

1 **Phased T2T genome assemblies facilitate the mining of disease-resistance genes**
2 **in *Vitis davidii***

3
4 Yuanyuan Luo^{1,†}, Zhenya Liu^{2,†}, Zhongxin Jin³, Peng Li^{1,4}, Xibei Tan¹, Shuo Cao², Xu
5 Wang², Zhongqi Liu², Xiaoya Shi², Siyang Huang², Liyuan Gu¹, Xiucan Fan¹, Jianfu
6 Jiang¹, Lei Sun¹, Yongfeng Zhou^{2,3}, Chonghuai Liu¹, Xiaodong Xu^{2,*}, Zhiyao Ma^{2,*},
7 Ying Zhang^{1,4,5,*}

8
9 **Affiliations:**

10 ¹Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences,
11 Zhengzhou 450000, China

12 ²National Key Laboratory of Tropical Crop Breeding, Shenzhen Branch, Guangdong
13 Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology,
14 Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at
15 Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China

16 ³National Key Laboratory of Tropical Crop Breeding, Tropical Crops Genetic
17 Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou,
18 China

19 ⁴Chuxiong Yunguo Agriculture Technology Research Institute, Chuxiong, China.

20 ⁵Zhongyuan Research Center, Chinese Academy of Agricultural Sciences, Henan
21 China.

22
23
24
25
26 © The Author(s) 2024. Published by Oxford University Press. This is an Open Access
27 article distributed under the terms of the Creative Commons Attribution License
28 <https://creativecommons.org/licenses/by/4.0/>, which permits unrestricted reuse,
29 distribution, and reproduction in any medium, provided the original work is properly
30 cited.

31 [†]These authors contributed equally to this work.

32 *Correspondence authors:

33 Zhiyao Ma: mazhiyao@caas.cn

34 Xiaodong Xu: xuxiaodong@caas.cn

35 Ying Zhang: zhangying05@caas.cn

36

37

ORIGINAL UNEDITED MANUSCRIPT

- 38 **Running title:** Explore resistance genes in Chinese wild grape
- 39 **Yuanyuan Luo:** National Key Laboratory for Germplasm Innovation and Utilization of
 40 Horticultural Crops, Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural
 41 Sciences, Zhengzhou 450000, China.
- 42 **Zhenya Liu:** National Key Laboratory of Tropical Crop Breeding, Shenzhen Branch, Guangdong
 43 Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of
 44 Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of
 45 Agricultural Sciences, Shenzhen, China.
- 46 **Zhongxin Jin:** National Key Laboratory of Tropical Crop Breeding, Tropical Crops Genetic
 47 Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China.
- 48 **Peng Li:** National Key Laboratory for Germplasm Innovation and Utilization of Horticultural
 49 Crops, Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences,
 50 Zhengzhou 450000, China; Chuxiong Yunguo Agriculture Technology Research Institute,
 51 Chuxiong, China.
- 52 **Xibei Tan:** National Key Laboratory for Germplasm Innovation and Utilization of Horticultural
 53 Crops, Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences,
 54 Zhengzhou 450000, China.
- 55 **Shuo Cao:** National Key Laboratory of Tropical Crop Breeding, Shenzhen Branch, Guangdong
 56 Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of
 57 Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of
 58 Agricultural Sciences, Shenzhen, China.
- 59 **Xu Wang:** National Key Laboratory of Tropical Crop Breeding, Shenzhen Branch, Guangdong
 60 Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of
 61 Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of
 62 Agricultural Sciences, Shenzhen, China.
- 63 **Zhongqi Liu:** National Key Laboratory of Tropical Crop Breeding, Shenzhen Branch, Guangdong
 64 Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of
 65 Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of
 66 Agricultural Sciences, Shenzhen, China.
- 67 **Xiaoya Shi:** National Key Laboratory of Tropical Crop Breeding, Shenzhen Branch, Guangdong
 68 Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of
 69 Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of
 70 Agricultural Sciences, Shenzhen, China.
- 71 **Siyang Huang:** National Key Laboratory of Tropical Crop Breeding, Shenzhen Branch,
 72 Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology,
 73 Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese
 74 Academy of Agricultural Sciences, Shenzhen, China.

Liyuan Gu: National Key Laboratory for Germplasm Innovation and Utilization of Horticultural Crops, Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou 450000, China.

Xiucui Fan: National Key Laboratory for Germplasm Innovation and Utilization of Horticultural Crops, Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou 450000, China.

Jianfu Jiang: National Key Laboratory for Germplasm Innovation and Utilization of Horticultural Crops, Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou 450000, China.

Lei Sun: National Key Laboratory for Germplasm Innovation and Utilization of Horticultural Crops, Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou 450000, China.

Yongfeng Zhou: National Key Laboratory of Tropical Crop Breeding, Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China; National Key Laboratory of Tropical Crop Breeding, Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China.

Chonghuai Liu: National Key Laboratory for Germplasm Innovation and Utilization of Horticultural Crops, Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou 450000, China.

Xiaodong Xu: National Key Laboratory of Tropical Crop Breeding, Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China.

Zhiyao Ma: National Key Laboratory of Tropical Crop Breeding, Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China.

Ying Zhang: Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou 450000, China; Chuxiong Yunguo Agriculture Technology Research Institute, Chuxiong, China.

