De novo assembling a high-quality genome sequence of Amur grape (*Vitis amurensis* Rupr.) gives insight into *Vitis* divergence and sex determination

Pengfei Wang[1#], Fanbo Meng[2#], Yiming Yang[3], Qian Mu[1], Tingting Ding[1], Huiping Liu[1], Fengxia Wang[1], Ao Li[1], Qingtian Zhang[1], Shutian Fan[2], Bo Li[1], Zhiyao Ma[4], Tianhao Zhang[4], Yongfeng Zhou[4], Hongjun Zhao[1]*, Xiyin Wang[5]*

1.Shandong Academy of Grape, Jinan 250100, China

2.State Key Laboratory of Southwestern Chinese Medicine Resources, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China

3.Institute of special animal and plant sciences of CAAS, Changchun 130000, China

4.Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, 518000, China

5.North China University of Science and Technology, Tangshan 063000, China

Contribute equally to this work

*Corresponding author, Hongjun Zhao, 13964190239@139.com; Xiyin Wang, wangxiyin@vip.sina.com

Short title: Amur grape genome sequence

Author Contributions

P.W., Y.Z., Y.Y., H.Z. and B.L. designed the experiments. Q.M., T.D., H.L., F.W., S.F. and Q.Z. performed the experiments and wrote the manuscript. F.M., A.L., Z. M. and T. Z. analyzed the data. J.J., Y.Z. and X.W. edited the manuscript.

One Sentence Summary

We presented a high-quality *Vitis amurensis* genome sequence and conducted an in-depth study of grape genome variation, sex determination, and cold resistance, contributing to understanding of grape biology and breeding fine grapes.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/pphys/pages/General-Instructions) is Xiyin Wang.

Abstract

To date, there is no high-quality sequence for genomes of the East Asian grape species, hindering biological and breeding research efforts to improve grape cultivars. This study presents a ~522 Mb of the *Vitis amurensis* (*Va*) genome sequence containing 27,635 coding genes. Phylogenetic analysis indicated that *V. riparia* (Vr) may firstly split from the other two species, *Va*, *V. Vinifera* (*Vv*; Pinot Noir: PN40024 and Cabernet Sauvignon). Much divergent gene reservation among three grape duplicated gene sets suggests that the core eudicot common hexaploidy (ECH), 130 million years ago (Mya), has still played a non-negligible role in grape species divergence and biological innovation. Prominent accumulation of sequence variants might have improved cold resistance in *Va*, resulting in a more robust cold resistance gene regulatory

network than those in *Vv* and *Vr*. In contrast, *Va* preserved much fewer NBS disease resistance genes than the other grapes. Notably, multi-omics analysis identified one trans-cinnamate 4-monooxygenase gene positively correlated to the resveratrol accumulated during *Va* berry development. A selective sweep analysis revealed a hypothetical *Va* sex-determination region (SDR). Besides, a PPR-containing protein-coding gene in the hypothetical SDR may be related with sex determination in *Va*. The content and arrangement order of genes in the putative SDR of female *Va* were similar to the SDR of female *Vv*. However, the putative SDR of female *Va* lost one Flavin-containing monooxygenases (FMO) and contained one extra uncharacterized protein-coding gene. These findings will improve the understanding of *Vitis* biology and contribute to the improvement of grape breeding.

Keywords: *Vitis amurensis*, genome, structural variation, nutrient accumulation, sex determination, selective sweep

Introduction

Grape genus (*Vitis*) contain about 60 species of vining plants. Eurasian grape, East Asian grape, American grape, and other intergeneric hybrid grapes are some of the most common grape varieties globally (Foria *et al.*, 2022).

As a wild plant, originated in eastern Asia, Amur grape (*Vitis amurensis*), also called "Chinese wild grape" or "Shanputao" (Wang *et al.*, 2019; Wang *et al.*, 2021), is one of the East Asian grape species, resembling the wine grape (one of the Eurasian grapes). Wild Amur grapes are mainly found in the Yellow River basin and Songhua River basin in China, Siberia in Russia, Korea, Japan, and many other areas around the world (Chen *et al.*, 2018). Amur grapes have female, male, or bisexual plants (Figure 1b-d). The Latin name "*Vitis amurensis* Rupr." of the Amur grape was first proposed in 1857 by botanist Franz Josef

Ruprecht, and the cultivation of *Vitis amurensis* can be traced back to 1907 (wine-world.com). A female Amur grape cultivated variant (cv), called "Zuoshan No. 1" is vastly planted in the Northern China. A bisexual Amur grape cv. "Shuangyou" is also widely cultivated.

Amur grape has huge commercial potential, for containing abundant phenolic compounds, procyanidins, oligostilbenes, and stilbenes (Chen *et al.*, 2018). Its fruits are dark purple with small berries, more acidic than *Vv*, and are used as raw materials for claret, sweet red wine, semi-dry wine, semisweet red wine, and brandy (wine-world.com). Young twigs of the Amur grape have spider silk-like hairs (Figure 1a). Amur grape is extremely cold-tolerant and can survive temperatures as low as -40°C (Wan *et al.* 2008; Xu *et al.* 2014). Now, Amur grape has become a valuable germplasm resource for grape breeding and wine production. Cv. "Beibinghong", a cross of *Vitis amurensis* and *Vitis vinifera,* is a good example in cold resistance, having ability to tolerate temperatures of -37 °C and therefore used for preparing fancy red ice wine. Recent pharmacological studies showed that Amur grape possesses anti-inflammatory, antimicrobial, antioxidant, and anticancer properties (Chen *et al.*, 2018).

The genome sequence of *Vitis Vinifera* cv. "Pinot Noir (PN40024)", a kind of Eurasian grape species, was deciphered (Jaillon *et al.,* 2007; Canaguier *et al.,* 2017) and has been used as the standard grape genome for grapes' gene analyses (Liang *et al.,* 2019; Ma *et al.* 2018). A high-quality genome sequence of *Vv* cv. Cabernet Sauvignon was sequenced in 2021 with the Hi-C technology. The genome for *V. riparia*, an American grape species, was recently decoded (Girolletet al., 2019). *Va* IBCAS1988 genome was sequenced in 2021 with a genome size 604.56 Mb and N50 282,256 bp (Wang *et al.,* 2021). To date, there is no high-quality sequence for genomes of the East Asian grape species, hindering biological and breeding research efforts to improve grape cultivars.

Here, we present a high-quality sequence for the *Vitis amurensis* cv. "Zuoshan No.1" genome. The availability of the genome sequence of Amur grape will lay a solid foundation to understand grape biology and develop new breeding lines, by deepening our insight into grapes' evolution and genetic variability, the molecular basis for cold adaptation and sex determination of grapes, etc.

Results

A high-quality Amur grape reference genome

The study used the female Amur grape "Zuoshan No.1" (Figure 1E-G) as the material. The Amur grape genome sequence was *de novo* assembled with 52.69 Gb (~101.46 fold coverage) of Nanopore reads, 98.79 Gb (~189.23 fold coverage) of MGISEQ short reads, and 58.12 Gb (~111.34 fold coverage) of Hi-C data. A k-mer analysis of Illumina reads estimated the Amur grape genome to be ~532 Mb, with DNA heterozygosity 1.20% (Table 1). The final assembled genome sequence reached to ~522 Mb, covering ~97.5% of the estimated genome (BUSCO value = 97.5), with a GC content 0.345 (Table 1). Moreover, the assembly comprised 56 contigs with 2.5 Mb of contig N50. The N50 of *Va* IBCAS1988 is much shorter (282,256 bp) (Wang *et al.,* 2021) than that of the *Va* Zuoshan No.1 genome, suggesting quality differences in the genome assemblies. The resultant 615 contigs accounted for 98.31% (~513.47Mb) of the total assembled genome, anchored onto 19 pseudo-chromosomes using Hi-C reads (Figure 2A and B). Repetitive elements accounts for 59.21% of the genome sequence (Table 1), much higher than the finding (47.06%) in *Va* IBCAS1988 genome (Wang *et al.,* 2021).

We predicted 27,635 protein-coding genes using an integrative strategy combining *de novo* gene prediction, protein-based homology search, and transcript data from RNA sequences of various tissues. The ~96.4% of annotated protein-coding genes could be annotated by at least one public database

(InterPro, Nr, GO, KOG, Swissprot, TrEMBL and KEGG) (Table S1). Moreover, 231 miRNAs, 651 tRNAs, 291 rRNAs, and 412 snRNAs were annotated. LTR accounts for the largest proportion of transposable elements, making up ~46.39% of the genome.

The homologous gene dotplotting analysis showed a big difference in arrangement order of homologous genes between the *Va* Zuoshan No.1 and *Va* IBCAS1988 genomes (Figure S1). Moreover, homologous gene dotplots and chromosome synteny analysis showed highly similar arrangement order of homologous genes across *Va* Zuoshan No.1, *Vv* PN40024, *Vv* Cabernet Sauvignon, and *Vr* genomes (Figure 2C-E). *Va* has large inversions on chromosomes 3, 9, 10, and 18 relative to *Vv* and *Vr* (Figure 2D, E). Besides, *Va* IBCAS1988 chromosomes have large inversions on *Va* Zuoshan No.1 chromosomes 1, 5, 15, 16, 18, and 19 (Figure S1). The homologous gene dotplots also showed a major difference between *Va* IBCAS1988 and *Va* Zuoshan No.1 genomes on chromosome 13. The homologous region between chromosome 13 of *Va* IBCAS1988 and *Va* Zuoshan No.1 is clearly divided into two parts. The anterior part of chromosome 13 (~19.6 Mb) of *Va* IBCAS1988 is homologous to the posterior part of *Va* Zuoshan No.1 chromosome while the posterior part of *Va* IBCAS1988 chromosome is homologous to the anterior part of *Va* Zuoshan No.1 chromosome (~12.6 Mb) (Figure S1). Previous studies on the *Va* IBCAS1988 genome indicated a major difference between *Va* IBCAS1988 and *Vv* PN40024 on chromosome 13 (Wang *et al.,* 2021). Here, we found a minimal difference between *Va* Zuoshan No.1 and *Vv* PN40024 on chromosome 13 (Figure 2C).

~~In order to discover the differences between the previous genome and our genome. The "putative structural variation (SV) analysis" of genomes based on the genome alignment showed big differences in the genome structures between *Va* IBCAS1988 and *Va* Zuoshan No.1 genomes. This study identified 171~~

~~inversions (95.46Mb), 794 translocations (9.86 Mb), and 481 duplications (3.41Mb) (Table S2) between the Amur grape Zuoshan No.1 and *Va* IBCAS1988 genomes. The SV analysis of the genome alignment proved the large inversions on chromosomes 1, 5, 15, 16, 18, and 19 shown by the homologous gene dot plots. Moreover, the SV analysis also indicated large inversions on chromosomes 3, 8, 9, 10, 11, 14, and 17. However, the 122 duplications between the Amur grape Zuoshan No.1 and *Va* IBCAS1988 genomes involved 97 protein-coding genes. The 99 translocations between the Amur grape and *Va* IBCAS1988 genomes involved 82 protein-coding genes.~~

~~As to the SNPs, a total of 1,382,913 SNPs were identified between two *Va* cultivars, Zuoshan No.1 and IBCAS1988 (Table S2), including 10,733 protein-coding genes. Besides, a total of 379,188 indels were identified, with insertions affecting 7654 protein-coding genes, and deletions 8356 ones.~~

~~The Amur grape Zuoshan No.1 genome had 12018 genomic-specific segments (17.51 Mb), and the *Va* IBCAS1988 had 8465 genomic-specific segments (12.25 Mb), including 2271 *Va*-specific PAV (presence-absence variation) and 237 *Va* IBCAS1988 specific PAV genes (Table S3). Many *Va* Zuoshan No.1-specific PAV genes include pentatricopeptide repeat-containing protein and TF genes (Table S4). Many *Va* IBCAS1988-specific PAV genes include hypothetical protein genes (Table S5).~~

Comparative genomics analysis of three grape species

Duplicated genes produced by the core-eudicot-common hexaploidy (ECH) (Jaillon *et al.*, 2007), common to major core eudicot, may have contributed to the divergence of the species. We found that *Vr* contains more duplicated genes (5231, 20.04%) than *Va* (4494, 16.26%) and *Vv* (3993, 12.49%) (Table S2), showing 800-1300 duplicated gene composition difference. This should have been caused by unbalanced gene losses in the ECH-produced regions.

Polyploidization was anticipated to result in genome instability, characterized by extensive DNA rearrangements and gene losses, especially in the early days after the tripling of the genome. Notably, the present finding suggest much divergent evolutionary patterns among these grapes and the ECH, as ancient as 130 million years ago (Mya), played and has still played a non-negligible role in grape species divergence and biological innovation.

The Ks distribution of duplicated genes generated by the ECH shows an up to 3.2% difference in their evolutionary rates. *Vv*, *Vr*, and *Va*, had peaks located at 1.29 (+/- 0.150), 1.27 (+/-0.165), and 1.25 (+/- 0.135), respectively, indicating that *Va* evolved the slowest among the three species (ANOVA, *P* Value=0.023) (Figure 3a, Supplementary Table 7). The Ks peaks derived from the orthologous gene pairs between species were located around 0.026 (Figure 3b, Table S3). Considering the proposed occurrence time of the ECH, and the Ks between paralogous generated by each grape ECH event was ~1.25 (Figure 3a), we calculated the three species should have diverged ~2.3-2.8 Mya based on the Ks between orthologous of every two grapes was ~0.025-0.026 (Figure 3b).

