

Haplotype-resolved assemblies provide insights into genomic makeup of the oldest grapevine cultivar (Munage) in China

Haixia Zhong^{1,†}, Xiaoya Shi^{2,3,4,†}, Fuchun Zhang^{1,†}, Jin Yao⁵, Xu Wang^{2,4} , Vivek Yadav¹ , Xiaoming Zhou¹, Shuo Cao⁴, Songlin Zhang¹, Chuan Zhang¹, Jiangxia Qiao¹, Zhongjie Liu⁴ , Yingchun Zhang⁴, Yuting Liu^{2,4} , Hao Wang¹, Hui Xue⁴ , Mengyan Zhang^{2,4}, Tianhao Zhang^{2,4}, Wenrui Wang^{2,3}, Ruoyan Zhao², Tianrong Fan², Zhongqi Liu², Jiawei Li² , Ruirui Liu⁶ , Yongfeng Zhou^{2,4,*} , Ling Tian^{5,*}, Xinyu Wu^{1,*} and Hua Xiao^{1,4,*} 

¹Institute of Fruits and Vegetables, Xinjiang Uygur Autonomous Region Academy of Agricultural Sciences, Biological Breeding Laboratory, Xinjiang Uygur Autonomous Region Academy of Agricultural Sciences, The State Key Laboratory of Genetic Improvement and Germplasm Innovation of Crop Resistance in Arid Desert Regions (Preparation), Key Laboratory of Genome Research and Genetic Improvement of Xinjiang Characteristic Fruits and Vegetables, Urumqi, China

²State Key Laboratory of Tropical Crop Breeding, Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China

³College of Enology, Heyang Viti-Viniculture Station, Ningxia Helan Mountain's East Foothill Wine Experiment and Demonstration Station, Northwest A&F University, Yangling, China

⁴State 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, China

⁵School of Management, Shenzhen Polytechnic University, Shenzhen, China

⁶Gansu Key Laboratory of Conservation and Utilization of Biological Resources and Ecological Restoration in Longdong, School of Agriculture and Bioengineering, Longdong University, Qingyang, China

* Corresponding authors. E-mail: xiaohua01@caas.cn; wuxy@xaas.ac.cn; tianling@szpu.edu.cn; zhouyongfeng@caas.cn

†These authors contributed equally to this work

Abstract

Munage, an ancient grape variety that has been cultivated for thousands of years in Xinjiang, China, is renowned for its exceptional fruit traits. There are two main types of Munage: white fruit (WM) and red fruit (RM). However, the lack of a high-quality genomic resources has impeded effective breeding and restricted the potential for expanding these varieties to other growing regions. In this study, we assembled haplotype-resolved genome assemblies for WM and RM, alongside integrated whole genome resequencing (WGS) data and transcriptome data to illuminate the origin, private mutations and selection in Munage. Our analyses suggest that Munage likely shares a common ancestor with Eurasian grapes that originated in West Asia. Selective analysis between Munage clones and Eurasian grapes mapped genomic signals of selection in Munage grapes, with genes enriched in processes including cell maturation, plant epidermal cell differentiation, and root epidermal cell differentiation. We also identified 283 somatic mutation sites between WM and RM, along with differential selection on genome and expressed genes. These findings provide crucial genetic resources for investigating the genetics of the ancient Chinese grape variety, Munage, and will facilitate the genetic improvement in grapevine using this ancient cultivar as a gene donor.

Introduction

The cultivated grapevine (*Vitis vinifera* L.) is closely associated with human civilization. Its extensive genetic diversity has enabled its use for a wide range of purposes, while the grapevine has also come to symbolize cultural identity in many regions due to its central role in culinary and social practices [1, 2]. The Eurasian grapes were initially domesticated and cultivated in a region extending from the eastern Mediterranean coast to the Caucasus,

situated between the Caspian and Black Seas [3–7], and were subsequently introduced to Europe, Asia, and other regions [8]. There is no naturally occurring Eurasian grape in China. Historical records indicate that around 100 BC, grapes were introduced to the ancient Chinese capitals via the Silk Road [9, 10]. However, grape remains unearthed from the Yanghai cemetery show that grapes had already been brought into Xinjiang 2300 years ago, which is 200 years earlier than the opening of the ancient Silk Road [11–13].

Received: 2 October 2024. Accepted: 11 October 2025. Published: 20 October 2025. Corrected and Typeset: 1 January 2026

© The Author(s) 2025. Published by Oxford University Press on behalf of the Nanjing Agricultural University.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Xinjiang has a long history of grape cultivation and has thus inherited a rich variety of ancient native grape germplasm resources with irreplaceable genes [14, 15]. Traditional native varieties include Thompson Seedless, Mare's Teat (also known as Milk), and Munage, among others. As one of the most representative and high-quality cultivars in Xinjiang [16, 17], Munage not only possesses a unique flavor but is also highly favored by consumers for its large berries, delicate skin, and juicy flesh, and is thus widely cultivated [18]. Throughout thousands of years of extensive selection and cultivation, many variations have accumulated, including both beneficial and deleterious ones. These variations in germplasm resources are invaluable for breeding and the development of new varieties. However, the trait-specific genes underlying these characteristics have yet to be systematically investigated and utilized.

Munage can be classified into two types based on skin color: red Munage (RM) and white Munage (WM). This ancient variety exhibits unique regional adaptability and stress resistance [19, 20], and its favorable traits and specific genetic resources warrant further exploration and utilization. In addition, it is widely believed that RM grapes originated as a bud mutation of WM, although clear scientific evidence is currently lacking. Bud sport mutation is a common phenomenon in fruit plants and plays a vital role in improving the overall quality of grapevines. For instance, a famous Chinese variety Jinzao Wuhe is a bud sport of Himrod Seedless, characterized by enlarged berrie [21]. Moreover, Thompson Seedless is also considered to be a shoot mutant of Sultanina, which was further distributed and maintained by cuttings [22]. Other bud sports that have been identified and successfully applied in grapevine breeding include early-ripening cultivars, such as 'Fengzao', 'Tiangong Moyu', and 'Nantaihutezao' [23–25], as well as berry skin color mutants such as 'Pinot Grigio', 'Pinot Blanc', and 'Benitaka' [26–28]. The variants observed between WM and RM grapes will provide a valuable opportunity to understand the effects of somatic mutations on grape traits. More recently, genome sequencing in grapes has greatly contributed to our understanding of the structure of the *Vitis* genome [22]. In the past decade, the utility of haplotype assemblies for identifying genes associated with agronomic traits has been well demonstrated in grapevine [22, 29–31].

In this study, we generated high-quality haplotype-resolved genome assemblies of the ancient Chinese grape variety Munage. Comparative genomics was employed to identify selected differential segments and structural variations (SVs) in the genomes of different varietal types of the ancient grape variety Munage. Combining the whole genome sequencing data of Munage and various Eurasian grapes allows us to elucidate the evolutionary relationships and explore the sweep selection on the genome of this ancient grape variety. Furthermore, we analyzed transcriptome data from multiple Munage tissues to identify differentially expressed genes associated with key traits, particularly those related to fruit color. This work improves our understanding of the origin and domestication history of Munage, provides valuable genomic resources for future grape genome research, and offers a foundation for developing new grape germplasm and advancing breeding strategies.

Results

Haplotype-resolved assemblies of Munage genome

The plant materials used in this study consisted of an 80-year-old clonally propagated individual of white Munage (WM) from Xinjiang, China, and a red Munage (RM) individual derived from a mother tree over 30 years old (Fig. 1A). To characterize WM and RM at the genomic level, we performed PacBio sequencing on both varieties, generating 64.27 Gb of HiFi reads for WM and 49.78 Gb of HiFi reads for RM. Hi-C data were generated using the Illumina HiSeq X Ten platform, yielding 52.35 Gb of Hi-C reads for WM and 52.67 Gb of Hi-C reads for RM.

The genome sizes of the two species, determined using the k-mer metric method with HiFi reads, indicated a genome size of ~450 974 828 bp for WM and ~461 596 620 bp for RM (Fig. S1A, Table S1). Then, we performed an initial assembly using HiFiasm, which yielded four haplotypes for the two species at the contig level. The contig N50 values were 24.32 Mb for haplotype 1 of WM (WM1), 26.33 Mb for haplotype 2 of WM (WM2), 15.55 Mb for haplotype 1 of RM (RM1), and 17.08 Mb for haplotype 2 of RM (RM2) (Tables S1 and S2).

Through Hi-C data integration, we obtained haplotype-resolved sequence assemblies for white and red Munage. The sizes of the assemblies were 488 103 118 bp for WM1, 489 354 377 bp for WM2, 489 516 681 bp for RM1, and 480 619 624 bp for RM2 (Table S2). The BUSCO completeness scores were 98.3%, 98.6%, 98.3%, and 98.5%, respectively, indicating high assembly quality (Fig. S1B). There is a highly evident collinearity across the entire genome of WM, RM, and PN_T2T (Fig. 1B), but we also identified some large variations among haplotypes, e.g. a large inversion in the 13- to 15-Mb region of chromosome 11 between two haplotypes of both RM and WM (Fig. 1C–E).