Abstract

Grape is an important fruit crop, and its production faces significant threat from diseases, resulting in substantial economic loss. Wild grape relatives are valuable resources for the restoration of disease-resistance loci. However, available resistance loci in wild grape genomes remain largely unexplored. In this study, we assembled two phased genomes, including a high resistant Chinese wild grape, *Vitis davidii* Föex, and a susceptible cultivar, *V. vinifera* L. cv. “Manicure Finger”. We detected a total of 36,688 structural variations (SVs), with the genes associated with heterozygous SVs showing an enrichment in allele-specific expression (ASE). Furthermore, we identified eight subgroups of R genes and found that 74.2% of R genes overlap with transposable elements (TEs). Among R genes, NBS-type genes exhibit higher expression profiles in the wild grape genome compared with those in the grape cultivar. Additionally, five specific NBS-type R gene clusters were identified in the wild grape genome that are absent in the cultivar. Through genetic mapping, we identified four quantitative trait loci (QTLs) associated with grape white rot resistance based on the *Vitis davidii* genome, within which six NBS-type R genes exhibit differential expression between wild and cultivated grapes. Overall, our study revealed the landscape of resistance genes in grape genomes, providing valuable genetic resources for further breeding programs.

Keywords: *Vitis davidii*, T2T genome, Plant resistance genes, White rot disease

130 Introduction

131 Grape is one of the most important fruit crops in the world¹. Recent research has
 132 provided genomic, historical, and geographical evidence of grape domestication and
 133 genetic diversity²⁻⁴. Disease is one of the significant threats to grape production, and
 134 severe infection often result in substantial economic loss^{5,6}. The primary objective in
 135 breeding is to enhance stress tolerance, addressing challenges from the ongoing 'arms
 136 race' in plant-pathogen interaction⁷⁻⁹. The most popular varieties used to produce wine
 137 and table grapes belong to the European grape (*V. vinifera* L.)^{10,11}. Although *V.*
 138 *vinifera* L. is adapted to a warm, dry climate during its growing season, their berries
 139 are highly susceptible to fungal diseases¹². Exploring the diversity of global wild
 140 grape relatives is critical, especially for identifying more resistance loci. At the same
 141 time, wild grapes serve as valuable genetic resources for grape breeding¹³, some of
 142 which have been incorporated into breeding programs to enhance both biotic and
 143 abiotic stress tolerance in modern cultivars^{4,14-16}.

144 Significant advances in sequencing technology have opened new avenues for
 145 deeper understanding of agronomic traits and resistance loci. In particular, high-
 146 Fidelity (HiFi) long reads sequencing have greatly improved the accuracy and
 147 reliability of genome assembly. A notable achievement in genome assembly is the
 148 telomere-to-telomere (T2T) approach, designed for assembling entire chromosomes.
 149 This approach provides a more comprehensive view of the complete genome,
 150 eliminating gaps and inaccuracies. For example, the genomes of several important
 151 grape varieties have been assembled (Table S1), including PN40024, Yan73¹⁷,
 152 Thompson Seedless¹⁸, Cabernet Franc¹⁹, Cabernet Sauvignon^{20,21}, Carménère²²,
 153 Chardonnay²³, Merlot²⁴, Nebbiolo²⁵, and Shine Muscat²⁶. The assembly of the wild
 154 accessions has attracted wide attention in the grape community, such as that of the *V.*
 155 *riparis*²⁷, *V. rotundifolia*, *V. acerfolia* and *V. aestivalis*²⁸. Additionally, genome
 156 assemblies for four Chinese wild grape species, including *V. amurensis*²⁹, *V.*
 157 *adenoclada*³⁰, *V. heyneana*³¹ and *V. retordii*¹⁶ have completed genome assembly, have
 158 also been completed.

159 Disease-associated resistance genes (R genes) often encode proteins recognizing

specific pathogens, instigating defense responses, but the number of R genes vary across species³². For example, there are 50-100 nucleotide-binding domain and leucine-rich repeat (NLR) genes characterized in the genomes of *Carica papaya*³³, cucumber³⁴, and watermelon³⁵. The number of NLR genes is highly variable across species and within the same family or genus³⁶⁻⁴¹. Notably, bread wheat stands out with the identification of more than 2000 NLR-encoding genes⁴². In the rice⁴³ and lemon⁴⁴ genome studies, gapless genomic data has greatly expanded understanding of plant genome structure and facilitates the breeding of resistant varieties.

Grape cultivars generally exhibit reduced resistance to biotic and abiotic stresses compared to their wild relatives. Grape white rot is a major fungal disease affecting grape yield and quality. Most cultivated grapes, such as Manicure Finger, a cultivar of *V. Vinifera*, have been shown to be susceptible to white rot⁴⁵. *V. davidii* Föex exhibits considerable white rot resistance⁴⁶⁻⁴⁸. These observations underscore the potential and importance of tapping into *V. davidii* grape resources for sustainable viticulture. To fully understand the landscape of R genes and identify disease resistance loci, we newly sequenced and constructed phase-resolved genomes of a wild grape *V. davidii* (0940) and a cultivar Manicure Finger. Our genomic analysis provides new insights into the evolution of R clusters and identifies new candidate resistance loci, thereby facilitating the development of resistant grape varieties.