By using the concatenated multiple sequence alignment of single-copy genes (5929) inferred by Orthofinder, between the three *Vitis* genomes and *Aquilegia coerulea*, acting as an outgroup, a phylogenetic tree was constructed and showed that the *Va* was more closer to *Vv* in evolution history (Figure 3C). When Arabidopsis or Boxwood was used as an outgroup, we obtained the same inference (Figure S2).

The SVs analysis revealed large DNA inversions on *Va* chromosomes 3, 9, 10, and 18, relative to *Vv* (both Cabernet Sauvignon and PN40024) and *Vr*, consistent with the homologous gene dotplots (Figure 2; Figure 4; Figure S3). Large DNA rearrangements contributed to the divergence of different grapes. DNA rearrangements in *Vitis* chromosomes 19 were inferred across the three species, large inversions were found between the *Vr* and *Vv* chromosomes 3, 7,

and 18, especially one large DNA segmental translocation was detected on *Vr* and *Vv* chromosome 18 (Figure 4; Figure S3).

Genome instability due to the ECH resulted in species-specific regions, further contributing to their biological divergence. A comparison of *Va* and *Vv* showed that Amur grape had 12105 specific genomic segments, accumulated to 15.46 Mb, whereas *Vv* had 8465 (11.73 Mb), with 1609 *Va*-specific PAV genes and 79 *Vv*-specific PAV genes positioned within these specific genomic segments (Table S4). The *Va*-specific PAV genes include those respectively encoding cold-responsive protein kinase 1, ethylene-responsive transcription factor, flavonoid 3',5'-hydroxylase 2, DELLA2, heat shock 70 kDa protein, methyl-CpG-binding domain protein, and WRKY transcription factor 14 genes (Table S5). In contrast, *Vv*-specific PAV genes contain several uncharacterized genes (Table S6). We found 16 *Va*-specific PAV genes were cold resistance related genes. The enriched GO terms of *Va*-specific PAV genes ($P<0.05$) comprise monolayer-surrounded lipid storage body (GO:0012511) and stress fiber (GO:0001725). The enriched KEGG pathway of *Va*-specific PAV genes ($P<0.05$) comprise Calcium signaling pathway (ko04020), and phosphonate and phosphinate metabolism (ko00440).

A comparison of Amur grape and *Vr* genomes found that *Va* had 7142 specific genomic segments (6.53 Mb), while *Vr* had 3489 segments (7.68 Mb), including 266 *Va*-specific (Table S7) and 88 *Vr*-specific PAV genes (Table S8). We found the Vitis03G0393, Vitis11G0466, Vitis16G0074, Vitis14G1585, and Vitis05G0085 from *Va*-specific PAV genes were cold resistance related genes. The enriched GO terms of *Va*-specific PAV genes ($P<0.05$) comprise negative regulation of stomatal complex development (GO:2000122), regulation of vacuolar transport (GO:1903335) and cellular water homeostasis (GO:0009992). The enriched KEGG pathway of *Va*-specific PAV genes

(P<0.05) comprise Cutin, suberine, and wax biosynthesis (ko00073), and fatty acid biosynthesis (ko00061).

We also identified 11503 *Vr* specific (19.93 Mb) and 5439 *Vv* genomic-specific segments (14.26 Mb), including 1556 *Vr*-specific and 224 *Vv*-specific PAV genes positioned within these specific segments.

We identified 349 cold resistance related genes (CRGs) in *Va*, 437 in *Vv* and 404 in *Vr*, accounting 1.26%, 1.37% and 1.01% of the total number of genes in each genome, respectively. An *in silico* interaction network of key cold resistance related genes in *Va*, *Vv*, and *Vr* was constructed as previously described (Song *et al.,* 2020). Though having the fewest CRG genes in *Va*, a research into the robustness of CRG regulatory networks showed that *Va* CRG regulatory network was more robust in *Va* than those in *Vv* and *Vr* (Table 2).

In addition to cold resistance related genes, we studied disease resistance genes. We identified totally 64 nucleotide binding site (NBS) genes in *Va*, which is less than in *Vv* (158 NBS family genes) and in *Vr* (172 NBS family genes). 34 *Va* NBS family genes (53.13% of *Va* NBS genes) were related to the ECH and 42 (65.63% of *Va* NBS family genes) were related to the tandem duplication. The Ks value of 4 pairs *Va* NBS genes is less than 0.025 (Figure S4), showing likely being generated after divergence of *Va* from the other grape species. We found 151 ancestral *vitis* NBS family genes based on the collinear relationship across three grapes and *A. coerulea* and also found that 124 of the ancestral NBS genes were lost in *Va*. In contrast, *Vv* lost 57 and *Vr* lost 54 genes. The evolutionary tree of NBS genes supports this viewpoint. Figure S5 showed in many branches which existed more than one *Vv* or *Vr* genes contain one or none *Va* gene.

The regulatory mechanisms of nutrient accumulation at four developmental stages of *Vitis amurensis*

Berries of *Va* are nutrient-richand contain many phenolic compounds, such as anthocyanins and procyanidins, and stilbenes such as resveratrol (Chen *et al.*, 2018). Besides, *Va* berries can be used in wine making because it is rich in sugars and organic acids.

To reveal the molecular mechanisms of nutrient accumulation in *Va* cv. Zuoshan No.1 berries, we profiled the RNA-seq (Table 3) and calculated the gene expression changes of key biological pathways during grape development, and correlated the gene expression changes and metabolite accumulation. Four development stages of berries, including Stage1 (S1, late period of berry expansion), Stage 2 (S2, Veraison), Stage 3 (S3, the period when berries change color completely), and Stage 4 (S4, maturity stage) (Figure 5a) were analyzed.

The total sugar content increased significantly, and titratable acidity content was decreased significantly from S1 to S4 (Figure S6). Widely targeted metabolomics analysis showed that succinic acid and 2-isopropyl malic acid contents decreased significantly from S1 to S4. However, there was no significant change in tartaric acid content in berries at the four stages.

Widely targeted metabolomics analysis showed that the content of D-glucose, a monosaccharide, significantly increased between S1 and S4. Trehalose and sucrose, two disaccharides, significantly increased at S3.

Here, RNA-seq analysis is used for joint analysis with the widely targeted metabolomic analysis. Two starch and sucrose metabolism pathway (Ko00500) key genes (metabolic pathways key genes were judged based on KEGG database, and the metabolic pathways key genes were the genes could be mapped in related KEGG pathways) for sucrose synthesis (sucrose-6-phosphatase gene (Vitis08G0202)) and trehalose synthesis (alpha, alpha-trehalase gene (Vitis02G0204)) were significantly up-regulated (fold > 2, P value < 0.05) at S4 compared to S1. At S3, the key genes for sucrose synthesis (sucrose-phosphate synthase (Vitis05G0995 and Vitis18G2061)) and trehalose

synthesis (alpha, alpha-trehalase gene (Vitis02G0204)) were significantly up-regulated. In contrast, the key gene of trehalose degradation beta-fructofuranosidase (Vitis16G0541) was significantly down-regulated (fold < 0.5, P value < 0.05) at S3. At S4, the key genes for sucrose synthesis (sucrose-6-phosphatase (Vitis08G0202)), sucrose-phosphate synthase (Vitis05G0995 and Vitis18G2061), trehalose synthesis (alpha, alpha-trehalase gene (Vitis02G0204)) were significantly up-regulated. In contrast, sucrose metabolism-related genes (beta-fructofuranosidase, Vitis16G0541, and Vitis02G0497) were significantly down-regulated at S4 than S2.

Widely targeted metabolomics analysis showed that the resveratrol content increased significantly from S1 to S4, but there was no significant change in the resveratrol content between berries sampled at S1 and S2. In the stilbenoid, diarylheptanoid, and gingerol biosynthesis pathway (Ko00945), stilbene synthase competitively convert the p-Coumaroyl-CoA synthesized by trans-cinnamate 4-monooxygenase to resveratrol. In contrast, shikimate O-hydroxycinnamoyltransferase and 5-O-(4-coumaroyl)-D-quinate 3'-monooxygenase could convert the same to caffeoylquinic acid. From S1 to S2, the key genes for caffeoylquinic acid synthesis, shikimate O-hydroxycinnamoyltransferase gene (Vitis11G0727) and 5-O-(4-coumaroyl)-D-quinate 3'-monooxygenase gene (Vitis08G0528) were significantly down-regulated. From S1 to S3, the trans-cinnamate 4-monooxygenase gene (Vitis06G0746) was significantly up-regulated, and the shikimate O-hydroxycinnamoyltransferase gene (Vitis11G0727) was significantly down-regulated. Meanwhile, the content of caffeoylquinic acid was significantly down-regulated, and the resveratrol content increased significantly. From S1 to S4, the trans-cinnamate 4-monooxygenase gene (Vitis06G0746) was significantly up-regulated, while the shikimate O-hydroxycinnamoyltransferase (Vitis11G0727) and 5-O-(4-coumaroyl)-D-quinate 3'-monooxygenase

(Vitis08G0528) genes were significantly down-regulated. Meanwhile, the content of caffeoylquinic acid decreased significantly, while the resveratrol content increased significantly. The canonical correlation analysis (CCA) showed that the trans-cinnamate 4-monooxygenase gene (Vitis06G0746), related to the resveratrol content, and shikimate O-hydroxycinnamoyltransferase (Vitis11G0727) and 5-O-(4-coumaroyl)-D-quinate 3'-monooxygenase (Vitis08G0528) gene are also related to the caffeoylquinic acid content (Figure 5b, C).

Divergently expressed genes (DEGs) were checked (Table 4) between S1-S4. They included the putative ripening-related gene (Vitis08G0132), the chalcone synthase 2 gene (Vitis05G0886), the flavonoid 3-O-glucosyltransferase gene (Vitis16G0156), and sugar transporter SWEET2a (Vitis10G0718) gene, the ethylene-responsive transcription factor 3-like gene (Vitis12G0453), the UDP-glucose: flavonol synthase/flavanone 3-hydroxylase gene (Vitis08G0277), and some bHLH, MYB transcription factors, and WD repeat-containing protein.

Exploring Amur grape "putative sex-determining region (SDR)"

A previous study showed that *Vv* SDR is located on the segment between 4,801,876 bp and 5,061,548 bp on chromosome 2 and involves 15 protein-encoding genes (Massonnet *et al.,* 2020). The 5' terminus of the SDR segment has a PPR-containing protein-coding gene and the 3'terminus has an APT3 gene (Massonnet *et al.,* 2020). Another study showed the SDR is located on the segment between 4,810,929 bp and 4,921,949 bp on chromosome 2 which overlapped with the previous SDR region (Badouin et al.,2020). We explored the *Vv* SDR syntenic region between 5055465 and 5198824 on Amur grape chromosome 2, showing good synteny between two species. To determine whether the Amur grape putative SDR region is related to sex, we

adopted the selective sweep method, which allows analysis of fewer samples than GWAS (Kim *et al.*, 2020).

Through re-sequencing and analyzing SNPs in 24 Amur grape individuals, including five male, ten female, and nine hermaphroditic individuals (including four tetraploids), we performed phylogeny analysis and examined genetic population structure among the Amur plants (Table S9). A set of 2588125 high-quality SNPs were inferred and explored to characterize the relationship among these 24 *Va* individuals and the *Va* reference genome (this study). A phylogenetic tree was constructed based on the SNPs of the 24 *Va* individuals and indicated that the individuals could be divided into two groups. However, the grouping proved unrelated to sex division (Figure 6A). Then, a genetic population structure analysis was performed and showed that the individuals could be divided into two groups, and the grouping pattern proved unrelated to sex division, either (Figure 6B). Alternatively, based on the SNPs in the SDR regions of the 24 individuals, we reconstructed a phylogenetic tree. Notably, we found that the tree had groups consistent to sex division (Figure 6C). Grossly, the above analysis showed that the genetic relationship at the population level is not related to sex division, while the SNP difference in SDR regions is related to gender.

Therefore, a selective sweep analysis was performed according to the sex groups rather than the population structure. One of the selected regions inferred (both $\theta\pi$ ratios and FST in the top 5% range) from the comparison of male and female accessions totally overlapped the Amur grape putative SDR. We found that the Amur grape putative SDR contains 16 genes (Vitis02G0494, Vitis02G0495, Vitis02G0496, Vitis02G0497, Vitis02G0498, Vitis02G0499, Vitis02G0500, Vitis02G0501, Vitis02G0502, Vitis02G0503, Vitis02G0504, Vitis02G0505, Vitis02G0506, Vitis02G0507, Vitis02G0508 and Vitis02G0509), respectively homologous to *Vv* SDR genes associated with sex

(Figure 6D). Another selected region inferred between male and hermaphroditic groups overlapped part of the Amur grape putative SDR gene Vitis02G0494 (Figure 6D). However, none of the selected regions inferred between female and hermaphroditic groups overlapped the Amur grape putative SDR (Figure 6D).

Previous studies showed that the SDR of *Vv* M and H haplotype genes include those respectively encoding PPR-containing proteins, a YABBY transcription factor (VvYABBY3), a VviSKU5 (Skewed5), a beta-fructofuranosidase, an aldolase, a trehalose-6-phosphate phosphatase (TPP), an inaperturate pollen1 (VviINP1), an exostosin family protein, KASIII, two TPR-containing protein, a PLATZ transcription factor, three Flavin-containing monooxygenases (FMO), a hypothetical protein (VviFSEX), a WRKY transcription factor, and an Adenine phosphoribosyltransferase (VviAPT3). The SDR of *Vv* F haplotypes included genes encoding a PPR-containing protein, a YABBY transcription factor (VviYABBY3), a VviSKU5, a beta-fructofuranosidase, an Aldolase, a Trehalose-6-phosphate phosphatase (TPP), an inaperturate pollen1 (VviINP1), an exostosin family protein, a KASIII, a PLATZ transcription factor. The F haplotypes also had four flavin-containing monooxygenases, a hypothetical protein (VviFSEX), a WRKY transcription factor, and an Adenine phosphoribosyltransferase (VviAPT3), respectively (Massonnet *et al.*, 2020).