Comparative genomic analysis

Although strong collinearity among the five genomes—WM1, WM2, RM1, RM2, and PN_T2T [32] was observed (Fig. 1B), certain differences were detected in gene composition (Fig. 2A and B). Gene annotation identified 33 942, 34 034, 35 292, and 35 007 genes in WM1, WM2, RM1, and RM2, respectively (Table S2). By comparing the gene annotations among the five haplotypes, 718 orthologous genes unique to PN_T2T were identified (Fig. 2A). These genes were mainly enriched in pathways related to secondary metabolism, defense response, asymmetric cell division, and secondary metabolite biosynthesis during embryonic development (Table S3, Fig. S2A). Further, we identified 22 unique orthologous genes in WM1, which were mainly enriched in pathways such as response to stimulus, response to external biotic stimulus, response to biotic stimulus, biological processes involved in interspecies interactions, and obsolete multi-organism processes (Fig. 2A, Fig. S2B, Table S4). There are 17 unique genes in WM2, but no GO enrichment pathway was identified. The functional annotation of these genes is primarily associated with RNA-mediated transposition, ribonuclease activity, or kinase activity (Table S5). RM1 contained

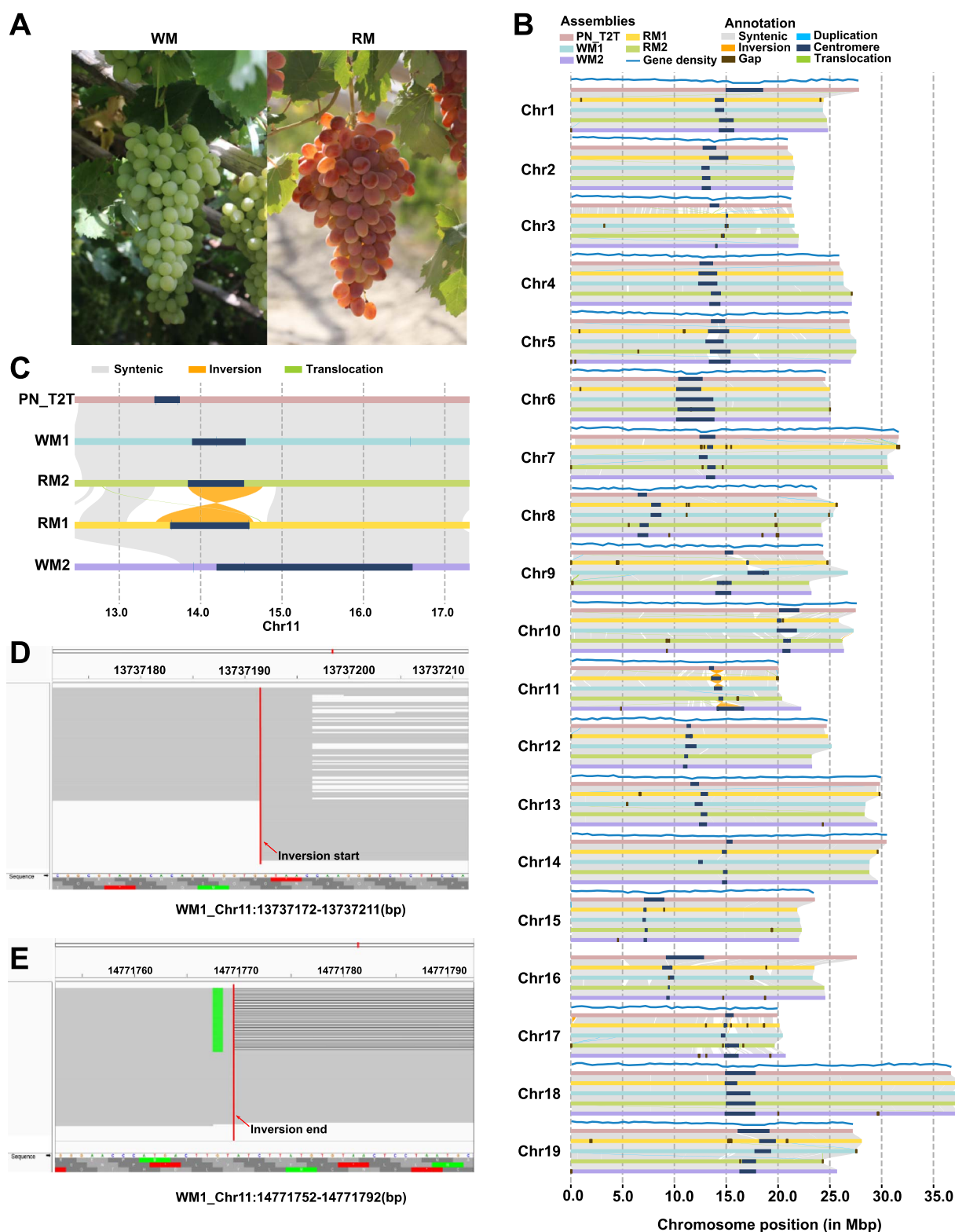


Figure 1 Haplotype-resolved assemblies of the RM and WM genomes. (A) Fruit images of RM and WM. (B) Collinearity alignments between the PN_T2T genome and the WM1, WM2, RM1, and RM2 genomes. (C) Schematic diagram of structural variations in the 13–15 Mb region of chromosome 11. (D-E) IGV visualization for inversion verification. The IGV screenshots show the inverted breakpoints of RM1, with the red vertical line indicating the locations of the breakpoints.

36 unique orthologous genes, which were primarily enriched in pathways such as obsolete cellular nitrogen compound metabolic processes, obsolete RNA polyadenylation, obsolete nitrogen

compound metabolic processes, RNA 3'-end processing, and RNA metabolic processes (Fig. 2A, Table S6, Fig. S2C). RM2 has 42 unique orthologous genes, which are mainly enriched in