Results

Phase-resolved wild and cultivated grape genomes

We generated phase-resolved genome assemblies of a *V. davidii* Föex (namely Vd) grape and a *V. vinifera* L. Manicure Finger (namely MF) grape incorporated PacBio HiFi and Hi-C sequencing data (Figure 1B). Average of approx. 30 Gb of PacBio HiFi reads (55-fold coverage) and approx. 54 Gb of Hi-C paired reads (64-fold coverage) were generated for Vd and MF genomes, respectively (Table S2). We estimated heterozygosity for two grape genomes based on the k-mer method using HiFi reads. We found a lower heterozygosity in the wild grape genome (0.702%) than the cultivated grape (1.290%) genome (Figure S1). Subsequently, we used HiFi and Hi-C reads for genome assembly based on a haplotype-resolved strategy (Figure 1C, Figure S2). As results, we obtained two haplotypes Vd_H1 and Vd_H2 (518.48 Mb and 513.30 Mb) of the wild grape genome and two haplotypes MF_H1 and MF_H2 (489.74 Mb and 486.96 Mb) of the cultivated grape genome (Figure 1A). We found that contributions from seven chromosomes (2, 9, 12, 13, 14, 16, and 19) mainly expanded the size of the Vd genome relative to the MF genome and the GC content of these haplotype assemblies ranges from 34.91% to 34.98% (Table S3). We identified the gene structures for each haplotype, with the number ranging from 31,368 to 33,153 (Figure 2). The evaluation based on Benchmarking Universal Single-Copy Orthologs (BUSCOs) showed that the integrity of these haplotypes reached more than 98% (Table S4). The TE sequences account for a larger proportion in the Vd haplotypes (256.9 Mb and 252.3 Mb) than in the MF haplotypes (235.8 Mb and 232.3 Mb) (Table 1, Table S5). Our analysis found that TEs explain 74.46% of the increase in genome size. Therefore, TEs constitute a significant portion of grape genomes, and contribute to the expansion of grape genome size. In addition, we identified centromeres of all haplotypes by utilizing a previously published 107-nucleotide monomer⁴⁹ (Figure 2, Table S6). The comparative genomic analysis showed that the position of the centromere was conserved, but there were differences in the number of repeats. For example, the length of centromere in chromosome 1 of haplotype Vd_H1 (1.86 Mb) was two times large than the haplotype MF_H1 (0.82 Mb). Through the

analysis of 'TTTAGGG/CCCATTT' distribution, we identified the telomere positions on each haplotype: Vd_H1, 17/19; Vd_H2, 16/19; MF_H1, 17/19 and MF_H2, 17/19 (Figure 1A). Overall, our analyses provided two high-quality phase-resolved grape genomes for further utilization in grape breeding.

The PNT2T genome was the first telomere to telomere genome of the grape⁵⁰. Using PNT2T as a reference, we constructed a genome-wide variation map including 3.24 million SNPs and 36,688 reliable SVs, of which 52.4% were deletions (DELS) and 47.3% were insertions (INSs) (Figure S3, Table S7). Additionally, we identified a 4.04 Mb inversion (INV) on chromosome 14 of Vd genome, which was confirmed to exist on the chromosome through Hi-C reads validation (Figure 1D). We genotyped these SVs and found that the number of heterozygous SVs in the MF genome (5,095) is higher than the Vd genome (2,171) (Figure S4).

Table 1. The genomic features of Vd and MF.

	Vd_H1	Vd_H2	MF_H1	MF_H2	PNT2T
Genome size (Mb)	518.48	513.30	489.74	486.96	494.87
GC content (%)	34.95	34.94	34.91	34.98	35.08
BUSCO (%)	98.4	98.0	98.4	98.4	98.5
Contig N50 (Mb)	26.50	26.13	24.09	24.06	26.90
Number of gene	31,711	31,368	33,351	33,153	37534
Telomeres annotated	36/38	35/38	36/38	36/38	36/38
Centromeres annotated	19/19	19/19	19/19	19/19	19/19
The length of centromere (Mb)	23.22	27.39	18.89	28.49	27.26
TE length ratio in the genome (%)	49.56	49.15	48.16	47.71	48.73
The length of TE (Mb)	257.0	252.3	235.8	232.2	241.1

