Pereira, L., Zhang, L., Suresh, H., Ramakrishnan, S., Maumus, F., and Ciren, D. (2020). Major impacts of widespread structural variation on gene expression and crop improvement in tomato. Cell **182:** 145–161.e23

Bân, N.T. (2003). Danh lục các loài thực vật Việt Nam. Tập. 1: 3.

Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Res. **27:** 573–580.

Branties N. (1978). Pollinator attraction of *Vitis vinifera* subsp. *silvestris*.

Cantalapiedra, C.P., Hernandez-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). EggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol. Biol. Evol. **38:** 5825–5829.

Carrasco, D., Zhou-Tsang, A., Rodriguez-Izquierdo, A., Ocete, R., Revilla, M.A., and Arroyo-García, R. (2022). Coastal wild grapevine accession (*Vitis vinifera* L. ssp. *sylvestris*) shows distinct late and early transcriptome changes under salt stress in comparison to commercial rootstock richter 110. Plants **11:** 2688.

Chang, C.C., Chow, C.C., Tellier, L., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience **4:** 16.

Chen, I., and Manchester, S.R. (2007). Seed morphology of modern and fossil *Ampelocissus* (Vitaceae) and implications for phytogeography. Am. J. Bot. **94:** 1534–1553.

Chen, J., Wang, Z., Tan, K., Huang, W., Shi, J., Li, T., Hu, J., Wang, K., Wang, C., Xin, B., et al. (2023). A complete telomere-to-telomere assembly of the maize genome. Nat. Genet. **55:** 1221–1231.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics **34:** i884–i890.

Chen, Z., Ren, H., Wen, J., and Wu, C. (2007). Vitaceae. Flora China **12:** 173–222.

Chin, C.S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G. T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. Nat. Methods **13:** 1050–1054.

Choat, B., Jansen, S., Brodribb, T.J., Cochard, H., Delzon, S., Bhaskar, R., Bucci, S.J., Feild, T.S., Gleason, S.M., and Hacke, U.G. (2012). Global convergence in the vulnerability of forests to drought. Nature **491:** 752–755.

Collins, R.L., Brand, H., Karczewski, K.J., Zhao, X., Alföldi, J., Francioli, L.C., Khera, A.V., Lowther, C., Gauthier, L.D., and Wang, H. (2020). A structural variation reference for medical and population genetics. Nature **581:** 444–451.

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., et al. (2011). The variant call format and VCFtools. Bioinformatics **27:** 2156–2158.

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

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

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

Edelman, N.B., Frandsen, P.B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R.B., García-Accinelli, G., Van Belleghem, S.M., Patterson, N., Neafsey, D.E., et al. (2019). Genomic architecture and introgression shape a butterfly radiation. Science **366:** 594–599.

Excofffier, L., Marchi, N., Marques, D.A., Matthey-Doret, R., Gouy, A., and Sousa, V.C. (2021). Fastsimcoal2: Demographic inference under complex evolutionary scenarios. Bioinformatics **37:** 4882–4885.

Fang, Z., Pyhäjärvi, T., Weber, A.L., Dawe, R.K., Glaubitz, J.C., González, J.d.J.S., Ross-Ibarra, C., Doebley, J., Morrell, P.L., and Ross-Ibarra, J. (2012). Megabase-scale inversion polymorphism in the wild ancestor of maize. Genetics **191:** 883–894.

Fitak, R.R. (2021). OptM: Estimating the optimal number of migration edges on population trees using Treemix. Biol. Meth. Protoc. **6:** 6.

Garcia-Molina, A., Altmann, M., Alkofer, A., Epple, P.M., Dangl, J.L., and Falter-Braun, P. (2017). LSU network hubs integrate abiotic and biotic stress responses via interaction with the superoxide dismutase *FSD2*. J. Exp. Bot. **68:** 1185–1197.

Gaut, B.S., Seymour, D.K., Liu, Q., and Zhou, Y. (2018). Demography and its effects on genomic variation in crop domestication. Nat. Plants **4:** 512–520.

Goel, M., and Schneeberger, K. (2022). Plotsr: Visualizing structural similarities and rearrangements between multiple genomes. Bioinformatics **38:** 2922–2926.

Goel, M., Sun, H.Q., Jiao, W.B., and Schneeberger, K. (2019). SyRI: Finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. **20:** 13.