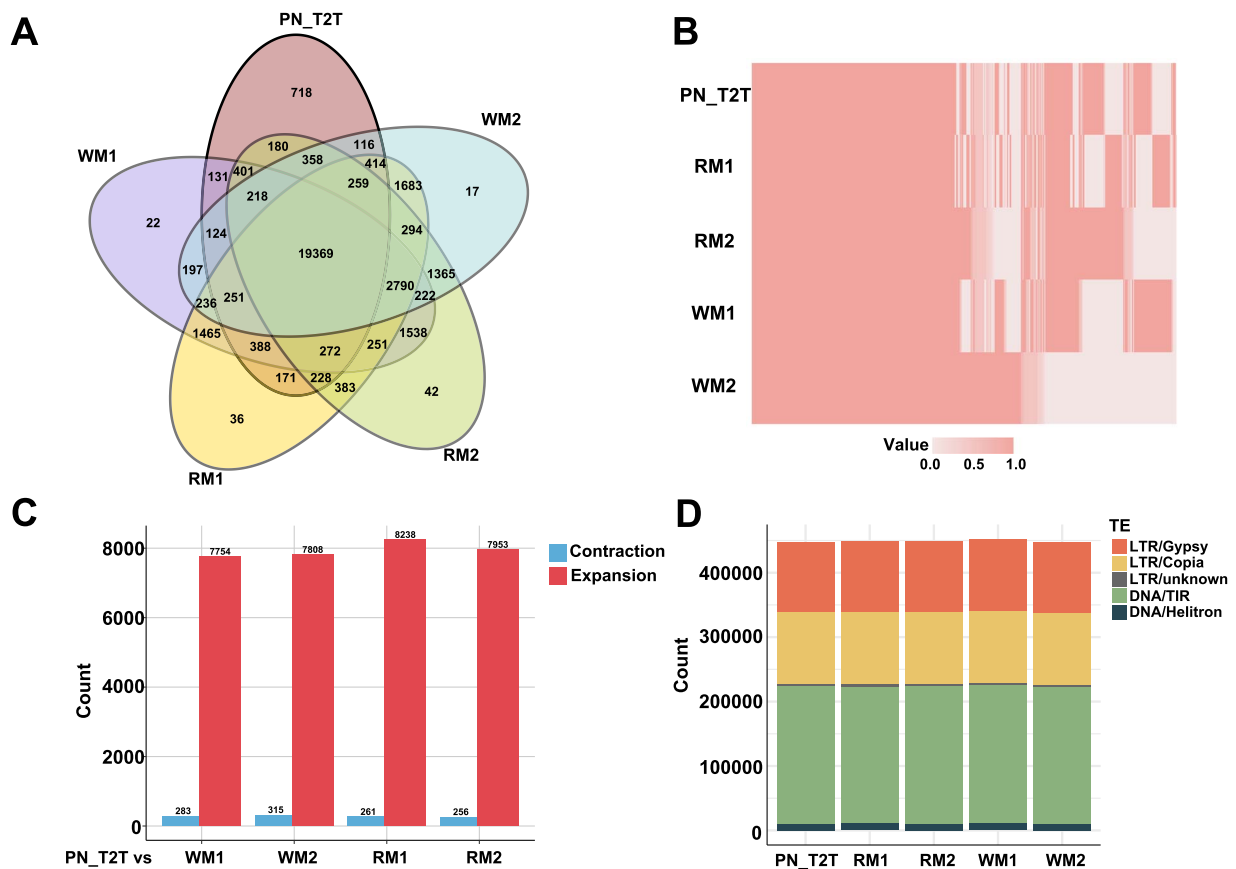


Figure 2 Comparisons among the PN_T2T, RM, and WM Genomes. (A) The number of shared and unique gene families among the PN_T2T, WM1, WM2, RM1, and RM2 genome assemblies. (B) Presence and absence of gene families in the PN_T2T, RM1, RM2, WM1, and WM2 genomes. The pink gradient indicates the abundance of each gene family when present in the genome, with the lightest color representing the absence of the gene family in that genome. (C) The numbers of contracted and expanded gene families in WM1, WM2, RM1, and RM2 compared to PN_T2T. (D) The number of major TE types across the whole genomes of WM1, WM2, RM1, RM2, and PN_T2T.

pathways including plant organ development, RNA splicing via transesterification reactions with bulged adenosine as a nucleophile, RNA splicing via transesterification reactions, and RNA splicing (Fig. 2A, Fig. S2D, Table S7). Compared with PN_T2T, the numbers of expanded gene families in WM1, WM2, RM1, and RM2 were 7754, 7808, 8238, and 7953, respectively, while the numbers of contracted gene family were 283, 315, 261, and 256, respectively (Fig. 2C). We also conducted statistics on transposable element (TE) annotations, with WM1, WM2, RM1, and RM2 showing proportions of 67.04%, 67.15%, 67.02%, and 66.56% of the whole genome, respectively. Additionally, the proportions of different TE types among the five haplotypes were similar, with TIR being the major TE (~23%) (Fig. 2D).

Population subdivision

Munage is a locally cultivated variety widely grown in Xinjiang. Predictions based on climate and sampling locations indicate that its most suitable growing regions are in northwestern China (Fig. 3A). There are no naturally occurring Eurasian grape species in China, and it is widely believed that the ancestors of Xinjiang's local grape varieties, including Munage, originated from West Asia, the domestication center of Eurasian cultivated grapes. To investigate the relationship of Munage with other Eurasian grapes, we analyzed resequencing data from 59 grape accessions.

These included 10 wild grapes (*V. vinifera* ssp. *sylvestris*) from Europe (EU), 10 wild grapes (*V. vinifera* ssp. *sylvestris*) from the Middle East and Caucasus region (ME), 20 cultivated grapes (*V. vinifera* ssp. *Vinifera*, also known as Eurasia grapes), and 9 Munage grapes from Xinjiang (Table S8), together with three *Vitis californica* samples from North America serving as the outgroup (OUT) (Table S8).

The phylogenetic tree analysis using FastTree software, with PN_T2T as the reference genome, revealed that wild and cultivated grapes form two distinct clades, reflecting an early divergence between the two groups (Fig. 3B). The 20 domesticated grape varieties, including those used for fresh eating (Table) and wine making (Wine), along with Munage, formed a well-supported monophyletic cluster. Among these, the cultivar most closely related to Munage was ISRAEL-U.I, a table grape originating from Israel (Fig. S3A) [4]. This finding suggests that Munage shares a common origin with domesticated Eurasian grapes. Principal component analysis (PCA) further confirmed the genetic differentiation among these populations. At the PC1 level, Munage and other Eurasian cultivated grapes are positioned on the left, in contrast to wild grapes on the right. However, at the PC2 level, Munage and other Eurasian cultivated grapes are separated (Fig. 3C), indicating that the Munage genome experienced different early selection pressures compared to other cultivated grapes during domestication.

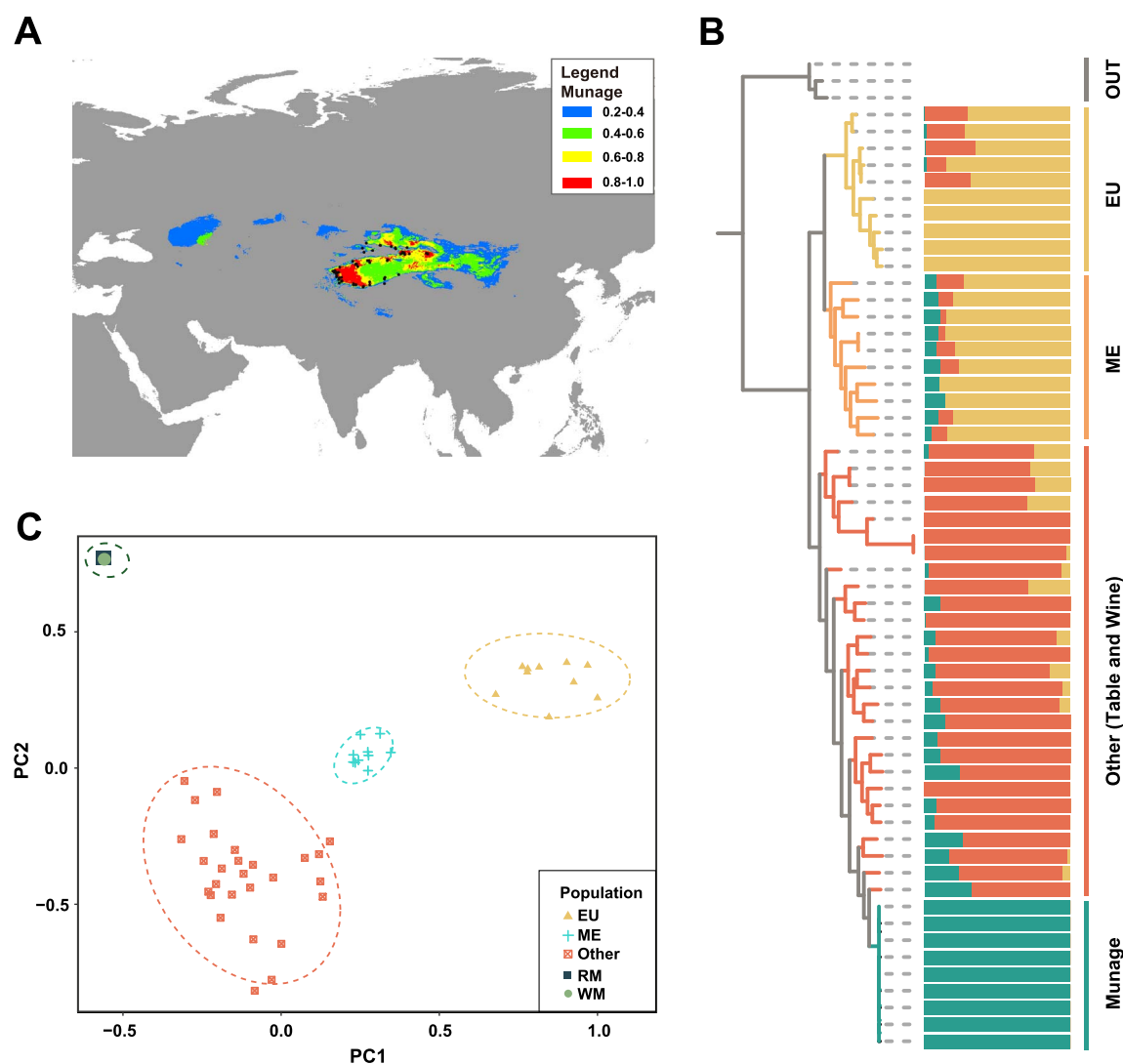


Figure 3 Population sub-division of Munage grapes. (A) Predicted distribution area of Munage, with black dots indicating the primary planting regions. The color from blue to red means the possibility from lower to higher. The color gradient from blue to red indicates the likelihood of suitability, ranging from lower to higher. (B) Phylogenetic tree with admixture analysis, where branches represent different group. The admixture plot shows $K = 4$. (C) A PCA supporting the designation of four groups. EU: European wild grapes; ME: Wild grapes from the Middle East and Caucasus region; Other: 20 Eurasia cultivated grapes; Munage: white Munage and red Munage grapes; OUT: *Vitis californica*.