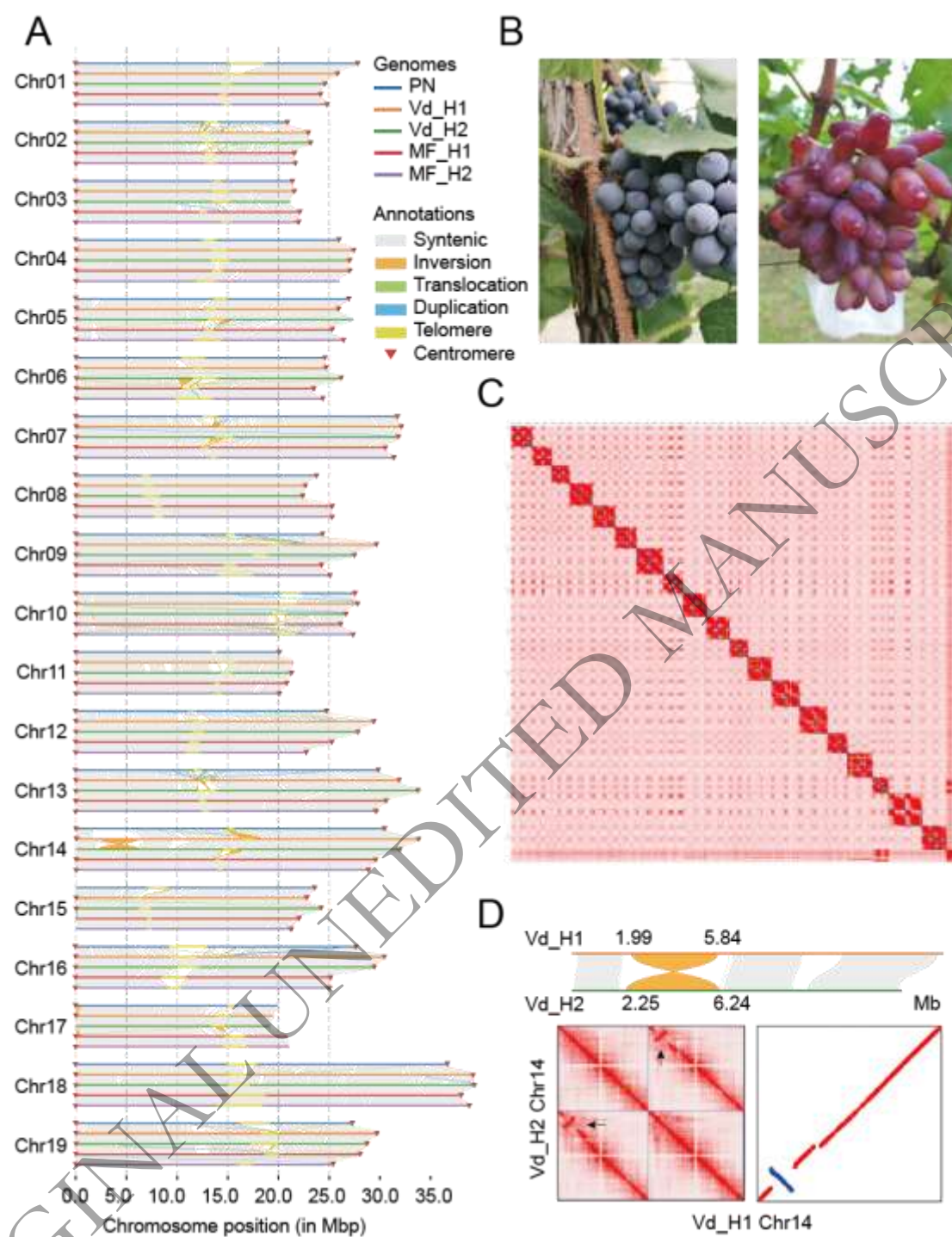


Figure 1. Grape genomes assembly and structural variations. A. Syntenic analysis results among five haplotypes (PNT2T, Vd_H1, Vd_H2, MF_H1 and MF_H2), and the telomeres and centromeres highlight the chromosomes. B. The grape images of a wild grape *V. davidii* (left) and a cultivated grape *V. vinifera* L. Manicure Finger (right). C. Hi-C interactions of chromosomes based on Vd assembly data. The blue square represents pseudo-chromosomes/chromosomes, the green square represents

different contigs. From top to bottom are chromosomes 1 to 19 respectively, and every two blue squares are the same chromosome of Haplotype 1 (upper left) and Haplotype 2 (lower right), respectively. D. A 4.04 Mb heterozygous inversion on chromosome 14 from the Vd genome.