In *Va*, the F haplotypes included 16 genes encoding a PPR-containing protein, a YABBY transcription factor (VviYABBY3), a VviSKU5, a beta-fructofuranosidase, a fusion protein of aldolase and trehalose-6-phosphate phosphatase (TPP), an inaperturate pollen1 (VviINP1), an uncharacterized protein (homologous to VIT_202s0154g00120 of *Vv* PN40024), an exostosin family protein, a KASIII, and a PLATZ transcription factor. Besides, the F haplotypes contained three flavin-containing monooxygenases, a hypothetical protein (VviFSEX), a WRKY transcription factor, and an adenine

phosphoribosyltransferase (VviAPT3), respectively. The SDR of *Va* F haplotypes had an additional uncharacterized protein gene (Vitis02G0500) and lost one FMO gene, as compared to the *Vv* Cabernet Sauvignon F haplotypes (Figure 7).

In *Vv*, the female-specific DNA polymorphisms at -13, -5, and +2 bp may reduce transcription and/or alter mRNA decay of the female PLATZ transcription factor allele on *Vv* SDR (Iocco-Corena *et al.,* 2021). However, *Va* lacks these DNA polymorphisms upstream of the ATG of the PLATZ transcription factor gene in the Amur grape putative SDR. In other words, there is no difference of DNA polymorphisms between female, hermaphrodite, and male *Va* individuals within 20 bp upstream of the PLATZ transcription factor gene.

We found some other selected regions may be related to sex-determination ($\theta\pi$ ratios and FST estimates in the top of 5%). For example, the region (chr3: 6730001bp-8840000 bp) from the comparison of male and female accessions, the region (chr16:3565001bp-4830000bp) from the comparison of male and female accessions, the region (chr16:15590001bp-18015000bp) from the comparison of male and female accessions, and the region (chr17:11450001bp-13785000bp) from the comparison of male and hermaphroditic accessions. In these regions, there are the MADS-box protein AGL62 genes (Vitis03G0709, Vitis03G0710, Vitis03G0711 and Vitis03G0712) from chr3: 6730001bp-8840000 bp. We found that there are the Pentatricopeptide repeat-containing protein genes in chr16:3565001bp-4830000bp (Vitis16G0214 and Vitis16G0215), in chr16:15590001bp-18015000bp (Vitis16G0545), and in chr17:11450001bp-13785000bp (Vitis17G0924). These Pentatricopeptide repeat-containing genes are the homologs of putative SDR gene Vitis02G0494 (a Pentatricopeptide repeat-containing gene).

Discussion

Vitis mainly include the Eurasian grape, East Asian grape, and the American grape. The sequence of the *Vv* genome (PN 40024), of a Eurosian grape, was published in 2007 (Jaillon *et al.*, 2007) and incurred researches to understand grape biology (Wang *et al.*, 2018a; Wang *et al.*, 2018b; Wang *et al.*, 2019; Liang et al., 2019). Given that its genome has only undergone one polyploidization, the ECH, after split from the basal endicots, it is often used as outgroup reference for studying the other endicot genomes (Wang *et al.*, 2018a). The sequence of *Vr* RGM genome, of an American grape, was published in 2019 (Girollet *et al.*, 2019). The Cabernet Sauvignon genome sequence, a *Vv* variety, was later published, being assembled using the Hi-C technology (Massonnet *et al.*, 2020).

Here, we assembled the genomes of *Vitis amurensis*. The above analysis showed that the assembled *Va* genome sequence reported here is much improved as to the previous one. Based on the present genome sequence, we discovered a likely assemble error in chromosome 13 of the previous grape genome sequence.

The availability of multiple grape genome sequences allows for comparing and analyzing gene sequences of several grape species so that we could detect the evolution history and other important comparative genomics information across the grapes. In the present study, gene synteny analysis showed the arrangement order of orthologous genes between *Va*, *Vv*, and *Vr* genomes are much conservative. We have to note here, genome instability due to the ECH, as ancient as 130 Mya, contributed to the divergence of the grape genomes. Actually, a characterization of homologous genes showed that *Vr* is the most conservative one among the three. *Vr* contains more paralogous genes produced by the ECH, comparing with *Va* and *Vv*, showing the conservativeness of the *Vv*

genome, and implying that *Vr* may much resemble their common ancestor. Moreover, more orthologous genes were preserved between *Vr* and the other two grapes (*Vv* and *Va*) than between *Vv* and *Va*. Besides, phylogenetic analysis supported that *Vr* was the first one to split from the other grapes. These show that *Vr* genome may be taken as a better reference for *Vitis* biology and even all the eudicots. The present finding that a 130-Mya polyploidization contributed to the species divergence of grapes reminded us that it may has played a non-negligible role in grape divergence and genetic innovation, as to a similar finding in rice and other grasses (Wang *et al.,* 2011). Possibly, the genomic feature of thousands of duplicated genes in crops could be manipulated to breed high-yield and/or high-quality crops.

Species-specific genomic segments also make up a large proportion of the genomes for all the three species, implying significant genomic differences among the three species. SV may generate phenotypic differences (Chawla *et al.,* 2021), for example, *Va* being more cold-resistant than the other two species. Biased sequence preservation in *Va* may have contributed to its capability to resist coldness. Sequence variations, especially incurred by the ECH, may generate phenotypic differences. Here, we found that the *Va*-specific PAV genes contain cold resistance-related genes (CRG), such as those encoding the ethylene-responsive transcription factors and cold-regulated 413 inner membrane proteins. The robustness of CRG regulatory networks in *Va* is higher than those in *Vv* and *Vr*. This showed that the stronger cold resistance of *Va* may be related to the preferential preservation of cold-related genes, and *Va* cold-related genes may constitute a more robust and effective interaction network. A study of CRGs in angiosperms, including *Vv*, supported the expansion of them was significantly related to the occurrence of polyploidization (Song *et al.,* 2020). The present study provides further evidence that the ECH,

though occurred 130-150 million ago, still contribute to enhance resistance to coldness in plants.

Multi-omics analysis of the *Va* was performed to understand the mechanism underlying the regulation of nutrient accumulation in the *Vitis amurensis* over its developmental period. *Va* is sourer than *Vv*, a factor that needs urgent resolution by breeders, in that high sourness makes the fruits taste unpleasant, and the wine tastes unbalanced. The present analysis revealed that the total sugar increased continuously during the "Zuoshan No.1" berry development, while the total sugar content in this species/variety was almost equal to that in *Vv*. Further analyses revealed that the total acid content decreased significantly during "Zuoshan No.1" berry development. Wide target metabolome analysis showed no decrease in tartaric acid content during berry of "Zuoshan No.1" development, while those of citric acid and several malic acid types decreased significantly in "Zuoshan No.1". Combined transcriptome and metabolome analyses revealed that sucrose-6-phosphatase, alpha, alph-trehalase, and beta fructofuranosidase genes jointly regulate the accumulation of sucrose and trehalose during "Zuoshan No.1" development. The DEGs during S1-S4 comprises the ripening-related protein-like gene and chalcone synthase gene, which is the first enzyme activated the flavonoid biosynthetic pathway (Waki *et al.,* 2020), UDP-glucose: flavonoid 3-O-glucosyltransferase gene, which promotes anthocyanin accumulation in grape (Yamazaki *et al.,* 2021), SWEET gene, which is a sugar transporter and uniporter gene (Chong *et al.,* 2014), ethylene-responsive transcription factor 3-like gene, certaining bHLH, MYB transcription factor genes, and WD repeat-containing protein genes. Most of these genes may be involved in *Va* berry ripening. Therefore, we infer that these DEGs may play a central role in the process of *Va* berry ripening.

The *Va* genome was further analyzed to understand the relationship between different *Vitis amurensis* phenotypes and genotypes. Given the scarcity of

hermaphroditic *Vitis amurensis* in nature, we explored the sex determination region in this grape. Previous studies have uncovered the SDR of *Vv* (Massonnet *et al.,* 2020). The three *Vv* types (varieties) are determined by the genotype at the SDR. Males are heterozygous for the male and female haplotypes (MF), females are homozygous (FF), and cultivated *Vv* hermaphrodites are either homozygous for hermaphrodite haplotypes (HH) or heterozygous (HF). Previous studies have shown that the SDR of *Vv* M and H haplotypes have 2 TPR-containing protein genes and 3 Flavin-containing monooxygenases (FMO), and the SDR of *Vv* F haplotypes contain 4 Flavin-containing monooxygenases (FMO) without TPR-containing protein gene (Massonnet *et al.,* 2020). This implies that the type or number of genes in the SDR of grapes vary between sexes.

Further analyses to clarify whether SDR is also present in *Va* revealed that one fragment from *Va* is homologous to *Vv* SDR in *Vitis amurensis* and call it the Amur grape putative SDR. In the present study, we sequenced the female *Va* genome. Therefore, further analyses were performed to explore whether the types and numbers of genes in female *Va* putative SDR were consistent with those in the female *Vv* SDR. Selective sweep analysis of male, female, and hermaphroditic individuals revealed one selective region from the comparison of male and female accessions overlapped total the Amur grape putative SDR including 16 genes all homologous to *Vv* SDR genes (such as APT3 gene, PLATZ transcription factor gene and TPP gene). A previous study inferred that VaAPT3 gene (orthologous to Vitis02G0508 in the present *Va* genome), located in Amur grape putative SDR, was associated to sex determination (Men *et al.,* 2021). Recently, the PLATZ transcription factor gene has been found to determine the grape's sexuality (Iocco-Corena *et al.,* 2021). The TPP gene also located in Amur grape putative SDR and TPP genes in many species have also been associated with sex determination, flower development, reproduction, and synthesis of plant hormones, among other functions (Kataya *et al.,* 2020; Qiu *et*

556 *al.,* 2020). PtTPPs displayed a specific expression pattern in seven
557 developmental stages of Populus male and female floral buds (Gao *et al.,* 2021).
558 Trehalose-6-phosphate phosphatase (TPP) controls inflorescence architecture in
559 maize through sugar signal modification (Satoh-Nagasawa *et al.,* 2006). This
560 suggests that the hypothetical *Va* SDR may be related to sex determination. One
561 selected region was from the comparison of male and hermaphroditic accessions
562 overlapped part of Amur grape putative SDR including only one gene
563 (PPR-containing gene), which is homologous to the *Vv* SDR gene
564 PPR-containing gene. Additionally, some selective regions may be related to
565 *Va* sex determining. Previous study indicated that some PPR-containing
566 proteins genes could restore fertility to cytoplasmic male-sterile plants
567 (Bentolila *et al.,* 2002; Koizuka *et al.,* 2003; Kazama *et al.,* 2003; Hu *et al.,*
568 2012). A study on RNA-Seq analysis of three flower sex types in grapevine
569 showed some grape PPR-containing genes could be essential for carpel
570 development (Ramos *et al.,* 2014). The study also inferred that PPR-containing
571 gene may be essential for the perfect development of sexual floral organs. Here,
572 we found that some other selected regions comprises many PPR-containing
573 protein genes, allowing for further exploration of their likely contribution to
574 sex determination.

575 However, no selected region overlapped the Amur grape putative SDR
576 from the comparison of male and hermaphroditic accessions was found. We
577 found that the type and arrangement order of all genes in the putative SDR of
578 female *Va* and the SDR of *Vitis vinifera* were similar. However, compared with
579 *Vv* F haplotypes, the putative SDR of female *Vitis amurensis* has lost one FMO
580 gene and contains one extra Uncharacterized protein. The Uncharacterized
581 protein gene has never been found in SDR of *Vv* Cabernet Sauvignon and *Vv.*
582 *sylvestris* (Massonnet *et al.,* 2020). The role of the Uncharacterized protein gene
583 in *Vitis amurensis*sex determination needs further exploration.

In the other selective region (chr17:11450001bp-13785000bp), we found from the comparison of *Va* male and hermaphroditic accessions, there are many MADS-box protein AGL62 genes. Previous study inferred MADS-box protein genes may be also related to developing male and female flowers (Hardenack *et al.*, 1994; Ainsworth *et al.*, 1995). We consider that the region (chr17:11450001bp-13785000bp) may be also related to *Va* sex-determining. Plus those selective regions that have been found to contain the PPR-containing proteins genes, it should be explored whether *Vitis amurensis* sex may be determined by several different genomic regions in the future.

Materials and methods

Plant materials and DNA sequencing

Fresh leaves and stems of *Vitis amurensis* cv. "Zuoshan No. 1" were sampled for DNA extraction and sequencing. Total genomic DNA was extracted using the CTAB (cetyltrimethylammonium Ammonium Bromide)) method (Tel-Zur *et al.*, 1999). The library for ONT sequencing (Oxford Nanopore Technology, Oxford, UK) was constructed using large (>15 kb) DNA fragments with the SQK-LSK109 Ligation Sequencing Kit and sequenced using the ONT platform. Adapters and low-quality nucleotides (with <7 mean quality score) were trimmed off. Paired-end libraries with 350 bp insert sizes were constructed following the manufacturer's protocols and sequenced using the MGIseq 2000 platform (MGI Tech Co. LTD, Guangdong, China). The MGIseq reads were filtered using the SOAPnuke1.5.6 online software (https://github.com/BGI-flexlab/SOAPnuke). For the high-throughput chromosome conformation capture (Hi-C) analysis, fresh leaves and stems of *Vitis amurensis* cv. "Zuoshan No. 1" were treated following previous methods (Yang *et al.*, 2020).