Selective sweep in the Munage grape

To identify the selective sweep region in Munage, we performed composite likelihood ratio (CLR) analysis on nine Munage grapevine samples using SweeD software. Several selected genomic regions were detected, particularly on chromosomes 2 and 10 (Fig. 4A). On chromosome 2, the *RPM1* gene [33] was located at the selection region. *RPM1* is a classical plant resistance gene (R gene) that encodes a protein capable of recognizing pathogen effectors and activates the plant immune system to defend against pathogen invasion. Another identified gene is *BIG* [34], which participates in the regulation of plant hormone signaling, including auxin and gibberellin pathways. *BIG* modulates cellular sensitivity and response to these hormones, thereby influencing plant growth and development. On chromosome 10, a gene cluster comprising *petA*, *petB*, *petD*, *petG*, and *petL* [35] was identified; these genes encode proteins associated with the cytochrome b6f complex of the photosynthetic electron

transport chain, playing essential roles in photosynthesis. In addition, *CRK8*, *CRK1*, and *CRK25* [36] were detected; these cysteine-rich receptor-like kinases (CRKs) are involved in plant immune responses to various pathogens and environmental stresses. We further performed GO enrichment analysis on the top 1% of genomic regions with the highest composite likelihood ratios. These regions were predominantly enriched in pathways including camalexin metabolic process, regulation of shoot system development, anthocyanin-containing compound biosynthetic process, glutathione metabolic process, regulation of auxin polar transport, cell wall disassembly, endosperm development, cell wall modification involved in abscission, toxin catabolic process, and proteasome assembly (Fig. 4B, Table S9). These pathways are closely associated with the regulation of normal physiological activities in plants.

We also performed Population Branching Structure (PBS) analysis to identify differential divergent selective regions between Munage and other cultivated grapes, revealing significant

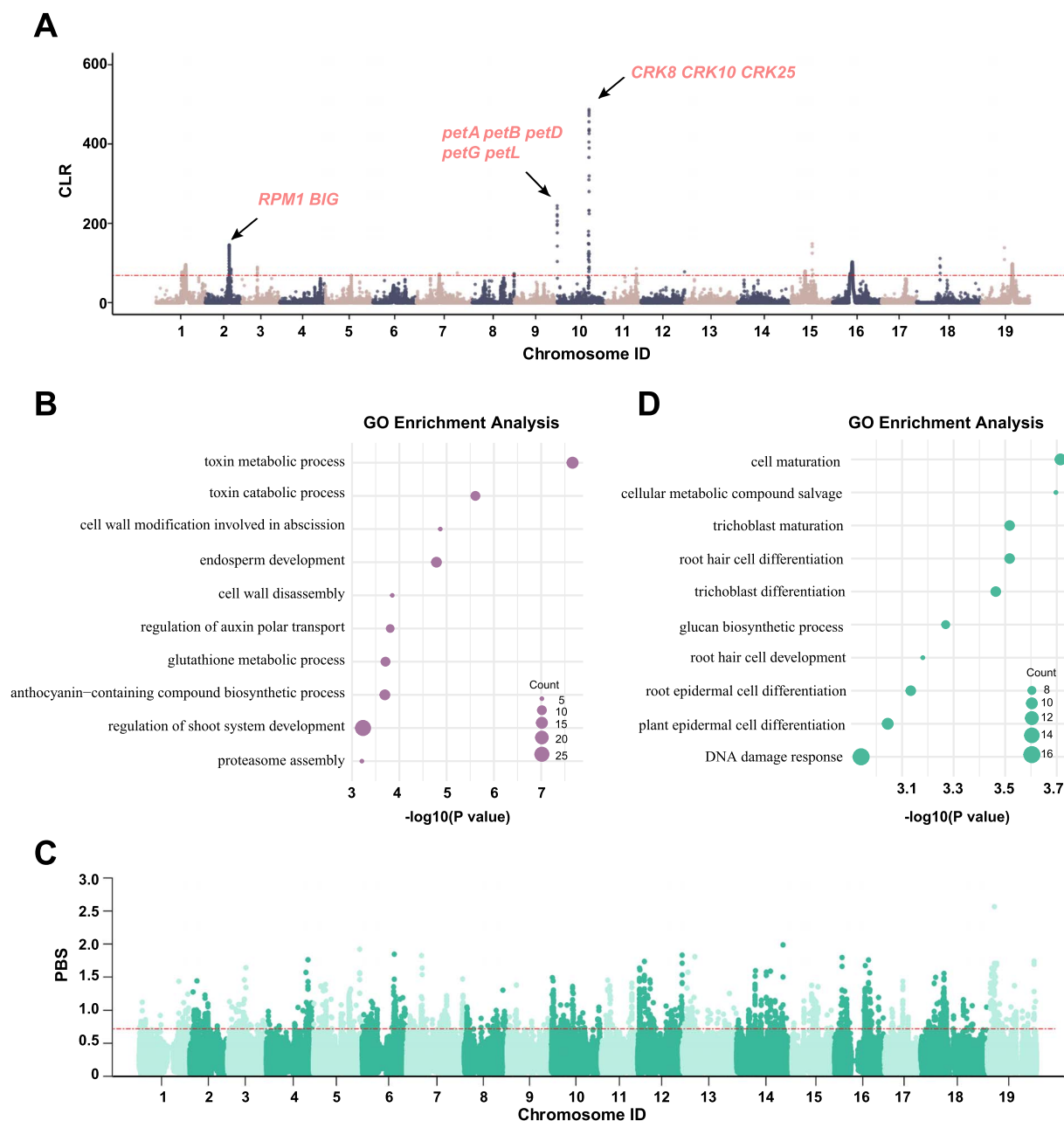


Figure 4 Selective sweep in the Munage grape. (A) SweepD analyses of the Munage group. (B) GO term enrichment of the biological process in the top 1% of genes in the SweepD analyses. (C) PBS analysis based on SNP variant information, using the Munage, Other, and ME groups. (D) GO term enrichment of the biological process in the top 1% of genes in the PBS analyses.

differentiation across numerous regions distributed across 19 chromosomes (Fig. 4C). We selected the top 1% of regions with the highest PBS values for gene ontology (GO) enrichment analysis. These regions were primarily enriched in pathways, such as cell maturation, plant epidermal cell differentiation, root epidermal cell differentiation, root hair cell development, glucan biosynthetic process, trichoblast differentiation, root hair cell differentiation, trichoblast maturation, cellular metabolic processes, and compound salvage (Fig. 4D, Table S10). These pathways associated with essential physiological processes and the maintenance of vital activities in plants, and are likely linked to adaptation to the specific environment in Xinjiang.

The variant and differential expression genes in grapes

Although RM and WM are genetically identical clones (IBD, 0.98; Fig. S3B), there are significant differences between them. WM features green shoots with a slight brownish tint and produces crispy, sweet fruit flesh, whereas RM shows green shoots with a reddish-purple hue, reddish branches, and crispy, slightly tart fruit flesh. Several distinct selection signals were observed between WM and RM (Figs S4 and S5, Table S11). Genes located within the top 1% of the selected regions in RM were enriched in pathways related to anthocyanin-containing compound biosynthesis, positive regulation of anthocyanin biosynthesis, positive regulation of