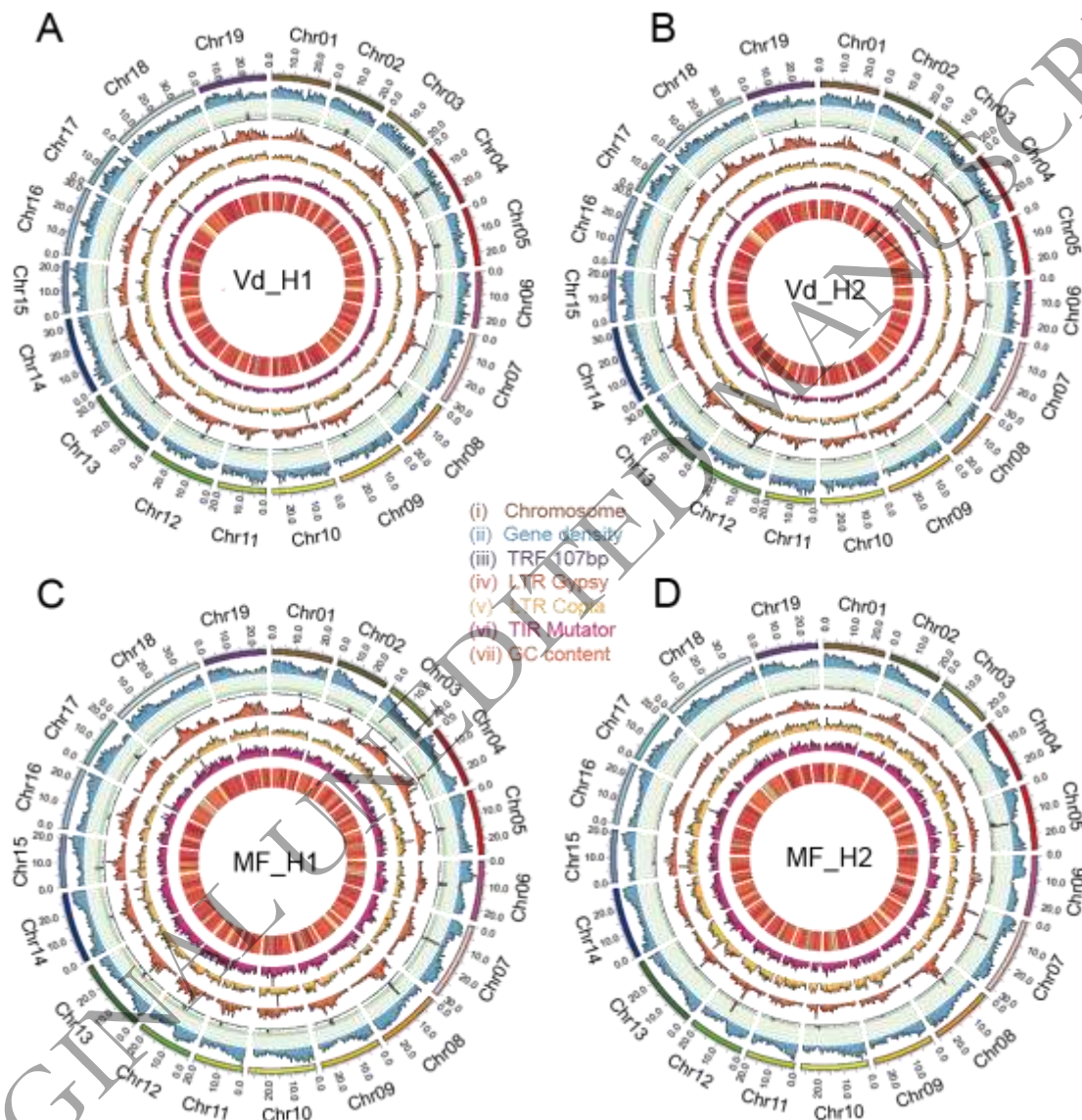


Figure 2. Grape genomes annotation. Genomic features of Vd_H1 (A), Vd_H2 (B), MF_H1 (C) and MF_H2 (D). Tracks from inside to outside are the length of the chromosomes, gene density, distribution of tandem repeats, distribution of LTR/Gypsy, distribution of LTR/Copia, distribution of TIR/Mutator, and GC content were calculated using a 50 kb window.

The comparative genomics and SVs in grapes

Through four assembled haplotypes from Vd and MF grapes, coupled with the PNT2T genome⁵⁰, we investigated gene loss and gain across distinct grapes. To reduce potential annotation errors, we employed the same gene annotation process as that used for the PNT2T genome. Based on the comparative genome analysis and genome-wide collinear blocks, we identified 27,122 genes acting as conserved genes in these grape haplotypes (Figure 3B). We found that the gene number varied across different haplotypes, ranging from 2,613 to 3,430. At the same time, the divergence between two haplotypes from the MF genome were higher than the Vd genome, which were consistent with the higher variable gene number (Figure S5). Furthermore, we used homology clustering to analyze gene families. We classified these gene families into core families, soft-core families, and specific families within the genome. Notably, 16,920 core gene families were identified, with variable soft-core gene families ranging from 304 to 564 (Figure 3A). PNT2T exhibited the highest variability in gene families, potentially associated with the near homozygous genetic background. We identified that the specific gene families across five haplotypes ranged from 76 to 122.

Previous studies revealed that SVs are related to the presence and absence of genes in the genome⁵¹. We focused on heterozygous haplotype-specific insertions or deletions in Vd and MF genomes, respectively. Notably, extensive SVs could impact the gene structure, rendering them hemizygous. For instance, a 82-kb region on chromosome 7 in the wild genome contained four hemizygous genes (Figure 3C), this phenomenon also has been found in MF (Figure S6). Approx. 10.4% genomic regions in cultivated grape genome were influenced by these heterozygous variations and resulted in 3,468 (10.9%) hemizygous genes, and a lower proportion of hemizygous gene detected in the wild grape genome (6.7% hemizygous genes) (Figure S7).

SVs manifested in the promoter region may influence the expression level, thereby contributes to phenotypes⁵². Transcriptome data from fruits were collected and used for identifying allele specific expression (ASE) genes between two haplotypes in wild and cultivated grapes, respectively (Table S8). We conducted a

correlation analysis between ASE and heterozygous SVs situated within the promoter region (2-kb) (Figure 3D). The outcomes revealed notable positive correlations in both wild and cultivated grape genomes (P value $< 7.7\text{e-}12$). In addition, the cultivated grape genome exhibited more ASE genes compared to the wild grape. Specifically, our investigation identified two ASE genes in the cultivated grape genome, *PN003575* encoded a B-box domain protein and *PN020576* encoded an AUXIN RESPONSE FACTOR11 homologous protein (Figure 3E, Figure S8). Two distinct deletions (303 bp and 1034 bp) in their promoter regions might be associated with the different allele expression. The expression of the specific allele harboring a 303 bp deletion in the promoter region is markedly repressed (Figure 3F).