Genome size and heterozygosity estimation

The genome size was estimated from the MGIseq reads using k-mer analysis, and the k-mer depth-frequency distribution was generated using the jellyfish software (Marçais and Kingsford, 2011). The genome size and heterozygosity were calculated using the genomeScope software (Vurture *et al.*, 2017).

***De novo* genome assembly**

Long ONT reads were corrected using Canu (https://github.com/marbl/canu/releases), and *de novo* assembled using the Smartdenovo software (https://github.com/ruanjue/smartdenovo). The Racon (https://github.com/isovic/racon) and Medaka (https://github.com/nanoporetech/medaka) software were applied to polish the assembled contigs. The polished, assembled data were then corrected using Pilon (v.1.22, https://github.com/broadinstitute/pilon/). Next, the Haplomerger2 pipeline reduced the assembly to ~522 Mb. After that, the purged assembled contigs were anchored into 19 pseudochromosomes using the Juicer with the parameters "-g draft -s MboI" and 3D de novo assembly (3D-DNA) pipeline with the parameters "-m haploid -r 2" (Durand *et al.*, 2016; Dudchenko *et al.*, 2017).

Genome annotation and gene prediction

Transposable elements (TEs) were predicted using combined homology-based comparisons with RepeatMaskerv4.0.7, RepeatProteinMask v4.0.7, and *de novo* approaches with Piler (Edgar and Myers, 2005) (http://www.drive5.com/piler/), RepeatScout, and RepeatModeler. Tandem repeats were identified using the Tandem Repeats Finder v4.09 (Benson., 1999) (http://tandem.bu.edu/trf/trf.html) and LTR_FINDER v1.06 (Zhao and Hao, 2007) (http://tlife.fudan.edu.cn/ltr_finder/). Furthermore, the *Va* protein-coding

gene set was deduced by *de novo*, homology, and evidence-based gene prediction (transcriptome data) (Guo *et al.,* 2021). The transcript evidence included transcripts assembled from the RNA-Seq data of different tissues (leaf, stem, flowers, tendril, and fruit; Table S9). Next, the predicted genes were functionally annotated using a previous method (Guo *et al.,* 2021). Moreover, tRNA was identified using tRNAscan-SE 1.3.1 (Lowe and Eddy, 1997) (http://lowelab.ucsc.edu/tRNAscan-SE/), BLASTN identifiedrRNA, while INFERNAL (http://infernal.janelia.org/) identified miRNA and snRNA.

Genome homology inference

Protein sequences from one plant were searched against themself and those another plant genome using BLASTP (Camacho *et al.,* 2009) to find the best, secondarily best, and the other matches with E_value<1E-5. Dot plots were produced using the WGDI package of Python (Sun *et al.,* 2021). Colinear genes were inferred using the -icl subprogram contained in WGDI package with default parameters. Nucleotide substitution rates were estimated between colinear homologous genes using the YN00 program in the PAML (v4.9h) package implementing the Nei-Gojobori approach (Yang, 2007).

Species tree

A rooted species tree was inferred using the Orthofinder (version 2.5.4) with the "-M msa" parameter (Emms and Kelly, 2015; Emms and Kelly, 2019) based on 5929 single-copy genes.

Identification and characterization of cold-related genes

The cold-related genes in the three grape species and gene network robustness were determined following previous methods (Song *et al.,* 2020).

Structural variations analysis

The SV analysis was performed using the NUCmer program embedded in MUMmer with the parameters "-mumreference -g 1000 -c 90 -l 40". (Yu *et al*., 2021). Genomic-specific segment analysis was performed as follows. First, the query genome sequences was split into 500bp windows with a overlapping step size of 100bp. and then all the 500bp windows subsequnces was aligned against the references genome by BWA-mem with the parameters "-w 500 -M". The 500bp window that failed to align or aligned with less than 25% coverage were defined as the genomic-specific segments. The genes within the genomic-specific segments, were defined as the specific-genes of the query genome (species-specific PAV genes). Structural variation plots was drawn using ggplot2.

Widely targeted metabolomic analysis

Berries from four growth stages were used for this study, including Stage1 (S1, the late period of berry expansion), Stage 2 (S2, Veraison), Stage 3 (S3, the period when berries change color completely), and Stage 4 (S4, maturity stage). Three biological replicates from each stage were used for metabolome analysis. The Metware Biotechnology Co., Ltd. (Wuhan, China) performed the metabolome extraction and analysis using previous methods (Ma *et al.,* 2021). The metabolites were identified using the Metware database (MWDB), and metabolite abundances were determined according to the metabolite peak areas. Metabolites were considered as differentially accumulated when the variable importance in projection (VIP) was ≥1, and the absolute log2 (fold change) was ≥1. Metabolic pathways were constructed according to the KEGG database.

Total sugar and titratable acidity content analysis

Berries from four growth stages (S1, S2, S3 and S4) were grind into homogenate used for this study. Three biological replicates from each stage were used for metabolome analysis. We analyzed the total sugar content using sulfuric acid-anthrone colorimetric method and the instrument used is Lanbda 365 ultraviolet/visible spectrophotometer. We analyzed the titratable acidity content using acid-base titration and the standard solution is Sodium hydroxide standard solution (0.05 mol/L).

RNA-seq

Just as we analyzed the metabolome, berries from four stages, S1, S2, S3, and S4, were used for the RNA-seq analysis. Three biological replicates from each berry stage were used for RNA-seq analysis. Novogene Co. LTD performed the RNA-seq experiments on the Illumina Novaseq 6000 platform (Illumina, CA, USA) following the manufacturer's instructions and previously described steps (Zhang *et al.,* 2021). All paired-end reads were mapped to the *Va* genome using Hisat2 v2.0.5. Expression levels were calculated using the fragments per kilobase of exon model per million mapped fragments (FPKM). The DEGSeq R package (1.20.0) was used to identify differentially expressed genes (DEGs). Genes with adjusted *P* value < 0.05 and fold changes ≥ 2 or ≤ -2 were identified as differentially expressed genes (DEGs).

Integrated analysis of metabolome and RNA-seq

For the combined analysis of all metabolome and transcriptome, the canonical correlation analysis (CCA) was performed on the metabolome and RNA-seq data using CCA package in the R statistical environment (Gonzalez *et al.,* 2008). Next, the WGCNA (Weighted correlation network analysis) was performed on all RNA-seq and metabolome data using the WGCNA package in the R statistical environment.

723

724 **Population genetic analysis**

725 Essentially, 24 Amur grape genotypes were re-sequenced based on the *Va*
726 genome assembled in this studyand used for calling SNPs. Concurrently, the
727 ANNOVAR package was employed for population genetic analysis. Ten folds
728 *Va* genome size of clean reads data of 24 kind of Amur grapewere separately
729 obtained on the MGI2000 platform. The SNPs in LD were filtered using PLINK
730 with a window size of 50 SNPs (advancing 5 SNPs at a time) and a 0.5 r^2
731 threshold. The PCA was conducted using GCTA (v1.25.2) (Yang *et al.,* 2011).
732 Furthermore, the population structure was analyzed using FRAPPE, and the
733 MLtree was constructed using SNPhylo (Lee *et al.,* 2014) to clarify the
734 phylogenetic relationships.

735

736 **Selective sweep analysis**

737 Selective sweep analysis was performed using a previous method (Lin *et al.,*
738 2022). In briefly, Fst and θπ were used to detect candidate selective regions
739 between the three grape populations, Fst and θπ were calculated using
740 PopGenome (Brigida *et al.,* 2016; Pfeiferet *et al.,* 2014); regions with both θπ
741 ratios and FST estimates in the top of 5% were considered as the selection
742 regions.

743

744 ## Accession number

745 The *V. amurensis* genome project was deposited at NCBI under BioProject
746 number PRJNA868106 and BioSample SAMN30308461.

747

748 ## Acknowledgements

749 This research was supported by the Key Research and Development Project of
750 Shandong Province (2021LZGC025; 2022CXGC010605), Improved Variety

Program of Shandong Province (2020LZGC008), Shandong Natural Science Funding (ZR2020QC145), Scientific Research Guide Foundation of Shandong Academy of Grape (SDAG2021A01; SDAG2021A02), Shandong Academy of Agricultural Sciences, Introduction and Training of High-level Talents (CXGC2022E15), National Horticulture Germplasm Resources Center-Amur Grapevine Germplasm Resources Branch Center (NHGRC2022-NH04) and Species and variety Germplasm resources protection project (2130135).

Conflict interest statement

No conflict of interest declared.

References

Ainsworth C, Crossley S, Buchanan-Wollaston V, Thangavelu M, Parker J (1995) Male and female flowers of the dioecious plant sorrel show different patterns of MADS box gene expression. Plant Cell **7**:1583-1598.

Badouin H, Velt A, Gindraud F, Flutre T, Dumas V, Vautrin S, Marande W et al (2020) The wild grape genome sequence provides insights into the transition from dioecy to hermaphroditism during grape domestication. Genome Biol **21**:223.

Benson, G (1999) Tandem repeats finder: a program to analyze DNA se quences. Nucleic Acids Res **27**: 573-580.

Bentolila S, Alfonso AA, Hanson MR. A (2002) pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. Proc Natl Acad Sci U S A **99**:10887-10892.

Brigida G, Jan S, Troels P, Leahm S, Veerle S, Beatriz, HM, Adriaan M，et al(2016) Domestication and divergence of Saccharomyces cerevisiae beer yeasts. Cell **166**: 1397–1410.

778 **Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K,**
779 **Madden TL** (2009) BLAST+: architecture and applications. BMC
780 bioinformatics **10**: 421.
781 **Canaguier A, Grimplet J, Gaspero GD, Scalabrin S, Duchêne E, Choisne N,**
782 **Mohellibi N et al** (2017) A new version of the grapevine reference genome
783 assembly (12x.v2) and of its annotation (VCost.v3). Genomics Data **14**:
784 56-62.
785 **Chawla HS, Lee H, Gabur I, Vollrath P, Tamilselvan-Nattar-Amutha**
786 **S, Obermeier C, Schiessl SV et al** (2021) Long-read sequencing re
787 veals widespread intragenic structural variants in a recent allopolyplo
788 id crop plant. Plant Biotechnol J **19**: 240-250.
789 **Chen Q, Diao L, Song H, Zhu X** (2018) *Vitis amurensis* Rupr: A review of
790 chemistry and pharmacology. Phytomedicine **49**: 111-122.
791 **Chong J, Piron MC, Meyer S, Merdinoglu D, Bertsch C Mestre P** (2
792 014) The SWEET family of sugar transporters in grapevine: VvSWE
793 ET4 is involved in the interaction with Botrytis cinerea. J Exp Bot
794 **65**: 6589-6601.
795 **Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand**
796 **NC, Shamim MS et al** (2017) De novo assembly of the Aedes ae
797 gypti genome using Hi-C yields chromosome-length scaffolds. Scienc
798 e **356**: 92-95.
799 **Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander**
800 **ES Aiden EL** (2016) Juicer Provides a One-Click System for Analy
801 zing Loop-Resolution Hi-C Experiments. Cell Syst **3**: 95-98.
802 **Edgar RC, Myers EW** (2005) PILER: identification and classification of
803 genomic repeats. Bioinformatics **21 Suppl 1**: i152-158.

Emms DM Kelly S (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Bio **16**: 157.

Emms DM Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol **20**: 238.

Foria S, Magris G, Jurman I, Schwope R, Candido MD, Luca ED, Ivanišević D et al (2022) Extent of wild–to–crop interspecific introgression in grapevine (*vitis vinifera*) as a consequence of resistance breeding and implications for the crop species definition. Hortic Res **9**: uhab010.

Gao Y, Yang X, Yang X, Zhao T, An X Chen Z (2021) Characterization and expression pattern of the trehalose-6-phosphate synthase and trehalose-6-phosphate phosphatase gene families in Populus. Int J Biol Macromol **187**: 9-23.

Girollet N, Rubio B, Lopez-Roques C, valière S, Ollat N, Bert PF (2019) *De novo* phased assembly of the *Vitis riparia* grape genome. Sci Data **19**: 127.

Gonzalez I, Dejean S, Martin P Baccini A (2008) CCA: An R Package to Extend CanonicalCorrelation Analysis. J Stat Softw **23**: 1-14.

Guo X, Fang D, Sahu SK, Yang S, Guang X, Folk R, Smith SA et al (2021) Chloranthus genome provides insights into the early diversification of angiosperms. Nature Communications **12**: 6930.

Hardenack S, Ye D, Saedler H, Grant S (1994) Comparison of MADS box gene expression in developing male and female flowers of the dioecious plant white campion. Plant Cell **6**:1775-1787.

Hu J, Wang K, Huang W, Liu G, Gao Y, Wang J, Huang Q et al (2012) The rice pentatricopeptide repeat protein RF5 restores fertility in Hong-Lian cytoplasmic male-sterile lines via a complex with the glycine-rich protein GRP162. Plant Cell **24**:109-22.

Iocco-Corena P, Chaïb J, Torregrosa L, Mackenzie D, Thomas MR, Smith HM (2021) VviPLATZ1 is a major factor that controls female flower morphology determination in grapevine. Nat Commun. **12**:6995.

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449**: 463-468.

Kataya ARA, Elshobaky A, Heidari B, Dugassa NF, Thelen JJ Lillo C (2020) Multi-targeted trehalose-6-phosphate phosphatase I harbors a novel peroxisomal targeting signal 1 and is essential for flowering and development. Planta **251**: 98.

Kazama T, Toriyama K (2003) A pentatricopeptide repeat-containing gene that promotes the processing of aberrant atp6 RNA of cytoplasmic male-sterile rice. FEBS Lett. **544**: 99-102.

Kim MS, Hur YY, Ji HK, Jeong SC (2020) Genome resequencing, improvement of variant calling, and population genomic analyses provide insights into the seedlessness in the genus *vitis*. G3-Genes Genomes Genetics **10**: 3365-3377.