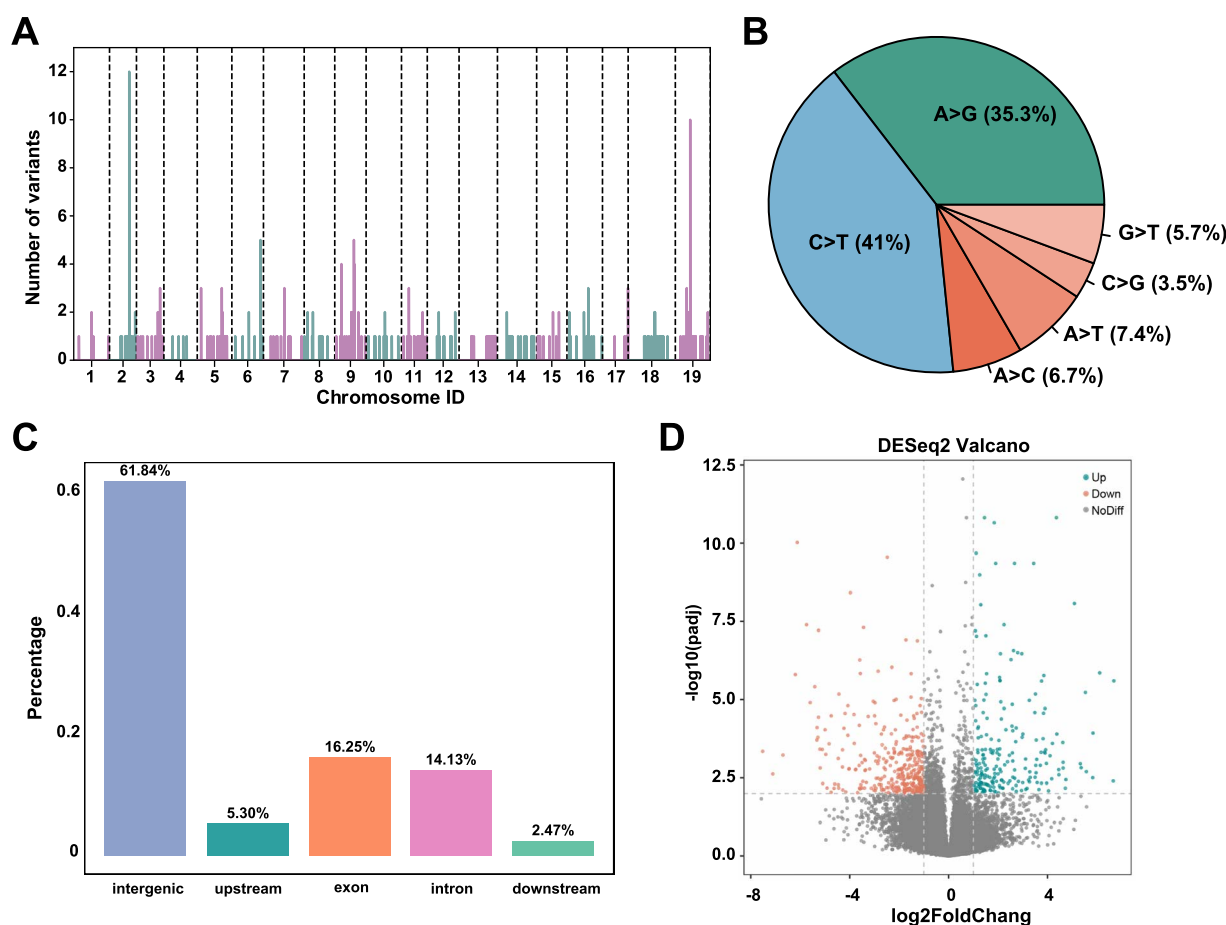


Figure 5 Variants and Differential Selection between RM and WM. (A) The number of somatic variants identified on each chromosome between WM and RM. (B) The proportion of somatic variant types present in WM and RM. The two transition types are shown in different shades of teal, while the four transversion types are shown in different shades of orange. (C) The percentage of somatic variants distributed at different positions within genes. (D) Differential gene expression analysis based on WM and RM transcriptome data.

anthocyanin metabolism, regulation of anthocyanin biosynthesis, and pigment biosynthesis. These processes show a strong association with berry color (Fig. S5B, Table S12).

To further explore the genetic differences between WM and RM, we examined sites that differed between the two cultivars but were consistent within each group. A total of 283 somatic mutation sites were identified. They are distributed throughout the entire genome, with slightly higher densities observed on chromosomes 2, 5, 9, and 19 (Fig. 5A). Among these, the two transition types, C > T (41%) and A > G (35.3%), were more prevalent than the four transversion types, each representing less than 10% of the variants (Fig. 5B). The majority of variants (61.8%) were located in intergenic regions, 30.4% were within genes (16.3% in exons and 14.1% in introns), and only 7.8% were situated in gene flanking regions (2 kb upstream and downstream) (Fig. 5C).

We also integrated RNA-seq data from RM and WM peels for temporal analysis of differentially expressed genes. A total of 578 differentially expressed genes were identified, including 216 up-regulated genes and 362 down-regulated genes in RM relative to WM (Fig. 5D). The up-regulated genes were primarily enriched in pathways related to water transport, fluid transport, transmembrane transport, and solute transport (Fig. S6A, Table S13), as well as related to skin color, including pigment metabolism and biosynthesis. Conversely, down-regulated

genes were enriched in pathways involved in heat response, temperature stimulus response, pigment metabolism, and pigment biosynthesis processes (Fig. S6B, Table S14). These findings suggest that somatic mutations between WM and RM may influence gene function or expression in Munage grapes.

Discussion

In this study, we provide valuable resources for a deeper understanding of the genetic basis, domestication history, and breeding of Munage grapes through the construction of high-quality genome assemblies and population genetic analyses. Our results indicate that Munage shares a common ancestor with Eurasian cultivated grapes originating in West Asia; however, through long-term artificial selection, it has undergone genomic adaptation to the climatic conditions of Xinjiang and genetically diverged from other Eurasian grape varieties. Although clonal propagation preserves the stability of favorable phenotypes and genotypes, the accumulation of somatic mutations has probably led to phenotypic variation between white Munage (WM) and red Munage (RM).

Genomes serve as indispensable resources for both basic research and breeding, as they provide comprehensive insights into genetic variations and thereby enable the development of improved traits and optimized crop varieties [37–41]. In 2023, the

first telomere-to-telomere grape genome assembly was published using PN40024 [32], marking a major advance in grapevine genomics. This complete assembly has substantially enhanced our understanding of grape genetics and has improved the precision of breeding efforts. However, most cultivated grapevines are highly heterozygous [42–46], which poses a challenge in studying their genomes. With the advent of chromium sequencing technology and high-throughput chromatin conformation capture (Hi-C) [47], an increasing number of high-quality plant genomes have been assembled, making diploid genome assembly and annotation a routine practice. Nonetheless, for ancient grape germplasm resources, phylogenetic studies of Eurasian species of native grapes are even more challenging to carry out due to the lack of high-quality reference genome sequences and genome annotation information. Thus, we generated haplotype-resolved high-quality genomes for this renowned cultivar; this resource will not only facilitate in-depth research on the cultivar itself but also provide critical genomic data support for studies on its closely related varieties, genetic improvement, and the evolutionary mechanisms of the entire grapes. The estimated genome size of *V. vinifera* is ~500 Mb⁴⁸ and our results indicate that the genome sizes of white Munage (WM) and red Munage (RM) range from 480 to 490 Mb, respectively in line with the recently published grape assemblies. Both genomes exhibited a heterozygosity level of 1.5%, reflecting the high degree of heterozygosity characteristic of grapevines, similar to cultivars such as Chasselas and Ugni Blanc [48]. A high BUSCO score (> 98.3) across all Munage haplotypes indicates that the genome assemblies are of high quality, with accurate annotations and minimal fragmentation or duplication. We annotated 33 942 (WM1), 34 034 (WM2), 35 292 (RM1), and 35 007 (RM2) genes, with TE annotations ranging between 67.15% and 66.56%. The annotated genes and TE annotations in previous grape assemblies, including Shanputao (*V. amurensis*), Noble (*M. rotundifolia*), PN40024 v2, and Sultanina v2, were lower than those in the current study, highlighting the importance of our research in terms of assembly quality [49].

According to historical records, the grapes domesticated in Xinjiang were introduced from Central Asia; however, there is a lack of genomic evidence, and the origin of Munage remains unclear. Based on population genetic analyses, we found that Munage clusters with Eurasian cultivated grapes and is closely related to U.I., a table grape originating from Israel [4]. According to Dong *et al.*, U.I belongs to group CG1 (Fig. S3A), which comprises table grape varieties from the West Asia region. Grapes within this group have a long domestication history, dating back ~11 000 years. Therefore, Munage likely shares common ancestry with the CG1 grapes that originated in West Asia, and its ancestors were subsequently introduced to Xinjiang via Central Asia. (Fig. 3B) [11, 50].

Although the ancestor of Munage originated in West Asia, its most suitable cultivation region has shifted to Northwest China, particularly Xinjiang (Fig. 3A). This indicates that adaptive selection occurred during the domestication process. The distinct selective sweeps observed are also supported by PCA and PBS analyses, highlighting the genomic differences between Munage and other cultivated grapes. (Figs 3C and 4C). The significant differences in selective regions can be observed across 19 chromosomes (Fig. 4C). During domestication, selection has been primarily driven by traits favored by humans [51], leading to the continuous accumulation of genetic variants that enhance these desirable traits and, over time, causing the genetic composition

of domesticated species to diverge from that of their wild relatives, thereby reflecting the selective pressures imposed during domestication [52–54]. In this study, the PBS analysis with multiple populations showed some selective sweep regions in the chromosomes. Significant differentiation across the chromosome region in the population provides strong evidence of natural selection over time in Munage.

Clonal propagation is widely used in grape production, providing significant advantages in maintaining the stability of favorable phenotypes and genotypes [11, 55–57]; however, the clonal process may also accumulate a large number of heterozygous somatic variations [58]. In this context, once beneficial mutations are identified during grape breeding, breeders can achieve rapid selection through clonal propagation [46, 59]. Recent genomic studies, such as those on ‘Yan73’ provided insights into the color development mechanism in grapevines [30]. White berries are assumed to have emerged from independent mutations [60]. It is assumed that RM originated by the bud mutation in WM, but distinct traits are present among them, especially in fruit peel color. In our findings, we provided scientific evidence to their clonal relationship through IBD analysis (0.98 between WM and RM). We hypothesize that certain somatic variants between WM and RM may underlie their phenotypic difference. To investigate this, we identified the somatic mutations between WM and RM by identifying specific variants in each group, and a total of 283 somatic mutation sites were identified between WM and RM. These variations were distributed across the genome, with several peaks observed, including a notable cluster on chromosome 2, a region previously reported to be associated with berry color (Fig. 5A) [43]. Only 30% of these variations are located on genes, indicating that these variations probably mainly affect gene expression rather than altering the genes themselves (such as inactivation or functional changes). As expected, a total of 578 differentially expressed genes were identified based on RNA-seq data from RM and WM peels. (Fig. 5D). Both the up-regulated genes and down-regulated genes were enriched in pathways related to skin color, including pigment metabolism and biosynthesis (Fig. S6, Tables S13 and S14). However, our recent work relies on bioinformatics approaches and transcriptomic data, necessitating follow-up experiments to validate these findings and identify the causal genes for color change in the future. These experiments include quantitative real-time polymerase chain reaction (qRT-PCR) to verify the expression levels of the candidate genes and performing functional studies, such as gene knockouts or overexpression, to confirm the roles of these candidate genes in the observed phenotype.