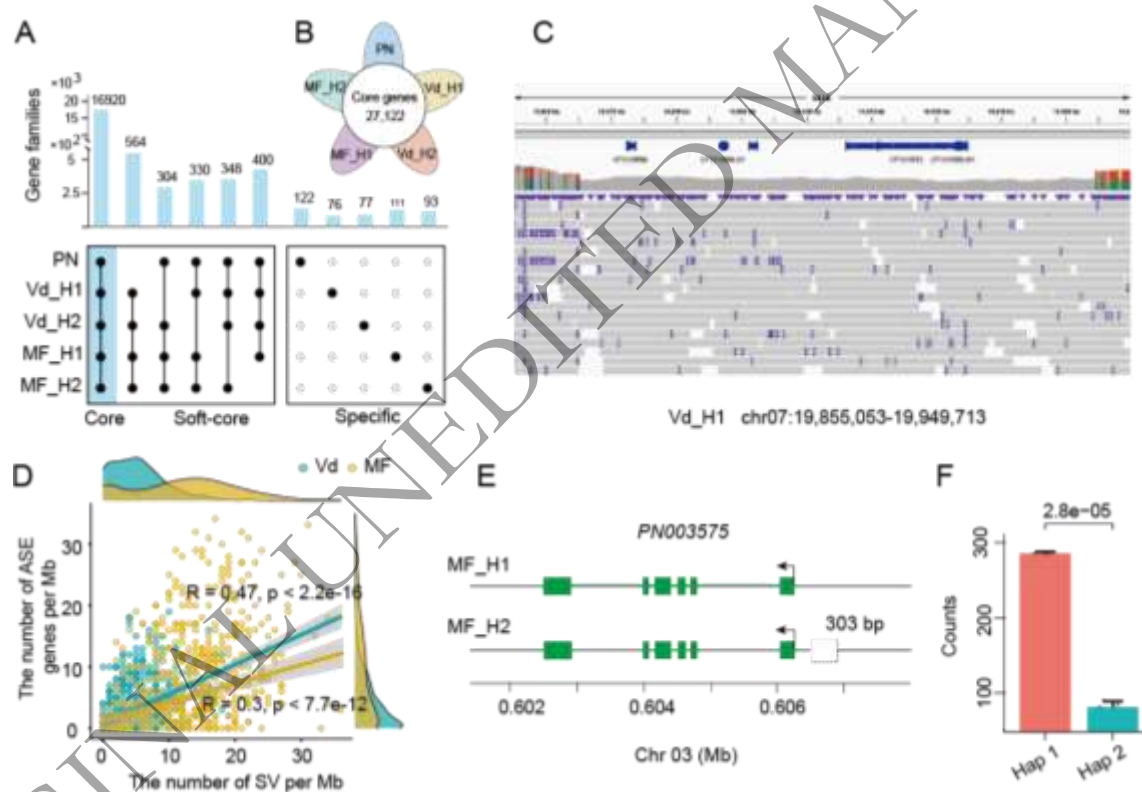


Figure 3. Variable genes and expression affected by SVs. A. The gene family in four newly assembled haplotypes and PNT2T genome. The core and variable (soft-core and specific) gene families were shown. B. The number of conserved genes in five haplotypes. C. A 82 kb heterozygous deletion led to four hemizygous genes in chromosome 7 in the Vd genome. D. Scatterplot between the number of SVs and allele-specific expression genes per Mb. E-F. Schematic representation of the gene structure for *PN003575* on two haplotypes in MF and the expression counts number

for two alleles. The 303 bp deletion was identified by comparing the defined promoter regions between two haplotypes.

R gene landscape in the grape genome

Plant resistance (R) genes are integral components of disease-associated genes that exhibit considerable variation both within and between species⁵³. The wild grape Vd demonstrated stronger resistance compared to the cultivated grape MF⁴⁸. To assess the R genes in grape genomes, we identify genome-wide R genes across eight major categories (CN, CNL, NBS, NL, TN, TNL, RN and RNL) (see Methods). These categories were established based on the single or combinations of protein domains. To validate the classification of R genes, we analyzed high-dimensional features derived from the protein sequence (Figure 4B). Our analysis identified 424 to 478 R genes in four haplotypes from grape Vd and MF genomes (Figure 4A, Table S9). Notably, the CNL, NBS, and NL gene categories constituted an average of 76.5% of R genes in these haplotypes, whereas RNL and RN categories were not included because they contain no more than three copies per haplotype. Although the total count of R genes in different haplotypes showed proximity, only 73% of homologous genes with collinear relationships were identified among the four haplotypes, indicating a considerable dynamic nature of R genes. This dynamism was evident not only between distinct genomes (e.g., MF and Vd), but also within different haplotypes of the same individual. We found that 84% of R genes in the Vd genome are biallelic, compared with 56% in the MF genome. Overall, our analysis sheds light on the dynamism of R genes within wild and cultivated grape genomes.

To insight into the landscape of R genes in grape genomes, we investigated the enrichment of transposable elements (TEs)⁵⁴ within a $\pm$ 50-kb region surrounding the R genes. We randomly selected genes to establish the genome background. Compared to the genome-wide background, we observed a significant enrichment of 74.2% TEs near R genes across all six categories (P value < 0.001) (Figure 4C, Figure S9 and Table S10). The type II transposons (LTR/Copia and LTR/Gypsy) play an important role, leading to an increase in the proportion of TEs.

320 Furthermore, we computed the proportion of intact LTR type elements
321 surrounding R genes across four haplotypes. The results revealed a significantly
322 higher proportion (P value < 0.01) of intact TEs enrichment compared to the genome-
323 wide level (Figure 4D). In addition, we calculated the identity of LTR in associated
324 intact TEs. Higher identity values indicate more recent TE activity. Our findings
325 demonstrated that the identity values of intact LTR/Gypsy and LTR/Copia around R
326 genes exceeds the genome background level (Figure 4E, Figure S10). Interestingly, R
327 genes in the wild Vd grape exhibited higher identity compared to the cultivated MF
328 grape in both intact LTR/Gypsy and LTR/Copia. In addition, we integrated
329 transcriptome data and performed expression analysis of the genome-wide R genes
330 from four haplotypes (Figure 4F). Most R genes exhibited low expression levels, and
331 we defined genes with significantly higher expression as outliers. Compared to
332 cultivated grapes, we found that NBS-type R genes in the wild Vd grape showed more
333 outliers expression.