Koizuka N, Imai R, Fujimoto H, Hayakawa T, Kimura Y, Kohno-Murase J, Sakai T et al (2003) Genetic characterization of a pentatricopeptide repeat protein gene, orf687, that restores fertility in the cytoplasmic male-sterile Kosena radish. Plant J. **34**:407–415.

Lee TH, Guo H, Wang X, Kim C Paterson AH (2014) SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genomics **15**: 162–7.

Liang Z, Duan S, Sheng J, Zhu S, Ni X, Shao J, Liu C et al (2019) Whole-genome resequencing of 472 *Vitis* accessions for grapevine diversity and demographic history analyses. Nat Commun **10*:*** 1190.

Lin P, Wang K, Wang Y, Hu Z, Yan C, Huang H, Ma X et al (2022) The genome of oil-Camellia and population genomics analysis provide insights into seed oil domestication. Genome Biol **23**: 14.

Lowe TM Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res**25**: 955-964.

Ma W, Xu L, Gao S, Lyu X, Cao X Yao Y (2021) Melatonin alters the secondary metabolite profile of grape berry skin by promoting *Vv*MYB14-mediated ethylene biosynthesis. Hortic Res **8**: 43.

Ma ZY, Wen J, Tian JP, Jamal A, Chen LQ Liu XQ (2018) Testing reticulate evolution of four vitis species from East Asia using restriction-site associated DNA sequencing. J Syst Evol **56**: 331-339.

Marcais G Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics **27**: 764-7 70.

Massonnet M, Cochetel N, Minio A, Vondras AM, Lin J, Muyle A, Garcia JF et al (2020) The genetic basis of sex determination in grapes. Nat Commun **11**: 2902.

Men Y, Li JR, Shen HL, Yang YM, Fan ST, Li K, Guo YS et al (2021) VaAPRT3 Gene is Associated With Sex Determination in Vitis amurensis. Front Genet **12**: 727260.

Pfeifer B, Wittelsbürger U, Onsins SER Lercher MJ (2014) PopGeno me: an efficient Swiss army knife for population genomic analyses in R. Mol Biol Evol **31**: 1929–36.

Qiu L, Wei XY, Wang SJ Wang JJ (2020) Characterization of trehalos e-6-phosphate phosphatase in trehalose biosynthesis, asexual develop ment, stress resistance and virulence of an insect mycopathogen. Pes tic Biochem Physiol **163**: 185-192.

887 **Ramos MJ, Coito JL, Silva HG, Cunha J, Costa MM, Rocheta M** (2
888 014) Flower development and sex specification in wild grapevine. B
889 MC Genomics. **5**:1095.
890 **Satoh-Nagasawa N, Nagasawa N, Malcomber S, Sakai H Jackson D**
891 (2006) A trehalose metabolic enzyme controls inflorescence architect
892 ure in maize. Nature **441**: 227-230.
893 **Song XM, Wang JP, Sun PC, Ma X, Yang QH, Hu JJ, Sun SR et al** (2020)
894 Preferential gene retention increases the robustness of cold regulation in
895 *Brassicaceae* and other plants after polyploidization. Hortic Res **7**: 20.
896 **Sun P, Jiao B, Yang Y, Shan L, Li T, Li XN, Xi ZX et al** (2021) WGDI: A
897 user-friendly toolkit for evolutionary analyses of whole-genome
898 duplications and ancestral karyotypes.
899 https://doi.org/10.1101/2021.04.29.441969
900 **Tel-Zur N, Abbo S, Myslabodski D Mizrahi Y** (1999) Modified CTAB
901 procedure for DNA isolation from epiphytic cacti of the genera Hy
902 locereus and Selenicereus (Cactaceae). Plant Mol Biol Rep **17**: 249–
903 254.
904 **Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, G**
905 **urtowski J Schatz MC** (2017) GenomeScope: fast reference-free ge
906 nome profiling from short reads. Bioinformatics **33**: 2202-2204.
907 **Waki T, Mameda R, Nakano T, Yamada S, Terashita M, Ito K, Ten**
908 **ma N et al** (2020) A conserved strategy of chalcone isomerase-like
909 protein to rectify promiscuous chalcone synthase specificity. Nat Co
910 mmun **11**: 870.
911 **Wan Y, Schwaninger H, Li D, Simon CJ, Wang Y He P** (2008) The
912 eco-geographic distribution of wild grape germplasm in China. *Vitis* **47**:
913 77-80.

914 **Wang JP, Yu JG, Li J, Sun PC, Wang L, Yuan JQ, Meng FB et al** (2018a)
915 Two likely auto-tetraploidization events shaped kiwifruit genome and
916 contributed to establishment of the actinidiaceae family. *iScience* **7**:
917 230-240.
918 **Wang P, Su L, Gao H, Jiang X, Wu X, Li Y, Zhang Q et al** (2018b)
919 Genome-Wide Characterization of bHLH Genes in Grape and Anal
920 ysis of their Potential Relevance to Abiotic Stress Tolerance and Sec
921 ondary Metabolite Biosynthesis. Front Plant Sci **9**: 64.
922 **Wang P, Yang Y, Shi H, Wang Y Ren F** (2019) Small RNA and degradome
923 deep sequencing reveal respective roles of cold-related micrornas across
924 Chinese wild grapevine and cultivated grapevine. *BMC* Genomics **20**: 740.
925 **Wang X, Tang H, Paterson AH** (2011) Seventy million years of concerted
926 evolution of a homoeologous chromosome pair, in parallel, in major
927 Poaceae lineages. Plant Cell **23**:27-37.
928 **Wang Y, Xin H, Fan P, Zhang J, Liu Y, Dong Y, Wang Z et al** (2021) The
929 genome of shanputao (*Vitis amurensis*) provides a new insight into cold
930 tolerance of grapevine. Plant J **105**: 1495-1506.
931 **Xu W, Li R, Zhang N, Ma F, Jiao Y Wang Z** (2014) Transcriptome profiling
932 of *Vitis amurensis*, an extremely cold-tolerant Chinese wild *Vitis* species,
933 reveals candidate genes and events that potentially connected to cold stress.
934 Plant Mol Biol **86**: 527-541.
935 **Yamazaki M, Ishida A, Suzuki Y, Aoki Y, Suzuki S Enoki S** (2021)
936 Ethylene Induced by Sound Stimulation Enhances Anthocyanin Accu
937 mulation in Grape Berry Skin through Direct Upregulation of UDP-
938 Glucose: Flavonoid 3-O-Glucosyltransferase. Cell **10**.
939 **Yang J, Lee SH, Goddard ME Visscher PM** (2011) GCTA: a tool for
940 genome-wide complex trait analysis. Am J Hum Genet **88**: 76–82.
941 **Yang Y, Sun P, Lv L, Wang D, Ru D, Li Y, Ma T et al** (2020) Pric

kly waterlily and rigid hornwort genomes shed light on early angios perm evolution. Nat Plants **6**: 215-222.

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586-91.

Yu Y, Guan J, Xu Y, Ren F, Zhang Z, Yan J, Fu J et al (2021) Population-scale peach genome analyses unravel selection patterns and biochemical basis underlying fruit flavor. Nat Commun **12**: 3604.

Zhang Q, Wang L, Wang Z, Zhang R, Liu P, Liu M, Liu Z et al (2021) The regulation of cell wall lignification and lignin biosynthesis during pigmentation of winter jujube. Hortic Res **8**: 238.

Zhao X Hao W (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res **35**: W265-8.

Tables

Table 1 Summary statistics for the *Va* Zuoshan No.1 genome assembly in comparison with the *Va* IBCAS1988 genome.

Genomic feature	Zuoshan No.1	IBCAS1988
Estimated genome size	532.35Mb	607Mb
Total assembly size	522.28Mb	604.56Mb
Number of contigs	615	—
Largest contig	14,947,757bp	2,623,866bp
Contig N50 length	2,512,611bp	282,256bp
Scaffold N50	26,519,999bp	748,673bp
Sequences anchored to chromosomes	513.47 Mb/98.31%	—
GC content	34.50%	34.37%
Complete BUSCOs	97.50%	—
Repeat(%)	59.21%	47.06%
Protein-coding genes	27,635	32,885
Heterozygous ratio(%)	1.20%	1.23%

Table 2 The assessment of interaction networks constructed using candidate CRGs in *Va*, *Vr* and *Vv*

	All nodes in a network	5	10	5%	10%	20%
Va	84	0.851	0.807	0.861	0.826	0.754
Vr	91	0.832	0.793	0.839	0.801	0.728
Vv	85	0.816	0.782	0.823	0.797	0.731

Network robustness was calculated after the removal of certain numbers (5 or 10) or percentages (5%, 10%, or 20%) of nodes.

Table 3 Summary of the RNA-seq

sample	clean_reads	mapping rate
S1_1	42378612	95.52%
S1_2	44413474	96.27%
S1_3	42453694	95.60%
S2_1	43634464	95.63%
S2_2	39853380	94.96%
S2_3	40757404	95.10%
S3_1	43282574	93.71%
S3_2	44118914	94.34%
S3_3	43169656	94.15%
S4_1	44205082	92.82%
S4_2	42005090	92.90%
S4_3	44053722	93.42%

Table 4 The number of DEGs

group	all	up	down
S2vsS1	3307	1367	1940
S3vsS1	4633	1862	2771
S4vsS1	5969	2209	3760
S3vsS2	2904	1196	1708
S4vsS2	4817	1904	2913
S4vsS3	1797	652	1145

Figure legends

Figure 1 Amur grape Atlas

(A) Spider silk-like hairs of Amur grape. (B) The bisexual flower. (C) The female flower. (D) The male flower. (E) Amur grape cv. Zuoshan No.1. (F) The leaf of Zuoshan No.1. (G) The flower of Zuoshan No.1.

Figure 2 Genomic feature of *Va* Zuoshan No.1 and comparison with other genomes.

(A) The landscape of genome assembly and annotation of *Va* Zuoshan No.1. Tracks from outside to the inner correspond to I, chromosomes; II, gene density; III, repeat density IV, Non-coding RNA density; V, GC content; VI-IX, Gene expression levels (FPKM) in S4-S1; and X, synteny information.

(B) Hi-C map of *Va* Zuoshan No.1.

(C) Dotplot of homologous genes between *Va* Zuoshan No.1 and *Vv* PN40024.

(D) Dotplot of homologous genes between *Va* Zuoshan No.1 and *Vv* Cabernet Sauvignon.

(E) Dotplot of homologous genes between *Vv* Cabernet Sauvignon and *Vr*.

Figure 3 Ks distribution of collinear genes and phylogenetic relationship of species.

(A) Ks distribution of homologous genes within a species.

(B) Ks distribution of homologous genes between any two species.

(C) Evolutionary tree of selected plants. The number is the step size and represents the size of the difference between two adjacent sequences.

Figure 4 Structural Variation analysis

(A) A comparison between *Va* and *Vv*, with *Va* as the reference. (B) A comparison between *Va* and *Vr*, with *Va* as the reference. (C) A comparison between *Vr* and *Vv*, with *Vr* as the reference.

Figure 5 Multiomics analysis of *Vitis amurensis* fruit at different stages

(A) Fruits at stages S1, S2, S3 and S4. (B) CCA analysis of key genes and metabolites in resveratrol metabolic pathway. (C) Schematic diagram of resveratrol synthesis and metabolism mechanism during fruit ripening of *vitis amurensis*. "*" represents that these genes expression or metabolites content of the sample changes significantly compared with S1($P<0.05$). CYP73A is the symbol of trans-cinnamate 4-monooxygenase, ST is the symbol of stilbene synthase, HTC is the symbol of shikimate O-hydroxycinnamoyltransferase, ROMT is the symbol of trans-resveratrol di-O-methyltransferase and CYP98A is the symbol o of 5-O-(4-coumaroyl)-D-quinate 3'-monooxygenase.

Figure 6 Population structure analysis and selective sweep analysis of 24 *Va* individuals.

(A) Population structure of 24 *Vitis amurensis* individuals (There are some individuals whose names use abbreviations here. Please refer to table s14 for the corresponding relationship between abbreviations and full names of individual names). (B) An evolutionary tree based on the whole-genome SNPs of 24 *Va* individuals. (C) An evolutionary tree based on the SNPs on the SDRs of 24 *Va* individuals. (D) A Manhattan diagram of the selective region. The top subfigure shows the selective region from the comparison of female and male groups comparison. The middle subfigure is the selective region from the comparison of female and hermaphroditic groups. The bottom subfigure shows the selective region from the comparison of male and hermaphroditic groups. The red dashed line represents the threshold of θπ ratios and FST (top 5%). The blue asterisk represents the selective region which harbored SDR genes.

Figure 7 Genes of *Vv* and *Va* SDRs

Genes *of Vv* H, F hap *SDRs* are from *Vv* Cabernet Sauvignon. Genes of *Vv* M hap SDR are from *Vv. sylvestris*. Genes of Va F hap are from Zuoshan No.1. The gene with a arrowhead is the *Va* Uncharacterized protein gene.

Supplementary tables

Table S1 Statistics of gene annotations for *Va* Zuoshan No.1 genome.

Table S2 Number of Homologous Genes Residing in Blocks within each Genome or between Genomes.

Table S3 Kernel Function Analysis of K_S Distribution Related to Duplication Events Within each Genome and Between Genomes.

Table S4 Statistics of PAV regions between *Va* genome and *Vv* genome or *Va* genome and *Vr* genome.

Table S5 *Va*-specific PAV genes between *Va* genome and *Vv* genome.

Table S6 *Vv*-specific PAV genes between *Va* genome and *Vv* genome.

Table S7 *Va*-specific PAV genes between *Va* genome and *Vr* genome.