In summary, our findings not only provide high-quality genomic resources for the Munage grapes but also reveal the roles of clonal propagation, selective pressures, and somatic mutations in shaping fruit phenotypic variation. These results further offer theoretical guidance and data support for future genetic improvement of grapes and functional gene studies.

Materials and methods

Plant materials and sampling

An ancient grape variety with specific traits, a standard mother plant of red Munage over 30 years old, and a standard mother plant of white Munage over 80 years old were selected from multiple

locations in Xinjiang, China. High-quality DNA from leaves to generate haplotype-resolved genome assemblies. To achieve this, we employed PacBio whole-genome high-fidelity (HiFi) sequencing at 100× coverage, high-throughput chromosome conformation capture (Hi-C) proximity ligation, and Illumina short-read sequencing. Genome annotation was performed using Illumina RNA-seq data from various Munage tissues. Details of the genetic resources used are provided in Table S8.

DNA sequencing and library preparation

The CTAB method was used to extract high-quality genomic DNA from leaf tissue samples. The concentration and purity of the DNA were assessed using a Qubit fluorometer from Thermo Fisher Scientific. Gel electrophoresis was performed to verify DNA integrity. To sequence the DNA of the standard mother plant of the Xinjiang native ancient grape variety Munage, third generation of PacBio Sequel CCS 100× with HiFi data volume of up to 50Gb and two cells were used. For PacBio HiFi sequencing, a standard SMRTbell library was constructed using the SMRTbell Express Template Prep Kit 2.0, following the manufacturer's recommendations (Pacific Biosciences, CA, USA). Additionally, the Munage genome haplotypes were assembled using second-generation 50× Hi-C combination sequencing technology.

Total RNA extraction, library preparation and sequencing

To perform RNA sequencing, we utilized the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina® (New England Biolabs, MA, USA) to extract total RNA from various plant parts, including the leaves, pulp, stalk, and root of the Munage grapevine. Subsequently, paired end reads of 150 base pairs were generated on the Illumina NovaSeq platform. The sample details are presented in Table S8.

De novo haplotype-resolved genome assembly and quality assessment

We first used the Hi-C Integrated Assembly mode of HiFiasm, combining Munage HiFi and Hi-C data, to generate two contig-level haplotype genomes. Genomic heterozygosity was estimated using GenomeScope (v2.0) [61], applying a k-mer-based strategy applied to the raw HiFi reads. Next, RagTag was used with the Cabernet Sauvignon genome as a reference to infer the approximate order of contigs along the chromosomes. Hi-C sequencing data were then utilized to anchor all contigs using Juicer (v1.5) [62]. The genome structure was further refined with 3D-DNA, and manual adjustments were performed in Juicebox (v1.11.08, <https://github.com/aidenlab/Juicebox>), followed by a secondary application 3D-DNA [63] to obtain scaffold-level genome. Genome quality was assessed using BUSCO [64] with the embryophyta_odb10 database to evaluate the genome's completeness.

Annotation of genes and transposable elements

We employed a comprehensive genome-wide annotation pipeline (<https://github.com/unavailable-2374/Genome-Wide-Annotation-Pipeline>) to annotate the Munage genome, utilizing

RNA-seq data from multiple tissues. Initially, RNA sequences were aligned to repeat-masked assemblies with Hisat2 (v2.10.2), followed by transcript assembly with StringTie (v1.3.0) [65]. Transcripts and UniProt (<https://www.uniprot.org/>) were used as supporting evidence for gene identification. Preliminary gene models were constructed using tools such as exonerate, genewise, and transfrag. Following the initial gene models construction, AUGUSTUS (v3.4.0) [66] was applied to refine predictions. Genes located in duplicated regions, with CDS regions shorter than 90 bases, with coding sequences shorter than 90 bp, or lacking supporting evidence were filtered out. Missing genes were supplemented where possible, and alternative splicing analysis was performed on the remaining intact genes. Final gene models were verified using a hidden Markov model derived from the Pfam database.

For transposable element (TE) annotation, a TE library was first constructed using RepeatModeler (open-2.0.3) with the -LTRStruct option [67]. Genome-wide TE annotation was then conducted using RepeatMasker (open-4.1.2, <https://github.com/rmhubble/RepeatMasker>) with the -e rmbblast, -lcambig, and slow models.

Genome comparison between haplotypes and reference genome

For comparative genomic analysis, the WM1, WM2, RM1, RM2, and PN_T2T genomes were first aligned using Minimap2 (v2.21) [68], and the resulting BAM files were indexed with SAMtools (v1.4) [69]. Synteny relationships and structural rearrangements among the genomes were subsequently identified using SyRI (v1.7.0) [70]. The alignment results were finally visualized with plotsr [71].

SNP calling and filtering

We utilized resequencing data from a total of 59 grape accessions, comprising 10 wild grapes (*Vitis vinifera* subsp. *sylvestris*) from Europe (EU), 10 wild grapes from the Middle East (ME), 27 cultivated grapes (Other), three white Munage, and six red Munage. Additionally, three Muscadine grape samples were included as an outgroup. Of these, nine Munage accessions were newly resequenced in this study, while the remaining datasets were retrieved from the National Center for Biotechnology Information (NCBI). Quality control of the resequencing data was performed using fastp (v0.21) [72] with default parameters. Subsequently, the quality-filtered paired-end data were aligned to the PN_T2T complete reference genome, and SNPs were identified across the 59 resequenced samples using GTX (v2.2.1). To reduce false-positive variant calls, we applied stringent filtering with VCFtools [73]. Specifically, we removed genotypes with a genotype quality less than 20 (option -minGQ 20), SNPs with more than two alleles (options -min-alleles 2--max-alleles 2), and SNPs with more than 20% missing genotypes (option -max-missing 0.8). Indels were also filtered by VCFtools using command --remove-indels.

Gene ontology enrichment analysis

We first converted the genome into protein sequences based on the gene annotation file, and then annotated the protein sequences using the eggNOG-mapper web server (eggNOG

v5.0) [74]. Each protein sequence under analysis was ultimately assigned a functional description corresponding to its ortholog. The final annotation results were used as the background, and GO enrichment analysis of the target genes was performed using TBtools [75].

Phylogeny and population structure

Principal component analysis (PCA) was performed using PLINK [76]. Phylogenetic trees were constructed using FastTree (v2.1.10) [77] based on GTR models. Population structure was inferred using Admixture (v1.3.0) [78] with K values ranging from 2 to 10, based on nuclear SNPs. IBD was calculated using PLINK with parameter: —genome.

Selective sweep

To detect signatures of selective sweeps in Munage populations, we applied the Composite Likelihood Ratio (CLR) method implemented in SweepD (v3.2.1) [79]. The top 1% of likelihood values were considered to identify potential selective sweeps within highly differentiated genomic regions. The analysis first involved partitioning the population and chromosomes. Chromosome lengths were obtained from the VCF file header, and scanning window sizes were set accordingly. Next, we determined the length of each chromosome and set the size of the scanning window. To ensure uniform segmentation, we divided each chromosome into 10 kb grids, 100 grids per megabase. SweepD was then executed, and the results were visualized following generation of the output files. We performed Population Branch Statistic (PBS) analysis using PBSscan (<https://github.com/thamala/PBSscan>) [80] to detect signatures of selective sweeps among the Munage, other cultivated grapes, and ME wild grapes. The PBS output was ranked in descending order, and the top 1% of values were designated as outliers for subsequent Gene Ontology (GO) enrichment analysis.

Transcriptome differential gene analysis

In this study, we analyzed RNA-seq data from three WM leaves (control), three WM peels, and nine RM peels, using WM2 as the reference genome for differential expression analysis. We first constructed an index of the haplotype reference genome with HISAT2's 'hisat2-build' (v.2.2.1) [65]. The transcriptome data were then aligned to the reference genome using HISAT2. The resulting SAM files were converted to BAM format, sorted, and indexed with SAMtools. In parallel, the gene annotation file in GFF3 format was converted to GTF format using 'gffread' (v.0.12.7) [81]. Next, we employed 'featureCounts' [82] to quantify gene expression by comparing the genomic positions of annotated features with the alignment positions of sequencing reads or fragments, thereby generating a gene expression matrix.