Gu, Z., Pan, S., Lin, Z., Hu, L., Dai, X., Chang, J., Xue, Y., Su, H., Long, J., and Sun, M. (2021). Climate-driven flyway changes and memory-based long-distance migration. Nature **591:** 259–264.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. Bioinformatics **29:** 1072–1075.

Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, Jr, R.K., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., and Town, C.D. (2003). Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. **31:** 5654–5666.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., and Lieber, M. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. **8:** 1494–1512.

Hager, E.R., Harringmeyer, O.S., Wooldridge, T.B., Theingi, S., Gable, J.T., McFadden, S., Neugeboren, B., Turner, K.M., Jensen, J.D., and Hoekstra, H.E. (2022). A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes. Science **377:** 399–405.

Hall, M., Lowry, D., and Willis, J. (2010). Is local adaptation in *Mimulus guttatus* caused by trade-offs at individual loci? Mol. Ecol. **19:** 2739–2753.

Hamala, T., and Savolainen, O. (2019). Genomic patterns of local adaptation under gene flow in *Arabidopsis lyrata*. Mol. Biol. Evol. **36:** 2557–2571.

Hämälä, T., Wafula, E.K., Guiltinan, M.J., Ralph, P.E., dePamphilis, C.W., and Tiffin, P. (2021). Genomic structural variants constrain and facilitate adaptation in natural populations of *Theobroma cacao*, the chocolate tree. Proc. Natl. Acad. Sci. U.S.A. **118:** e2102914118.

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

Holt, C., and Yandell, M. (2011). MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinf. **12:** 1–14.

Huang, K., Ostevik, K.L., Elphinstone, C., Todesco, M., Bercovich, N., Owens, G.L., and Rieseberg, L.H. (2022). Mutation load in sunflower inversions is negatively correlated with inversion heterozygosity. Mol. Biol. Evol. **39:** msac101.

Hudson, R.R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics **18:** 337–338.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernandez-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. **47:** D309–D314.

Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B.M., Vinckenbosch, N., Liang, Y., Yi, X., He, M., Somel, M., et al. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature **512:** 194–197.

Jiang, T., Liu, Y.Z., Jiang, Y., Li, J.Y., Gao, Y., Cui, Z., Liu, Y.D., Liu, B., and Wang, Y.D. (2020). Long-read-based human genomic structural variation detection with cuteSV. Genome Biol. **21:** 189.

Kawecki, T.J., and Ebert, D. (2004). Conceptual issues in local adaptation. Ecol. Lett. **7:** 1225–1241.

Kern, A.D., and Hahn, M.W. (2018). The neutral theory in light of natural selection. Mol. Biol. Evol. **35:** 1366–1371.

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

Kimura, M. (1968). Evolutionary rate at the molecular level. Nature **217:** 624–626.

Korf, I. (2004). Gene finding in novel genomes. BMC Bioinf. **5:** 1–9.

Kou, Y., Liao, Y., Toivainen, T., Lv, Y., Tian, X., Emerson, J., Gaut, B.S., and Zhou, Y. (2020). Evolutionary genomics of structural variation in Asian rice (*Oryza sativa*) domestication. Mol. Biol. Evol. **37:** 3507–3524.

Krupovic, M., and Koonin, E.V. (2015). Polintons: A hotbed of eukaryotic virus, transposon and plasmid evolution. Nat. Rev. Microbiol. **13:** 105–115.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. **33:** 1870–1874.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. Genome Biol. **5:** 9.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics **27:** 2987–2993.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data, P. (2009). The sequence alignment/Map format and SAMtools. Bioinformatics **25:** 2078–2079.

Liang, Z.C., Duan, S.C., Sheng, J., Zhu, S.S., Ni, X.M., Shao, J.H., Liu, C.H., Nick, P., Du, F., Fan, P.G., et al. (2019). Whole-genome resequencing of 472 *Vitis* accessions for grapevine diversity and demographic history analyses. Nat. Commun. **10:** 12.

Liu, S.Y., Zhang, L., Sang, Y.P., Lai, Q., Zhang, X.X., Jia, C.F., Long, Z. Q., Wu, J.L., Ma, T., Mao, K.S., et al. (2022). Demographic history and natural selection shape patterns of deleterious mutation load and barriers to introgression across populus genome. Mol. Biol. Evol. **39:** 16.