Table S8 *Vr*-specific PAV genes between *Va* genome and *Vr* genome.

Table S9 Information of 24 *Vitis amurensis* individuals.

Supplementary figure legends

Figure S1. Dotplot of homologous genes between *Va* Zuoshan No.1 and *Va* IBCAS1988 genomes

Figure S2. The evolutionary tree based on the 3 grapes and Arabidopsis or Boxwood.

Figure S3. Structural Variation analysis plot based on ggplot2.

Figure S4. Tree of NBS family genes of the three species.

Figure S5. The distribution of NBS family genes on the chromosomes of the three species. Lines represent Ks values of NBS gene pairs that are less than 0.025 (red), larger than 0.025 but less than 1 (orange), and larger than 1 but less than 1.5 (grey).

Figure S6. The total sugar and titratable acid content of *Vitis amurensis* fruit at different stages.

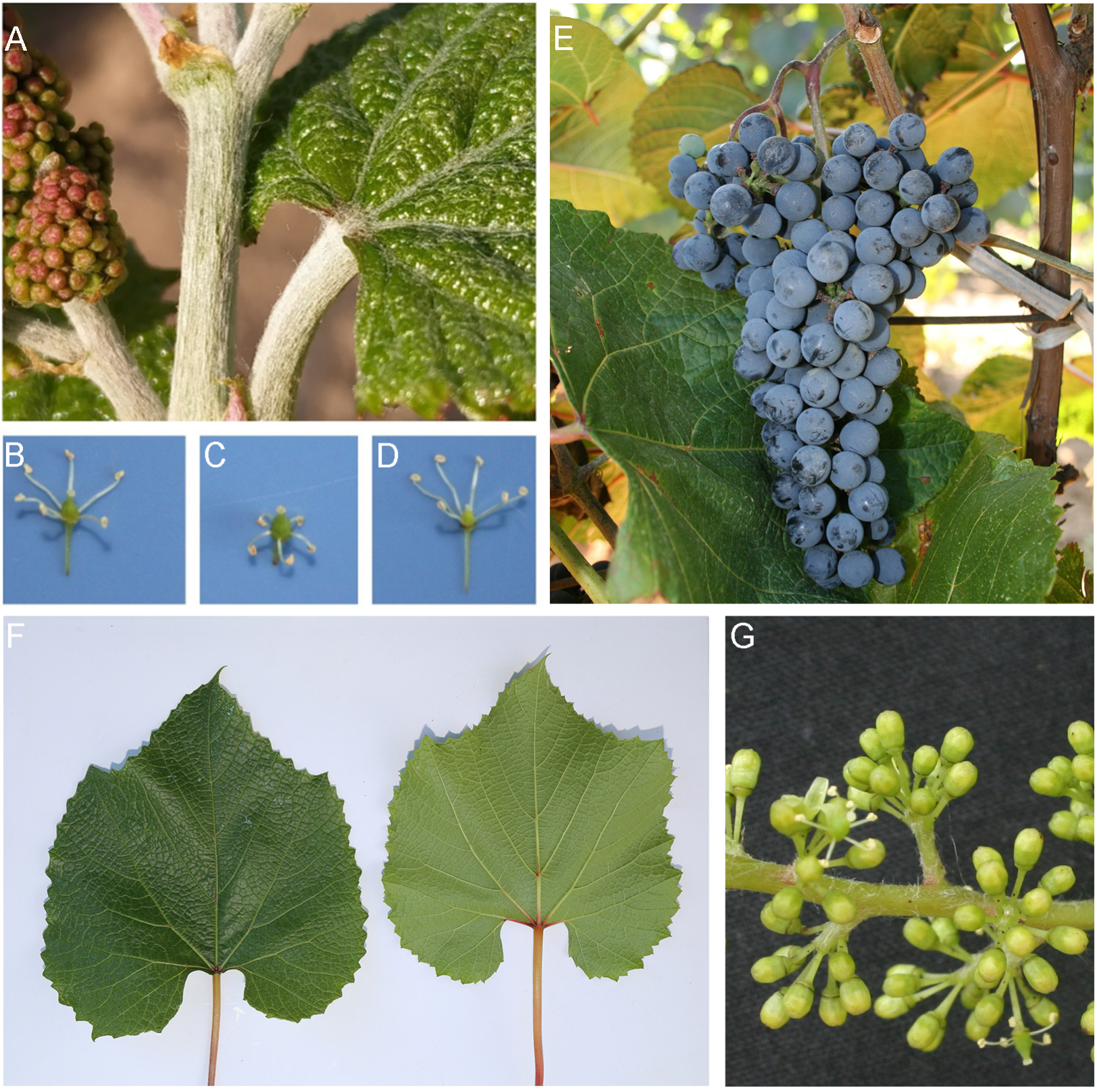

Figure 1 Amur grape Atlas.

(A) Spider silk-like hairs of Amur grape. (B) The bisexual flower. (C) The female flower. (D) The male flower. (E) Amur grape cv. Zuoshan No.1. (F) The leaf of Zuoshan No.1. (G) The flower of Zuoshan No.1.

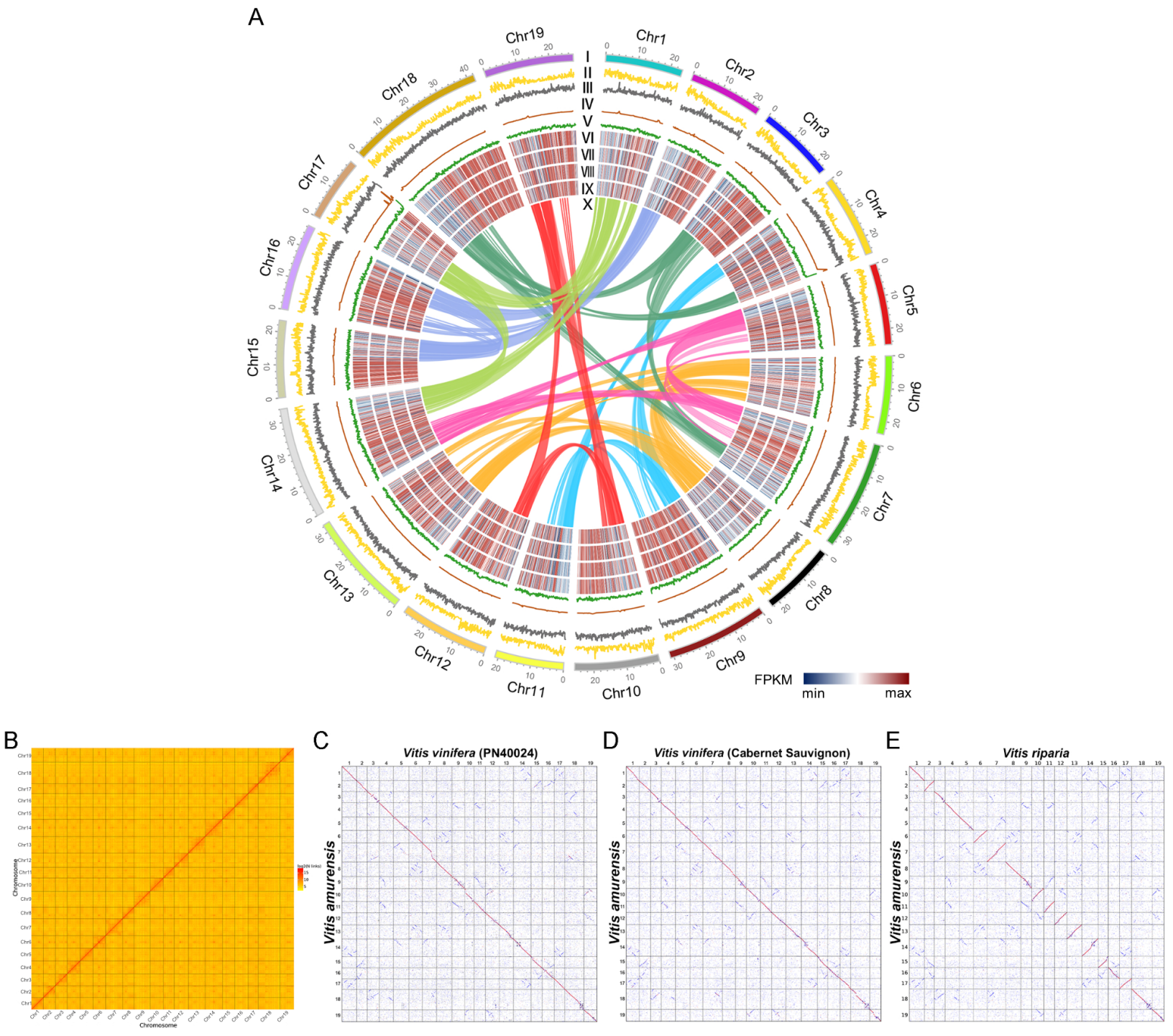

Figure 2 Genomic feature of *Va* Zuoshan No.1 and comparison with other genomes.
(A) The landscape of genome assembly and annotation of *Va* Zuoshan No.1. Tracks from outside to the inner correspond to I, chromosomes; II, gene density; III, repeat density IV, Non-coding RNA density; V, GC content; VI-IX, Gene expression levels (FPKM) in S4-S1; and X, synteny information. (B) Hi-C map of *Va* Zuoshan No.1. (C) Dotplot of homologous genes between *Va* Zuoshan No.1 and *Vv* PN40024. (D) Dotplot of homologous genes between *Va* Zuoshan No.1 and *Vv* Cabernet Sauvignon. (E) Dotplot of homologous genes between *Vv* Cabernet Sauvignon and *Vr*.

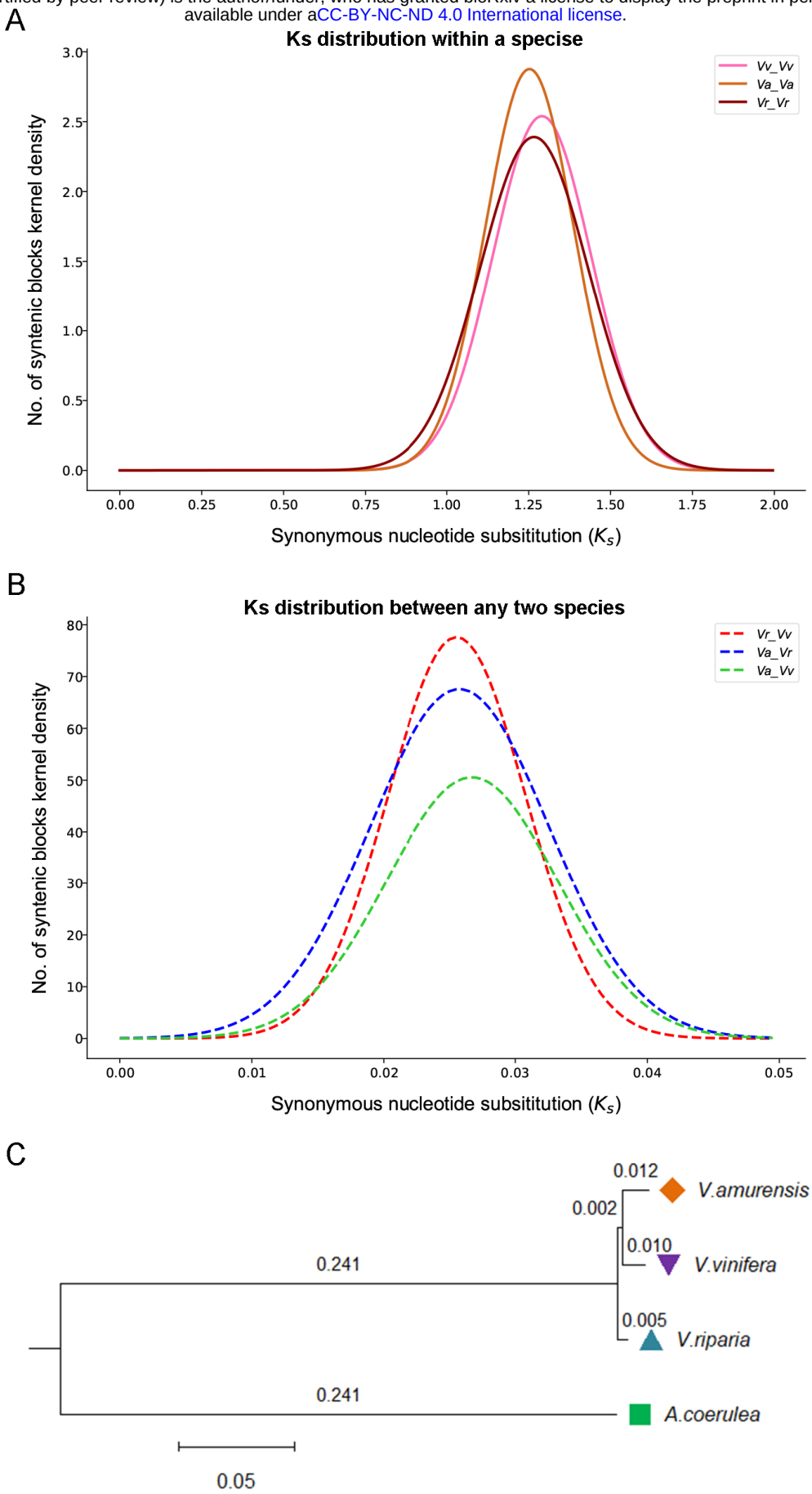

Figure 3 *Ks* distribution of collinear genes and phylogenetic relationship of species.
(A) *Ks* distribution of homologous genes within a species. (B) *Ks* distribution of homologous genes between any two species. (C) Evolutionary tree of selected plants.