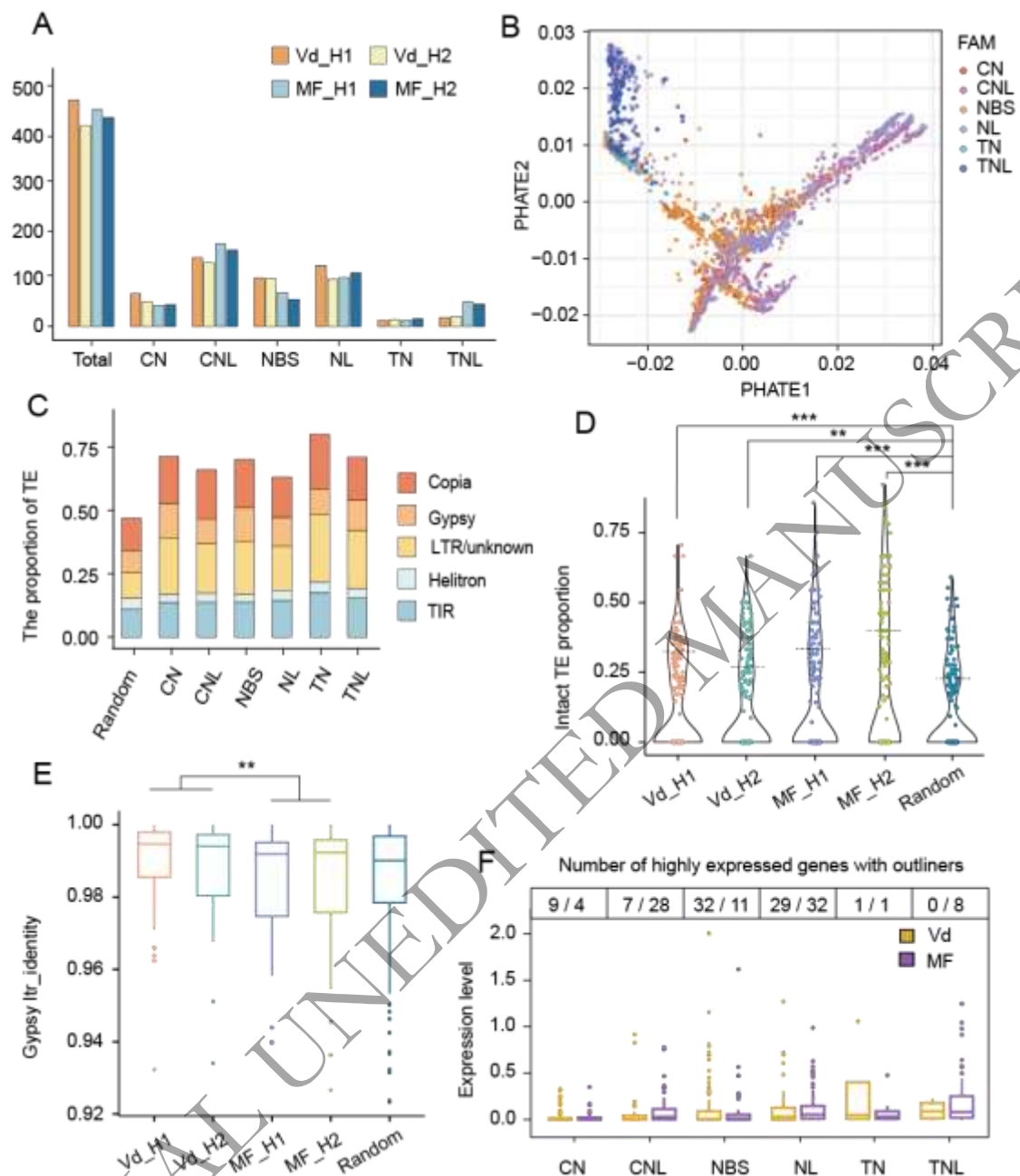


Figure 4. Distribution and pattern of R genes in wild and cultivated grapes. A. The number of six categories R gene on four newly assembled haplotypes. B. PCA analysis of six category R gene protein sequences. C. The proportion of TEs in genomic region (50-kb) nears R gene. The five classifications of TE were highlighted, and the 1% randomly selected genes were used as genome background level (random). D. The proportion of intact LTR type TEs in genomic region (50-kb) nears

R gene. E. The identify value of long terminal repeats in the intact LTR/Gypsy. F. The number of highly expressed R genes in Vd and MF genomes with outliers.

Evolution of R gene clusters

To further investigate the genomic patterns of R genes and their duplication in grapes, we constructed phylogenetic trees based on the protein sequences of different type R genes (Figure S11-12). We observed a prominent clustering effect within the same chromosome (Table S11). We also identified 35 R gene clusters across four haplotypes, mainly in the CNL, NL, and NBS families, with NBS-type R genes showing the highest number of clusters (Figure S13-18). These results suggested that 78% clusters were detected in both wild and cultivated grape genomes; however, within these clusters, only 45% showed homology among four haplotypes. In particular, the phylogeny analysis revealed that 85% of NBS-type R genes were variable, at least one haplotype lacking the homologous gene (Figure 5A).

Our findings suggest that genes from the same cluster were distributed in closer phylogenetic branches. To further elucidate the pattern of R gene clusters in the grape genome, we compared the distribution of R gene clusters between wild and cultivated grapes. We identified 59 and 48 clusters in the Vd and MF genomes, respectively. To compare the clusters from different haplotypes, we used the PNT2T genome as a reference and mapped these clusters. We found significant differences in the number of NBS-type R gene clusters between cultivated and wild grapes (Figure S15). There were five clusters (Cluster 1-5) enriched with multiple NBS-type R genes in the wild Vd grape, whereas in the corresponding positions, without cluster in the cultivated grape genome (Figure 5B). These clusters were distributed across chromosomes 9, 12, 14, and 18, with chromosome 12 containing two clusters, spanning 2.0 Mb (13.5-15.5 Mb) and 1.2 Mb (18.9-20.1 Mb), respectively. Notably, Cluster 1 in the wild Vd grape had 11 NBS-type R genes, while only one gene in the cultivated MF grape at the corresponding location.