Lowry, D.B., and Willis, J.H. (2010). A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. PLoS Biol. **8:** e1000500.

Ma, Z.Y., Nie, Z.L., Liu, X.Q., Tian, J.P., Zhou, Y.F., Zimmer, E., and Wen, J. (2022). Phylogenetic relationships, hybridization events, and drivers of diversification of East Asian wild grapes as revealed by phylogenomic analyses. J. Syst. Evol. **61:** 273–283

Ma, Z.Y., Wen, J., Ickert-Bond, S.M., Nie, Z.L., Chen, L.Q., and Liu, X.Q. (2018). Phylogenomics, biogeography, and adaptive radiation of grapes. Mol. Phylogenet. Evol. **129:** 258–267.

Malhi, Y., and Wright, J. (2004). Spatial patterns and recent trends in the climate of tropical rainforest regions. Philos. T. R. Soc. B. **359:** 311–329.

Malinsky, M., Matschiner, M., and Svardal, H. (2021). Dsuite - Fast D-statistics and related admixture evidence from VCF files. Mol. Ecol. Resour. **21:** 584–595.

Manni, M., Berkeley, M.R., Seppey, M., Simao, F.A., and Zdobnov, E.M. (2021). BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol. Biol. Evol. **38:** 4647–4654.

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

Mills, R.E., Walter, K., Stewart, C., Handsaker, R.E., Chen, K., Alkan, C., Abyzov, A., Yoon, S.C., Ye, K., and Cheetham, R.K. (2011). Mapping copy number variation by population-scale genome sequencing. Nature **470:** 59–65.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. **37:** 1530–1534.

Mitchell-Olds, T., Willis, J.H., and Goldstein, D.B. (2007). Which evolutionary processes influence natural genetic variation for phenotypic traits? Nat. Rev. Genet. **8:** 845–856.

Moore, M., and Wen, J., Committee, F. o. N. A. E. (2016). Vitaceae. FNANM. **12:** 3–13.

Newman, M., Ketphanh, S., Svengsuksa, B., Thomas, P., Sengdala, K., Lamxay, V., and Armstrong, K. (2007). A checklist of the vascular plants of Lao PDR: Royal Botanic Garden Edinburgh.

Ou, S., Chen, J., and Jiang, N. (2018). Assessing genome assembly quality using the LTR Assembly Index (LAI). Nucleic Acids Res. **46:** e126.

Ou, S.J., Su, W.J., Liao, Y., Chougule, K., Agda, J.R.A., Hellinga, A.J., Lugo, C.S.B., Elliott, T.A., Ware, D., Peterson, T., et al. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. **20:** 18.

Oziolor, E.M., Reid, N.M., Yair, S., Lee, K.M., Guberman VerPloeg, S., Bruns, P.C., Shaw, J.R., Whitehead, A., and Matson, C.W. (2019). Adaptive introgression enables evolutionary rescue from extreme environmental pollution. Science **364:** 455–457.

Pavlidis, P., Živković, D., Stamatakis, A., and Alachiotis, N. (2013). SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. Mol. Biol. Evol. **30:** 2224–2234.

Pertea, G., and Pertea, M. (2020). GFF utilities: GffRead and GffCompare. F1000Res. **9:** 304.

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

Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. **8:** 17.

Rausch, T., Zichner, T., Schlattl, A., Stutz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: Structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics **28:** I333–I339.

Santiago, J.P., and Tegeder, M. (2016). Connecting source with sink: The role of *Arabidopsis AAP8* in phloem loading of amino acids. Plant Physiol. **171:** 508–521.

Savolainen, O., Lascoux, M., and Merilä, J. (2013). Ecological genomics of local adaptation. Nat. Rev. Genet. **14:** 807–820.

Shackelford, G., Steward, P.R., Benton, T.G., Kunin, W.E., Potts, S. G., Biesmeijer, J.C., and Sait, S.M. (2013). Comparison of pollinators and natural enemies: A meta-analysis of landscape and local effects on abundance and richness in crops. Biol. Rev. **88:** 1002–1021.

Shang, L., He, W., Wang, T., Yang, Y., Xu, Q., Zhao, X., Yang, L., Zhang, H., Li, X., Lv, Y., et al. (2023). A complete assembly of the rice Nipponbare reference genome. Mol. Plant **16:** 1232–1236.