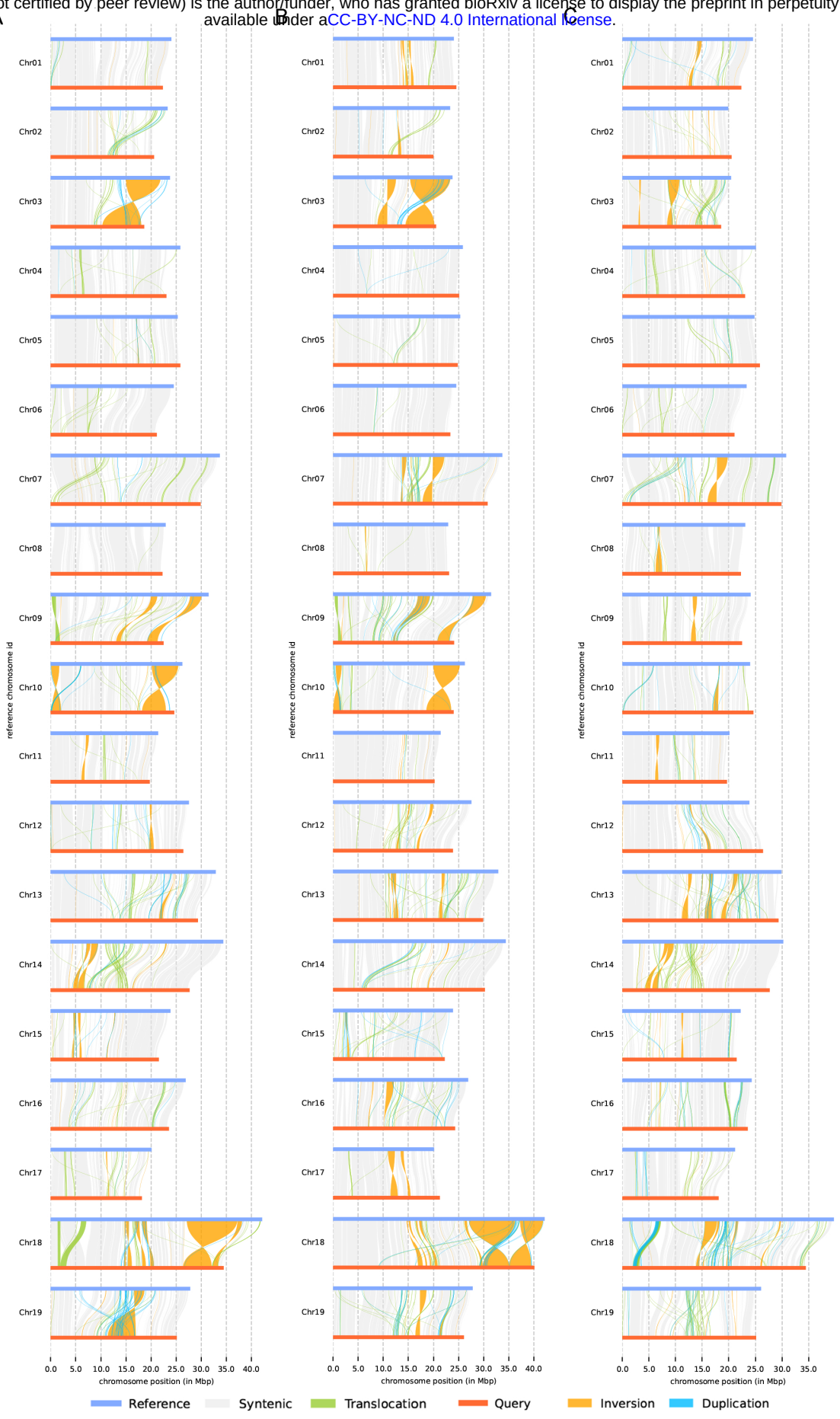

Figure 4 Structural Variation analysis.

(A) A comparison between *Va* and *Vv*, with *Va* as the reference. (B) A comparison between *Va* and *Vr*, with *Va* as the reference. (C) A comparison between *Vr* and *Vv*, with *Vr* as the reference.

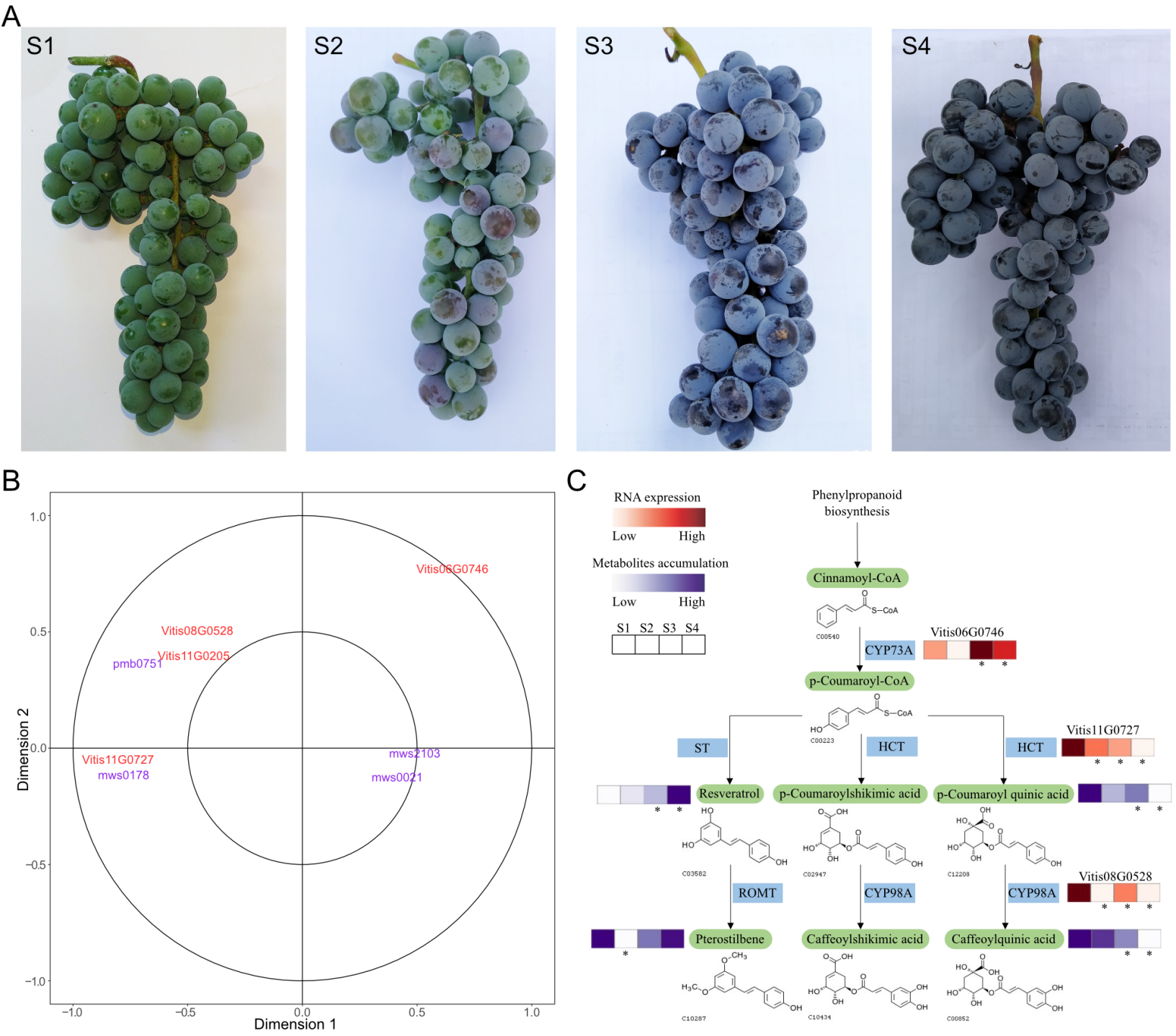

Figure 5 Multiomics analysis of *Vitis amurensis* fruit at different stages.
(A) Fruits at stages S1, S2, S3 and S4. (B) CCA analysis of key genes and metabolites in resveratrol metabolic pathway. (C) Schematic diagram of resveratrol synthesis and metabolism mechanism during fruit ripening of vitis amurensis. "*" represents that these genes expression or metabolites content of the sample changes significantly compared with S1(P<0.05). CYP73A is the symbol of trans-cinnamate 4-monooxygenase, ST is the symbol of stilbene synthase, HTC is the symbol of shikimate O-hydroxycinnamoyltransferase, ROMT is the symbol of trans-resveratrol di-O-methyltransferase and CYP98A is the symbol o of 5-O-(4-coumaroyl)-D-quinate 3'-monooxygenase.

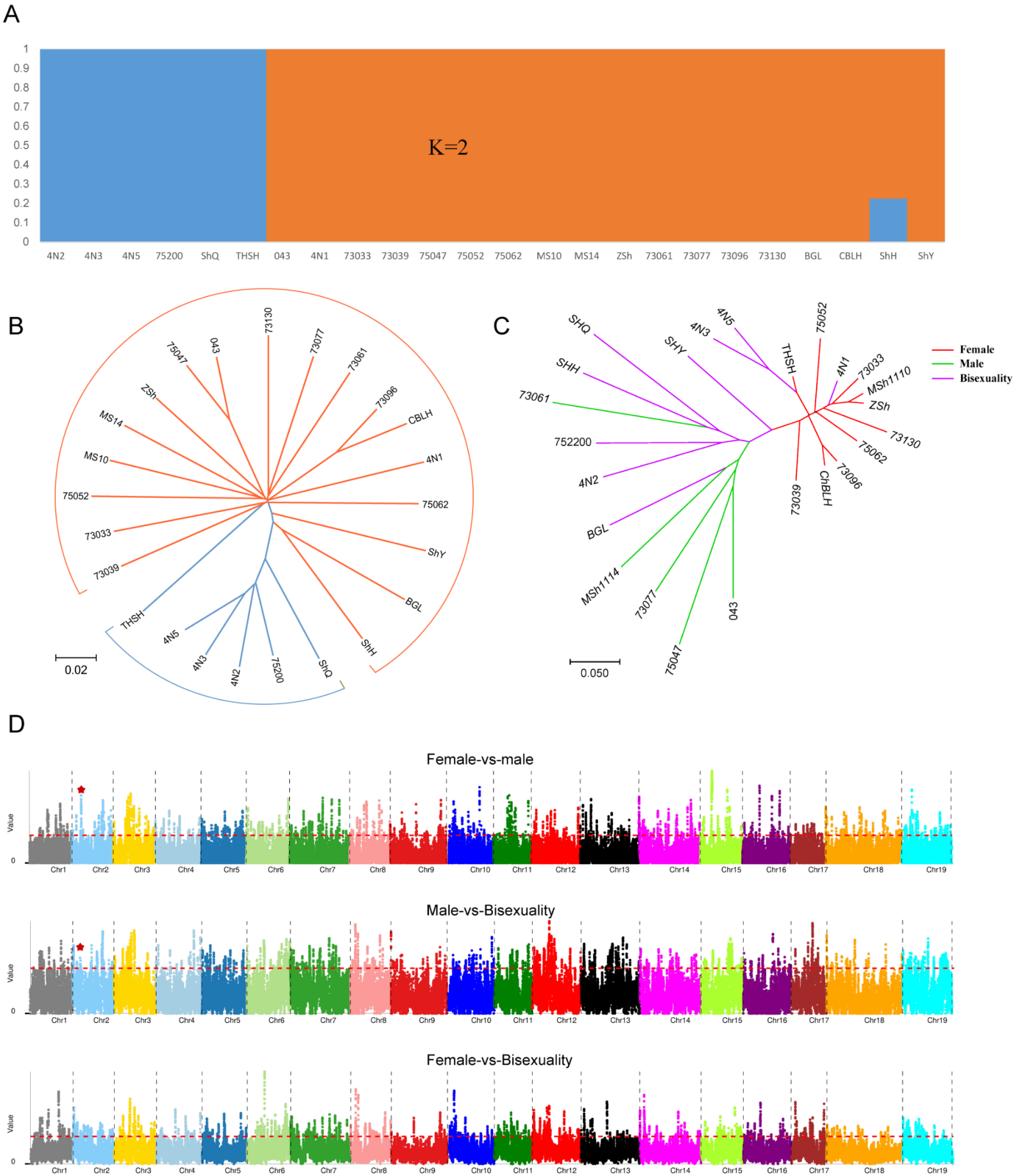

Figure 6 Population structure analysis and selective sweep analysis of 24 *Va* individuals.
(A) Population structure of 24 *Vitis amurensis* individuals (There are some individuals whose names use abbreviations here. Please refer to table s14 for the corresponding relationship between abbreviations and full names of individual names). (B) An evolutionary tree based on the whole-genome SNPs of 24 *Va* individuals. (C) An evolutionary tree based on the SNPs on the SDRs of 24 *Va* individuals. (D) A Manhattan diagram of the selective region. The top subfigure shows the selective region from the comparison of female and male groups comparison. The middle subfigure is the selective region from the comparison of female and hermaphroditic groups. The bottom subfigure shows the selective region from the comparison of male and hermaphroditic groups. The red dashed line represents the threshold of θπ ratios and FST (top 5%). The blue asterisk represents the selective region which harbored SDR genes.