Subsequently, we applied DESeq2 [83] to estimate differential fold changes based on the gene expression matrix, obtaining the differential fold changes and significance *p*-values for each gene. Genes were filtered using the criteria $|\log_2FC| \geq 2$ and $\text{padj} < 0.01$, with upregulated and downregulated genes classified as 'up' and 'down', respectively.

Acknowledgements

The financial assistance for this research is received from the Project of Fund for Stable Support to Agricultural Sci-Tech Renovation (xjnkwdzc-2025001-07, xjnkwdzc-2026001-18, xjnkwdzc-2025001-39), Tianshan Autonomous Region Talent Youth Top notch Talent—Young Science and Technology Innovation Talent Project (2024TSYCCX0097), the National Natural Science Foundation of China (No. 32160682, No. 32260732 and No. 32460722), Xinjiang Uygur Autonomous Region Tianchi Talent - Young Doctor approved for VIVEK YADAV (History and genetic mechanism of grape domestication in Asia), Xinjiang Uygur Autonomous Region Tianchi Talent—Special Expert Project (Whole genome design and breeding of grapes), Key research and development project of autonomous region (2022B02045-1-1, 2023B02029-1-1), National Key Laboratory of Tropical Crop Breeding (No. NKLTCC-RC202501), the China Agriculture Research System of MOF and MARA and Shenzhen Polytechnic University Research Project (6024330001K) to Tian Ling.

Author contributions

Hua Xiao, Xinyu Wu, Ling Tian, Yongfeng Zhou, Haixia Zhong, Xiaoya Shi, Fuchun Zhang (Conceptualization), Haixia Zhong, Xiaoya Shi, Fuchun Zhang, Jin Yao, Xu Wang, Vivek Yadav, Xiaoming Zhou (Data curation), Xiaoya Shi, Fuchun Zhang, Xu Wang, Shuo Cao, Songlin Zhang, Chuan Zhang, Jiangxia Qiao, Zhongjie Liu, Yingchun Zhang, Yuting Liu, Hao Wang, Hui Xue, Mengyan Zhang (Data analysis), Xinyu Wu, Haixia Zhong, Ling Tian, Yongfeng Zhou (Funding acquisition), Haixia Zhong, Xiaoya Shi, Fuchun Zhang, Tianhao Zhang, Wenrui Wang, Ruoyan Zhao, Tianrong Fan, Zhongqi Liu, Jiacy Li, Ruirui Liu (Methodology), Haixia Zhong (Project administration), Haixia Zhong, Fuchun Zhang, Xiaoming Zhou, Xinyu Wu (Resources), Haixia Zhong, Xiaoya Shi, Fuchun Zhang (Validation), Haixia Zhong, Xiaoya Shi, Vivek Yadav, Hua Xiao (Writing—review & editing), and Hua Xiao, Xinyu Wu, Yongfeng Zhou (Supervision)

Data availability

All raw data have been deposited in the NCBI Sequence Read Archive under project number PRJNA1158437 and the National Genomics Data Center (NGDC) Genome Sequence Archive (GSA) (<https://ngdc.cncb.ac.cn/gsa/>), with BioProject number PRJCA029988. The assembly and annotation have been deposited in zenodo: <https://doi.org/10.5281/zenodo.13732355>.

Conflicts of interest statement

The authors declare no competing interests.

Supplementary material

Supplementary material is available at *Horticulture Research* online.

References

1. Banilas G, Korkas E, Kaldis P. *et al.* Olive and grapevine biodiversity in greece and cyprus—a review. In: Lichtfouse E,

- ed. *Climate Change, Intercropping, Pest Control and Beneficial Microorganisms*. Springer: Dordrecht, 2009;401–28
2. This P, Lacombe T, Thomas MR. Historical origins and genetic diversity of wine grapes. *Trends Genet.* 2006;22:511–9
3. Jun W, Changqing D. Advance in research on domestication and taxonomy of Eurasian grape (*Vitis vinifera* L.). *Front Plant Sci.* 2010;11:572
4. Dong Y, Duan S, Xia Q. *et al.* Dual domestications and origin of traits in grapevine evolution. *Science.* 2023;379:892–901
5. Manen JF, Bouby L, Dalnoki O. *et al.* Microsatellites from archaeological *Vitis vinifera* seeds allow a tentative assignment of the geographical origin of ancient cultivars. *J Archaeol Sci.* 2003;30:721–9
6. Arroyo-García R, Ruiz-García L, Bolling L. *et al.* Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *sativa*) based on chloroplast DNA polymorphisms. *Mol Ecol.* 2006;15:3707–14
7. Grassi F, Mattia FD, Zecca G. *et al.* Historical isolation and quaternary range expansion of divergent lineages in wild grapevine. *Biol J Linn Soc.* 2008;95:611–9
8. Maghradze D, Melyan G, Salimov V. *et al.* Wild grapevine (*Vitis sylvestris* C.C.Gmel.) wines from the Southern Caucasus region. *OENO One.* 2020;54:809–22
9. Liu X. Viticulture and Viniculture in the Turfan Region.
10. Laufer B. SINO-IRANICA: chinese contributions to the history of civilization in ancient Iran. *Publ Field Mus Nat Hist Anthropol Ser.* 1919;15:i–630
11. Zhao M, Jiang H, Grassa CJ. Archaeobotanical studies of the Yanghai cemetery in Turpan, Xinjiang, China. *Archaeol Anthropol Sci.* 2018;11:1143–53
12. Jiang H, Zhang YB, Li X. *et al.* Evidence for early viticulture in China: proof of a grapevine (*Vitis vinifera* L., *Vitaceae*) in the Yanghai Tombs, Xinjiang. *J Archaeol Sci.* 2009;36:1458–65
13. Uchibayashi M. Introduction of grapes and alfalfa into China: a reflection on the role of Zhang Qian. *Yakugaku Zasshi.* 2005;125:895–8
14. Li S. Grapevine breeding and genetics in china: history, current status and the future. *Acta Hort.* 2025;1082:165–76
15. Zhang F, Zhong H, Zhou X. *et al.* Grafting with rootstocks promotes phenolic compound accumulation in grape berry skin during development based on integrative multi-omics analysis. *Hortic Res.* 2022;9:uhac055
16. Liu S, Zhong H, Zhang F. *et al.* Genetic diversity and core germplasm research of 144 munake grape resources using 22 pairs of SSR markers. *Horticulturae.* 2023;9:917
17. Mou Z, Yuan Y, Wei W. *et al.* Integrative metabolomic and transcriptomic analysis provides novel insights into the effects of SO₂ on the postharvest quality of ‘munage’ table grapes. *Foods.* 2024;13:3494
18. Zhang Z, Xing S, Yuan Y. *et al.* Sulfur dioxide fumigation promotes GABA accumulation and energy metabolism to delay quality deterioration in postharvest table grapes. *Postharvest Biol Technol.* 2024;215:113022
19. Yin X, Liu B, Maimaitiyming R. *et al.* The effect of non-*Saccharomyces cerevisiae* *Torulaspora delbrueckii* on the aroma composition of munage grape base-wine and the mechanism of the effect. *Fermentation.* 2024;10:266
20. Wang Y, Lv Y, Han T. *et al.* Post-harvest quality changes and molecular responses of epidermal wax in ‘munage’ grapes with *Botrytis cinerea* Infection. *Int J Mol Sci.* 2025;26:3468
21. Huang J, Zhang G, Li Y. *et al.* Integrative genomic and transcriptomic analyses of a bud sport mutant ‘Jinzaowu’ with the phenotype of large berries in grapevines. *PeerJ.* 2023;11:e14617
22. Wang X, Liu Z, Zhang F. *et al.* Integrative genomics reveals the polygenic basis of seedlessness in grapevine. *Curr Biol.* 2023;34:3763–3777.e5
23. Guo D-L, Zhang G-H. A new early-ripening grape cultivar - ‘fengzao’. *Acta Hort.* 2015;1082:153–56
24. Wei L, Cao Y, Cheng J. *et al.* Comparative transcriptome analyses of a table grape ‘Summer Black’ and its early-ripening mutant ‘Tiangong Moyu’ identify candidate genes potentially involved in berry development and ripening. *J Plant Interact.* 2020;15:213–22
25. Leng F, Ye Y, Zhu X. *et al.* Comparative transcriptomic analysis between ‘Summer Black’ and its bud sport ‘Nantaihuizao’ during developmental stages. *Planta.* 2021;253:23
26. Azuma A, Kobayashi S, Goto-Yamamoto N. *et al.* Color recovery in berries of grape (*Vitis vinifera* L.) ‘Benitaka’, a bud sport of ‘Italia’, is caused by a novel allele at the VvmybA1 locus. *Plant Sci.* 2009;176:470–8
27. Pan Q, Chen F, Zhu BQ. *et al.* Molecular cloning and expression of gene encoding aromatic amino acid decarboxylase in ‘Vidal blanc’ grape berries. *Mol Biol Rep.* 2012;39:4319–25
28. Negri AS, Prinsi B, Failla O. *et al.* Proteomic and metabolic traits of grape exocarp to explain different anthocyanin concentrations of the cultivars. *Front Plant Sci.* 2015;6:603
29. Calderón L, Carbonell-Bejerano P, Muñoz C. *et al.* Diploid genome assembly of the Malbec grapevine cultivar enables haplotype-aware analysis of transcriptomic differences underlying clonal phenotypic variation. *Hortic Res.* 2024;11:uhae080
30. Zhang K, du M, Zhang H. *et al.* The haplotype-resolved T2T genome of teinturier cultivar Yan73 reveals the genetic basis of anthocyanin biosynthesis in grapes. *Hortic Res.* 2023;10:uhad205
31. Shirasawa K, Hirakawa H, Azuma A. *et al.* De novo whole-genome assembly in an interspecific hybrid table grape, ‘Shine Muscat’. *DNA Res.* 2022;29:dsac040
32. Shi X, Cao S, Wang X. *et al.* The complete reference genome for grapevine (*Vitis vinifera* L.) genetics and breeding. *Hortic Res.* 2023;10:uhad061
33. Boyes DC, Nam J, Dangl JL. The *Arabidopsis thaliana* RPM1 disease resistance gene product is a peripheral plasma membrane protein that is degraded coincident with the hypersensitive response. *Proc Natl Acad Sci USA.* 1998;95:15849–54
34. Gil P, Dewey E, Friml J. *et al.* BIG: a calossin-like protein required for polar auxin transport in Arabidopsis. *Genes Dev.* 2001;15:1985–97
35. Cramer WA, Hasan SS. *Structure-Function of the Cytochrome b₆f Lipoprotein Complex*. Dordrecht: Springer Netherlands; 2016;84:1349–58
36. Yadeta KA, Elmore JM, Creer AY. *et al.* A cysteine-rich protein kinase associates with a membrane immune complex and the cysteine residues are required for cell death. *Plant Physiol.* 2016;173:771–87
37. Gao C. Genome engineering for crop improvement and future agriculture. *Cell.* 2021;184:1621–35