Shang, L.G., Li, X.X., He, H.Y., Yuan, Q.L., Song, Y.N., Wei, Z.R., Lin, H., Hu, M., Zhao, F.L., Zhang, C., et al. (2022). A super pan-genomic landscape of rice. Cell Res. **32:** 878–896.

Shi, X., Cao, S., Wang, X., Huang, S., Wang, Y., Liu, Z., Liu, W., Leng, X., Peng, Y., Wang, N., et al. (2023). The complete reference genome for grapevine (*Vitis vinifera* L.) genetics and breeding. Hortic. Res. **10:** uhad061.

Silverstone, A.L., Ciampaglio, C.N., and Sun, T.-p (1998). The *Arabidopsis RGA* gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. Plant Cell **10:** 155–169.

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

Stephan, W. (2010). Genetic hitchhiking versus background selection: The controversy and its implications. Philos. T. R. Soc. B. **365:** 1245–1253.

Sudmant, P.H., Rausch, T., Gardner, E.J., Handsaker, R.E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., and Hsi-Yang Fritz, M. (2015). An integrated map of structural variation in 2,504 human genomes. Nature **526:** 75–81.

Suzuki, T., Iizuka, Y., Matsuoka, K., Furukawa, T., Kamiyama, K., and Watanabe, O. (2002). Distribution of sea salt components in snow cover along the traverse route from the coast to Dome Fuji station 1000 km inland at east Dronning Maud Land. Antarctica. Tellus. B. **54:** 407–411.

Takou, M., Hämälä, T., Koch, E.M., Steige, K.A., Dittberner, H., Yant, L., Genete, M., Sunyaev, S., Castric, V., Vekemans, X., et al. (2021). Maintenance of adaptive dynamics and no detectable load in a range-edge outcrossing plant population. Mol. Biol. Evol. **38:** 1820–1836.

Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics **4:** 1–4.

Terhorst, J., Kamm, J.A., and Song, Y.S. (2017). Robust and scalable inference of population history froth hundreds of unphased whole genomes. Nat. Genet. **49:** 303–309.

Todesco, M., Owens, G.L., Bercovich, N., Légaré, J.-S., Soudi, S., Burge, D.O., Huang, K., Ostevik, K.L., Drummond, E.B., and Imerovski, I. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. Nature **584:** 602–607.

Vaser, R., Adusumalli, S., Leng, S.N., Sikic, M., and Ng, P.C. (2016). SIFT missense predictions for genomes. Nat. Protoc. **11:** 1–9.

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

Wang, Z., Xue, J.-Y., Hu, S.-Y., Zhang, F., Yu, R., Chen, D., Van de Peer, Y., Jiang, J., Song, A., and Ni, L. (2022). The genome of hibiscus hamabo reveals its adaptation to saline and waterlogged habitat. Hortic. Res. **9:** uhac067.

Wen-Tsai, W. (1979). Vitacearum novitates. J. Syst. Evol. **17:** 73–96.

Wu, T.Z., Hu, E.Q., Xu, S.B., Chen, M.J., Guo, P.F., Dai, Z.H., Feng, T.Z., Zhou, L., Tang, W.L., Zhan, L., et al. (2021). ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation **2:** 100141.

Xiao, H., Liu, Z., Wang, N., Long, Q., Cao, S., Huang, G., Liu, W., Peng, Y., Riaz, S., Walker, A.M., et al. (2023). Adaptive and maladaptive introgression in grapevine domestication. Proc. Natl. Acad. Sci. U.S.A. **120:** e2222041120.

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X.P., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T.S., et al. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. Science **329:** 75–78.

Zhang, C., Dong, S.S., Xu, J.Y., He, W.M., and Yang, T.L. (2019). PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics **35:** 1786–1788.

Zhang, Y., Peng, Y., Liu, J., Yan, J., Zhu, K., Sun, X., Bu, X., Wang, X., Ahammed, G.J., and Liu, Y. (2023). Tetratricopeptide repeat protein SIREC2 positively regulates cold tolerance in tomato. Plant Physiol. **192:** kiad085.

Zhou, Y., Massonnet, M., Sanjak, J.S., Cantu, D., and Gaut, B.S. (2017). Evolutionary genomics of grape (*Vitis vinifera* ssp. *vinifera*) domestication. Proc. Natl. Acad. Sci. U.S.A. **114:** 11715–11720.