Parsed Citations

Ainsworth C, Crossley S, Buchanan-Wollaston V, Thangavelu M, Parker J (1995) Male and female flowers of the dioecious plant sorrel show different patterns of MADS box gene expression. Plant Cell 7:1583-1598.
Google Scholar: Author Only Title Only Author and Title

Badouin H, Velt A, Gindraud F, Flutre T, Dumas V, Vautrin S, Marande W et al (2020) The wild grape genome sequence provides insights into the transition from dioecy to hermaphroditism during grape domestication. Genome Biol 21:223.
Google Scholar: Author Only Title Only Author and Title

Benson, G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27: 573-580.
Google Scholar: Author Only Title Only Author and Title

Bentolila S, Alfonso AA, Hanson MR. A (2002) pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. Proc Natl Acad Sci U S A 99:10887-10892.
Google Scholar: Author Only Title Only Author and Title

Brigida G, Jan S, Troels P, Leahm S, Veerle S, Beatriz, HM, Adriaan M, et al(2016) Domestication and divergence of Saccharomyces cerevisiae beer yeasts. Cell 166: 1397–1410.
Google Scholar: Author Only Title Only Author and Title

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC bioinformatics 10: 421.
Google Scholar: Author Only Title Only Author and Title

Canaguier A, Grimplet J, Gaspero GD, Scalabrin S, Duchêne E, Choisne N, Mohellibi N et al (2017) A new version of the grapevine reference genome assembly (12x.v2) and of its annotation (VCost.v3). Genomics Data 14: 56-62.
Google Scholar: Author Only Title Only Author and Title

Chawla HS, Lee H, Gabur I, Vollrath P, Tamilselvan-Nattar-Amutha S, Obermeier C, Schiessl SV et al (2021) Long-read sequencing reveals widespread intragenic structural variants in a recent allopolyploid crop plant. Plant Biotechnol J 19: 240-250.
Google Scholar: Author Only Title Only Author and Title

Chen Q, Diao L, Song H, Zhu X (2018) Vitis amurensis Rupr: A review of chemistry and pharmacology. Phytomedicine 49: 111-122.
Google Scholar: Author Only Title Only Author and Title

Chong J, Piron MC, Meyer S, Merdinoglu D, Bertsch C Mestre P (2014) The SWEET family of sugar transporters in grapevine: VvSWEET4 is involved in the interaction with Botrytis cinerea. J Exp Bot 65: 6589-6601.
Google Scholar: Author Only Title Only Author and Title

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS et al (2017) De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science 356: 92-95.
Google Scholar: Author Only Title Only Author and Title

Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES Aiden EL (2016) Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst 3: 95-98.
Google Scholar: Author Only Title Only Author and Title

Edgar RC, Myers EW (2005) PILER: identification and classification of genomic repeats. Bioinformatics 21 Suppl 1: i152-158.
Google Scholar: Author Only Title Only Author and Title

Emms DM Kelly S (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Bio 16: 157.
Google Scholar: Author Only Title Only Author and Title

Emms DM Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20: 238.
Google Scholar: Author Only Title Only Author and Title

Foria S, Magris G, Jurman I, Schwope R, Candido MD, Luca ED, Ivanišević D et al (2022) Extent of wild–to–crop interspecific introgression in grapevine (vitis vinifera) as a consequence of resistance breeding and implications for the crop species definition. Hortic Res 9: uhab010.
Google Scholar: Author Only Title Only Author and Title

Gao Y, Yang X, Yang X, Zhao T, An X Chen Z (2021) Characterization and expression pattern of the trehalose-6-phosphate synthase and trehalose-6-phosphate phosphatase gene families in Populus. Int J Biol Macromol 187: 9-23.
Google Scholar: Author Only Title Only Author and Title

Girollet N, Rubio B, Lopez-Roques C, valière S, Ollat N, Bert PF (2019) De novo phased assembly of the Vitis riparia grape genome. Sci Data 19: 127.
Google Scholar: Author Only Title Only Author and Title

Gonzalez I, Dejean S, Martin P Baccini A (2008) CCA: An R Package to Extend CanonicalCorrelation Analysis. J Stat Softw 23: 1-14.
Google Scholar: Author Only Title Only Author and Title

Guo X, Fang D, Sahu SK, Yang S, Guang X, Folk R, Smith SA et al (2021) Chloranthus genome provides insights into the early diversification of angiosperms. Nature Communications 12: 6930.
Google Scholar: Author Only Title Only Author and Title

Hardenack S, Ye D, Saedler H, Grant S (1994) Comparison of MADS box gene expression in developing male and female flowers of the dioecious plant white campion. Plant Cell 6:1775-1787.
Google Scholar: Author Only Title Only Author and Title

Hu J, Wang K, Huang W, Liu G, Gao Y, Wang J, Huang Q et al (2012) The rice pentatricopeptide repeat protein RF5 restores fertility in Hong-Lian cytoplasmic male-sterile lines via a complex with the glycine-rich protein GRP162. Plant Cell 24:109-22.
Google Scholar: Author Only Title Only Author and Title

Iocco-Corena P, Chaïb J, Torregrosa L, Mackenzie D, Thomas MR, Smith HM (2021) VviPLATZ1 is a major factor that controls female flower morphology determination in grapevine. Nat Commun. 12:6995.
Google Scholar: Author Only Title Only Author and Title

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449: 463-468.
Google Scholar: Author Only Title Only Author and Title

Kataya ARA, Elshobaky A, Heidari B, Dugassa NF, Thelen JJ Lillo C (2020) Multi-targeted trehalose-6-phosphate phosphatase I harbors a novel peroxisomal targeting signal 1 and is essential for flowering and development. Planta 251: 98.
Google Scholar: Author Only Title Only Author and Title

Kazama T, Toriyama K (2003) A pentatricopeptide repeat-containing gene that promotes the processing of aberrant atp6 RNA of cytoplasmic male-sterile rice. FEBS Lett. 544: 99-102.
Google Scholar: Author Only Title Only Author and Title

Kim MS, Hur YY, Ji HK, Jeong SC (2020) Genome resequencing, improvement of variant calling, and population genomic analyses provide insights into the seedlessness in the genus vitis. G3-Genes Genomes Genetics 10: 3365-3377.
Google Scholar: Author Only Title Only Author and Title

Koizuka N, Imai R, Fujimoto H, Hayakawa T, Kimura Y, Kohno-Murase J, Sakai T et al (2003) Genetic characterization of a pentatricopeptide repeat protein gene, orf687, that restores fertility in the cytoplasmic male-sterile Kosena radish. Plant J. 34:407–415.
Google Scholar: Author Only Title Only Author and Title

Lee TH, Guo H, Wang X, Kim C Paterson AH (2014) SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genomics 15: 162–7.
Google Scholar: Author Only Title Only Author and Title

Liang Z, Duan S, Sheng J, Zhu S, Ni X, Shao J, Liu C et al (2019) Whole-genome resequencing of 472 Vitis accessions for grapevine diversity and demographic history analyses. Nat Commun 10: 1190.
Google Scholar: Author Only Title Only Author and Title

Lin P, Wang K, Wang Y, Hu Z, Yan C, Huang H, Ma X et al (2022) The genome of oil-Camellia and population genomics analysis provide insights into seed oil domestication. Genome Biol 23: 14.
Google Scholar: Author Only Title Only Author and Title

Lowe TM Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res25: 955-964.
Google Scholar: Author Only Title Only Author and Title

Ma W, Xu L, Gao S, Lyu X, Cao X Yao Y (2021) Melatonin alters the secondary metabolite profile of grape berry skin by promoting VvMYB14-mediated ethylene biosynthesis. Hortic Res 8: 43.
Google Scholar: Author Only Title Only Author and Title

Ma ZY, Wen J, Tian JP, Jamal A, Chen LQ Liu XQ (2018) Testing reticulate evolution of four vitis species from East Asia using restriction-site associated DNA sequencing. J Syst Evol 56: 331-339.
Google Scholar: Author Only Title Only Author and Title

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

Massonnet M, Cochetel N, Minio A, Vondras AM, Lin J, Muyle A, Garcia JF et al (2020) The genetic basis of sex determination in grapes. Nat Commun 11: 2902.

Google Scholar: Author Only Title Only Author and Title

Men Y, Li JR, Shen HL, Yang YM, Fan ST, Li K, Guo YS et al (2021) VaAPRT3 Gene is Associated With Sex Determination in Vitis amurensis. Front Genet 12: 727260.
Google Scholar: Author Only Title Only Author and Title

Pfeifer B, Wittelsbürger U, Onsins SER Lercher MJ (2014) PopGenome: an efficient Swiss army knife for population genomic analyses in R. Mol Biol Evol 31: 1929–36.
Google Scholar: Author Only Title Only Author and Title

Qiu L, Wei XY, Wang SJ Wang JJ (2020) Characterization of trehalose-6-phosphate phosphatase in trehalose biosynthesis, asexual development, stress resistance and virulence of an insect mycopathogen. Pestic Biochem Physiol 163: 185-192.
Google Scholar: Author Only Title Only Author and Title

Ramos MJ, Coito JL, Silva HG, Cunha J, Costa MM, Rocheta M (2014) Flower development and sex specification in wild grapevine. BMC Genomics. 5:1095.
Google Scholar: Author Only Title Only Author and Title

Satoh-Nagasawa N, Nagasawa N, Malcomber S, Sakai H Jackson D (2006) A trehalose metabolic enzyme controls inflorescence architecture in maize. Nature 441: 227-230.
Google Scholar: Author Only Title Only Author and Title

Song XM, Wang JP, Sun PC, Ma X, Yang QH, Hu JJ, Sun SR et al (2020) Preferential gene retention increases the robustness of cold regulation in Brassicaceae and other plants after polyploidization. Hortic Res 7: 20.
Google Scholar: Author Only Title Only Author and Title

Sun P, Jiao B, Yang Y, Shan L, Li T, Li XN, Xi ZX et al (2021) WGDI: A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. https://doi.org/10.1101/2021.04.29.441969
Google Scholar: Author Only Title Only Author and Title

Tel-Zur N, Abbo S, Myslabodski D Mizrahi Y (1999) Modified CTAB procedure for DNA isolation from epiphytic cacti of the genera Hylocereus and Selenicereus (Cactaceae). Plant Mol Biol Rep 17: 249–254.
Google Scholar: Author Only Title Only Author and Title

Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J Schatz MC (2017) GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33: 2202-2204.
Google Scholar: Author Only Title Only Author and Title

Waki T, Mameda R, Nakano T, Yamada S, Terashita M, Ito K, Tenma N et al (2020) A conserved strategy of chalcone isomerase-like protein to rectify promiscuous chalcone synthase specificity. Nat Commun 11: 870.
Google Scholar: Author Only Title Only Author and Title

Wan Y, Schwaninger H, Li D, Simon CJ, Wang Y He P (2008) The eco-geographic distribution of wild grape germplasm in China. Vitis 47: 77-80.
Google Scholar: Author Only Title Only Author and Title

Wang JP, Yu JG, Li J, Sun PC, Wang L, Yuan JQ, Meng FB et al (2018a) Two likely auto-tetraploidization events shaped kiwifruit genome and contributed to establishment of the actinidiaceae family. iScience 7: 230-240.
Google Scholar: Author Only Title Only Author and Title

Wang P, Su L, Gao H, Jiang X, Wu X, Li Y, Zhang Q et al (2018b) Genome-Wide Characterization of bHLH Genes in Grape and Analysis of their Potential Relevance to Abiotic Stress Tolerance and Secondary Metabolite Biosynthesis. Front Plant Sci 9: 64.
Google Scholar: Author Only Title Only Author and Title

Wang P, Yang Y, Shi H, Wang Y Ren F (2019) Small RNA and degradome deep sequencing reveal respective roles of cold-related micrornas across Chinese wild grapevine and cultivated grapevine. BMC Genomics 20: 740.
Google Scholar: Author Only Title Only Author and Title

Wang X, Tang H, Paterson AH (2011) Seventy million years of concerted evolution of a homoeologous chromosome pair, in parallel, in major Poaceae lineages. Plant Cell 23:27-37.
Google Scholar: Author Only Title Only Author and Title

Wang Y, Xin H, Fan P, Zhang J, Liu Y, Dong Y, Wang Z et al (2021) The genome of shanputao (Vitis amurensis) provides a new insight into cold tolerance of grapevine. Plant J 105: 1495-1506.
Google Scholar: Author Only Title Only Author and Title

Xu W, Li R, Zhang N, Ma F, Jiao Y Wang Z (2014) Transcriptome profiling of Vitis amurensis, an extremely cold-tolerant Chinese wild Vitis species, reveals candidate genes and events that potentially connected to cold stress. Plant Mol Biol 86: 527-541.
Google Scholar: Author Only Title Only Author and Title

Yamazaki M, Ishida A, Suzuki Y, Aoki Y, Suzuki S Enoki S (2021) Ethylene Induced by Sound Stimulation Enhances Anthocyanin

Accumulation in Grape Berry Skin through Direct Upregulation of UDP-Glucose: Flavonoid 3-O-Glucosyltransferase. Cell 10.
Google Scholar: Author Only Title Only Author and Title

Yang J, Lee SH, Goddard ME Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88: 76–82.
Google Scholar: Author Only Title Only Author and Title

Yang Y, Sun P, Lv L, Wang D, Ru D, Li Y, Ma T et al (2020) Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution. Nat Plants 6: 215-222.
Google Scholar: Author Only Title Only Author and Title

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24: 1586-91.
Google Scholar: Author Only Title Only Author and Title

Yu Y, Guan J, Xu Y, Ren F, Zhang Z, Yan J, Fu J et al (2021) Population-scale peach genome analyses unravel selection patterns and biochemical basis underlying fruit flavor. Nat Commun 12: 3604.
Google Scholar: Author Only Title Only Author and Title

Zhang Q, Wang L, Wang Z, Zhang R, Liu P, Liu M, Liu Z et al (2021) The regulation of cell wall lignification and lignin biosynthesis during pigmentation of winter jujube. Hortic Res 8: 238.
Google Scholar: Author Only Title Only Author and Title

Zhao X Hao W (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res 35: W265-8.
Google Scholar: Author Only Title Only Author and Title