38. Yu, Ying-hui L, Qing-wen Y. *et al.* Genomics-based crop germplasm research: advances and perspectives. 2015;48:3333–53
39. Liu Z, Wang N, Su Y. *et al.* Grapevine pangenome facilitates trait genetics and genomic breeding. *Nat Genet.* 2024;56:2804–14
40. Zhang Z, Liu W, Zhang T. *et al.* Population genomics of *Vitis pseudoreticulata* reveals the genetic basis of fungal resistance in grapevine. *Hortic Plant J.* 2025;
41. Liu Y, Dong Y, Fang X. *et al.* Pangenomics and single-cell transcriptomics uncover the genetic basis of continuous bearing trait in grapevine. *Hortic Res.* 2025;uhaf228
42. Zhou Y, Minio A, Massonnet M. *et al.* The population genetics of structural variants in grapevine domestication. *Nat Plants.* 2019;5:965–79
43. Zhou Y, Massonnet M, Sanjak JS. *et al.* Cantu, D. & Gaut, B.S. Evolutionary genomics of grape (*Vitis vinifera* ssp. *vinifera*) domestication. *Proc Natl Acad Sci USA.* 2017;114:11715–20
44. Zhang T, Peng W, Xiao H. *et al.* Population genomics highlights structural variations in local adaptation to saline coastal environments in woolly grape. *J Integr Plant Biol.* 2024;66:1408–26
45. Gaut BS, Seymour DK, Liu Q. *et al.* Demography and its effects on genomic variation in crop domestication. *Nat Plants.* 2018;4:512–20
46. Xiao H, Liu Z, Wang N. *et al.* Adaptive and maladaptive introgression in grapevine domestication. *Proc Natl Acad Sci USA.* 2023;120:e2222041120
47. A Decade of 3c Technologies: Insights into Nuclear Organization References.
48. Djari A, Madignier G, di Valentin O. *et al.* Haplotype-resolved genome assembly and implementation of VitExpress, an open interactive transcriptomic platform for grapevine. *Proc Natl Acad Sci USA.* 2024;121:e2403750121
49. Wang Y, Ding K, Li H. *et al.* Biography of *Vitis* genomics: recent advances and prospective. *Hortic Res.* 2024;11:uhaf128
50. Li Y. Agriculture and palaeoeconomy in prehistoric Xinjiang, China (3000–200 bc). *Veg Hist Archaeobotany.* 2020;30:287–303
51. Lamoureux D, Bernole A, le Clainche I. *et al.* Anchoring of a large set of markers onto a BAC library for the development of a draft physical map of the grapevine genome. *Theor Appl Genet.* 2006;113:344–56
52. Kui L, Tang M, Duan S. *et al.* Identification of selective sweeps in the domesticated table and wine grape (*Vitis vinifera* L.). *Front Plant Sci.* 2020;11:572
53. Zhang M, Xu X, Zhang T. *et al.* The dynamics of wild *Vitis* species in response to climate change facilitate the breeding of grapevine and its rootstocks with climate resilience. *Hortic Res.* 2025;12:uhaf104
54. Hou T, Xu Y, Dong Y. *et al.* The cytonuclear interactions during grapevine domestication. *J Integr Plant Biol.* 2025;67:13968
55. Miller AJ, Gross BL. From forest to field: perennial fruit crop domestication. *Am J Bot.* 2011;98:1389–414
56. Forneck A, Benjak A, Rühl EH. Grapevine (*Vitis* ssp.): example of clonal reproduction in agricultural important plants. In: Schon I, Martens K, Dijk P, eds. *Lost Sex*. Heidelberg, Germany: Springer, 2009
57. Peng Y, Wang Y, Liu Y. *et al.* The genomic and epigenomic landscapes of hemizygous genes across crops with contrasting reproductive systems. *Proc Natl Acad Sci USA.* 2025;122:e2422487122
58. Xiao H, Wang Y, Liu W. *et al.* Impacts of reproductive systems on grapevine genome and breeding. *Nat Commun.* 2025;16:2031
59. Jaillon O, Aury J-M, Noel B. *et al.* The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature.* 2007;449:463–7
60. Kobayashi S, Goto-Yamamoto N, Hirochika H. Retrotransposon-induced mutations in grape skin color. *Science.* 2004;304:982–2
61. Ranallo-Benavidez TR, Jaron KS, Jaron KS. *et al.* GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun.* 2020;11:1432
62. Durand NC, Shamim MS, Machol I. *et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* 2016;3:95–8
63. Dudchenko O, Batra SS, Omer AD. *et al.* De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science.* 2016;356:92–5
64. Simão FA, Waterhouse RM, Ioannidis P. *et al.* BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31:3210–2
65. Kim D, Paggi JM, Park C. *et al.* Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019;37:907–15
66. Stanke M, Keller O, Gunduz I. *et al.* AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* 2006;34:W435–9
67. Flynn JM, Hubley R, Goubert C. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci.* 2020;117:9451–7
68. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2017;34:3094–100
69. Danecek P, Bonfield JK, Liddle J. *et al.* Twelve years of SAM-tools and BCFtools. *GigaScience.* 2020;10:giab008
70. Goel M, Sun H, Jiao W-B. *et al.* SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* 2019;20:277
71. Goel M, Schneeberger K. plotsr: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics.* 2022;38:2922–6
72. Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta.* 2023;2:e107
73. Danecek P, Auton A, Abecasis G. *et al.* The variant call format and VCFtools. *Bioinformatics.* 2011;27:2156–8
74. Huerta-Cepas J, Szklarczyk D, Heller D. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 2018;47:D309–14
75. Chen C, Chen H, Zhang Y. *et al.* TBtools—an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant.* 2020;13:1194–202
76. Purcell SM, Neale B, Todd-Brown K. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75
77. Piñeiro C, Abuín JM, Pichel JC. Very Fast Tree: speeding up the estimation of phylogenies for large alignments through

- parallelization and vectorization strategies. *Bioinformatics*. 2020;36:4658–9
78. Alexander DH, Novembre J, Lange KL. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19:1655–64
 79. Pavlidis P, Zivkovic D, Stamatakis A. *et al.* SweepD: likelihood-based detection of selective sweeps in thousands of genomes. *Mol Biol Evol*. 2013;30:2224–34
 80. Hämälä T, Savolainen O. Genomic patterns of local adaptation under gene flow in *Arabidopsis lyrata*. *Mol Biol Evol*. 2019;36:2557–71
 81. Pertea G, Pertea M. GFF Utilities: GffRead and GffCompare. *F1000 Res*. 2020;9:ISCB Comm J-304
 82. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2013;30:923–30
 83. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550