Zhou, Y., Minio, A., Massonnet, M., Solares, E., Lv, Y., Beridze, T., Cantu, D., and Gaut, B.S. (2019a). The population genetics of structural variants in grapevine domestication. Nat. Plants **5:** 965–979.

Zhou, Y., Muyle, A., and Gaut, B.S. (2019b). Evolutionary genomics and the domestication of grapes. The Grape Genome, Dario Cantu and M. Andrew Walker (Eds). **3:** 39–55.

Zhou, Y., Zhang, Z., Bao, Z., Li, H., Lyu, Y., Zan, Y., Wu, Y., Cheng, L., Fang, Y., Wu, K., et al. (2022). Graph pangenome captures missing heritability and empowers tomato breeding. Nature **606:** 527–534.

Zhou, Y.F., and Gaut, B.S. (2020). Large chromosomal variants drive adaptation in sunflowers. Nat. Plants **6:** 734–735.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article: http://onlinelibrary.wiley.com/doi/10.1111/jipb.13653/suppinfo

Figure S1. The *Vitis retordii* brief introduction of habitat and morphology
Figure S2. Estimate of the genome size and complexity of woolly grape by the k-mers method
Figure S3. Hi-C map in Juicebox of woolly grape two haplotypes
Figure S4. Display of chromosome centromere region
Figure S5. Telomere region (TTTAGGG)n repeats count and centromeric regions of woolly grape
Figure S6. The chromosome 18 Hi-C heat map between woolly grape haplotypes 1 and 2
Figure S7. Whole-genome distributions of nucleotide diversity (π) and Tajima's *D*
Figure S8. Decay of linkage disequilibrium (LD) of three populations measured by the genome-wide average r^2
Figure S9. PCA analysis with outgroup *Muscadinia* (Yuanye) population
Figure S10. Observed joint 2D-SFS for each pair of three populations
Figure S11. Two-population models simulated by fastsimcoal27 in population LA and LC
Figure S12. Two-population models simulated by fastsimcoal27 in population LA and SZ
Figure S13. Two-population models simulated by fastsimcoal27 in population LC and SZ
Figure S14. Population fluctuations during the demographic history by SMC++ estimated
Figure S15. Three population models simulated by fastsimcoal27
Figure S16. Observed and excepted joint 2D-SFS of the best fit model
Figure S17. TreeMix trees with one migration event from population LA (inland) to SZ (coastal)
Figure S18. Whole-genome scatter plot of f_{dM} values performed by Dsuite D-statistics
Figure S19. Distribution of the five SVs types of proportion between the coastal and inland populations
Figure S20. The five types of SVs proportion between the coastal and inland populations

Figure S21. GO enrichment analysis with outlier genes in coastal population by SVs
Figure S22. Observed population branch statistics (PBS) and *p*-value distributions
Figure S23. GO enrichment analysis with the same outlier genes of three populations
Figure S24. SVs and SNPs burdens between coastal and inland populations
Figure S25. Breakpoint locations of the outlier inversion in Chromosome 18
Figure S26. The inland populations had deletion events on chromosome 7 downstream of the *RGA1* gene
Figure S27. The coastal and inland population possible protein three-dimensional structural prediction of *RGA* genes
Figure S28. Identification of hemizygous genes
Table S1. Woolly grape collect information
Table S2. Woolly grape reference genome assembly information
Table S3. Types and percentages of different transpositions
Table S4. Distribution of centromere region on chromosome
Table S5. Comparison of demographic models analyzed with fastsimcoal27
Table S6. The best coalescent model parameters
Table S7. Enrichment results of inland population-specific selective genes by PBS analysis (sorted by *p*-value top 20)
Table S8. Enrichment results of coastal population-specific selective genes by PBS analysis (sorted by *p*-value top 20)
Table S9. Enrichment results of the same selective genes in coastal and inland populations by PBS analysis (sorted by *p*-value top 20)
Table S10. SVs and SNPs burden
Table S11. Enrichment results of the different selective genes between coastal and inland populations by F_{ST}_SV analysis (sorted by *p*-value top 20)
Table S12. Enrichment analysis of the centromeric region in woolly grape chromosome 18
Table S13. Woolly grape resequence information
Table S14. Woolly grape and recently published plants quality comparison of complete genome

Scan using WeChat with your smartphone to view JIPB online

Scan with iPhone or iPad to view JIPB on Twitter