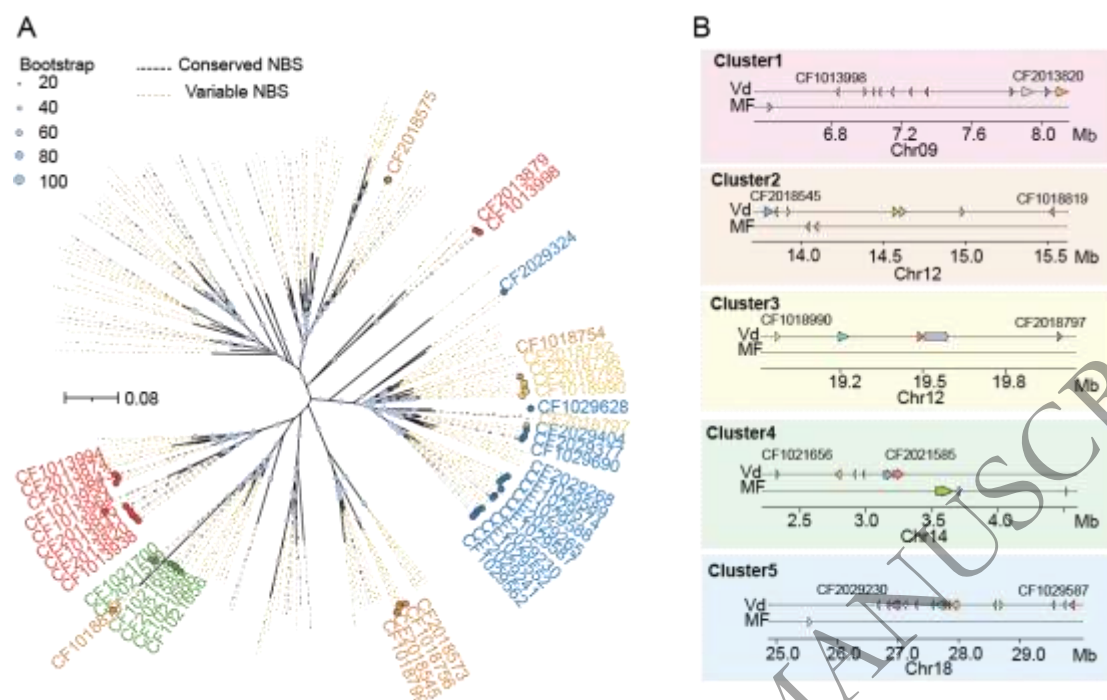


Figure 5. R gene cluster analysis of NBS genes. A. Phylogenetic tree of all NBS-type genes in the Vd genome. The variable NBS-type genes in four haplotypes were highlighted and NBS-type genes from five specific clusters were distinguished. B. Five specific NBS-type R gene clusters in the Vd genome and the syntenic regions in the MF genome.

Improving genetic mapping of R genes involved in white rot resistance.

To unravel the genetic architecture of R genes implicated in the immune response of grapes, we collected expression data from *Coniella diplodiella* (Speg.) Sacc. infection, a pathogen of grape white rot⁴⁸. The expression dataset encompassed experiments conducted on leaves and fruits of both wild Vd and cultivated MF grapes, with and without inoculation of white rot disease. Time series samples were collected at 24 hours and 48 hours post-infection. The phenotypic responses to white rot disease exhibited marked distinctions, with Vd demonstrating significantly higher resistance than MF (Figure 6A). To assess the candidate differentially expressed genes (DEGs) to resistance, we aligned these expression data to the Vd reference genome and compared expression profiles between infected samples at 24 hours and 48 hours and untreated samples (0 hours). The resulting joint expression matrix was used for

downstream analysis, revealing 2905, 5288, 2470, and 5417 genes with significantly changed expression in fruits and leaves of Vd and MF (Vd-berry, Vd-leaf, MF-berry, and MF-leaf) ($\log_{2}FC \geq 2$) (Figure 6B). Notably, the number of differentially expressed genes with increased expression in leaves was approximately twice that in fruits. Among these, 837 genes exhibited significant expression difference in both fruits and leaves of Vd, while 1201 genes showed significant expression difference in both fruits and leaves of MF. We also identified 550 genes with specifically increased expression in the wild Vd grape leaves but not induced in the cultivar MF. Enrichment analysis of these genes in Gene Ontology (GO) functional terms revealed associations with biological processes (Figure S19).

Furthermore, we collected previously sequenced data from the F₁ progeny derived from Vd and MF and aligned these reads to the Vd genome to construct a high-density genetic linkage map. The integration of three years of phenotypic data enabled the mapping of quantitative trait loci (QTLs) associated with the resistance to *C. diploide* (Rcd) (Figure 6C). As expected, the published locus Rcd1 exhibited the highest LOD values and encompassed a notably significant genomic region. In addition to the four previously identified QTLs (Rcd1, Rcd2, Rcd5, and Rcd7)⁴⁸, our investigation unveiled three novel QTLs situated on chromosomes 9, 10, and 17 (Rcd3, Rcd4 and Rcd6). These loci provided a moderate explanation for the phenotype, with Phenotype Variance Explanation (PVE) ranging from 5% to 10%. The cumulative genomic length of these significant regions amounted to 58.3 Mb, constituting 11.3% of the entire genome. Within these regions, we identified 72 NBS genes, representing 40.7% of all NBS-type genes in wild grape Vd genome (Figure S20). We investigated the QTL-associated NBS-type genes and analyzed their expression matrix, identifying six genes with significantly different expression levels (Figure 6D, Figure S21 and Table S14). Four NBS-type R genes exhibited significant down-regulation after inoculation, while two genes demonstrated significant up-regulation. Among them, *CF1069290* showed the most pronounced down-regulation fold change. Our results provide potential evidence for a narrowed set of candidate R genes involved in the response to white rot tolerance within the wild grape genome.

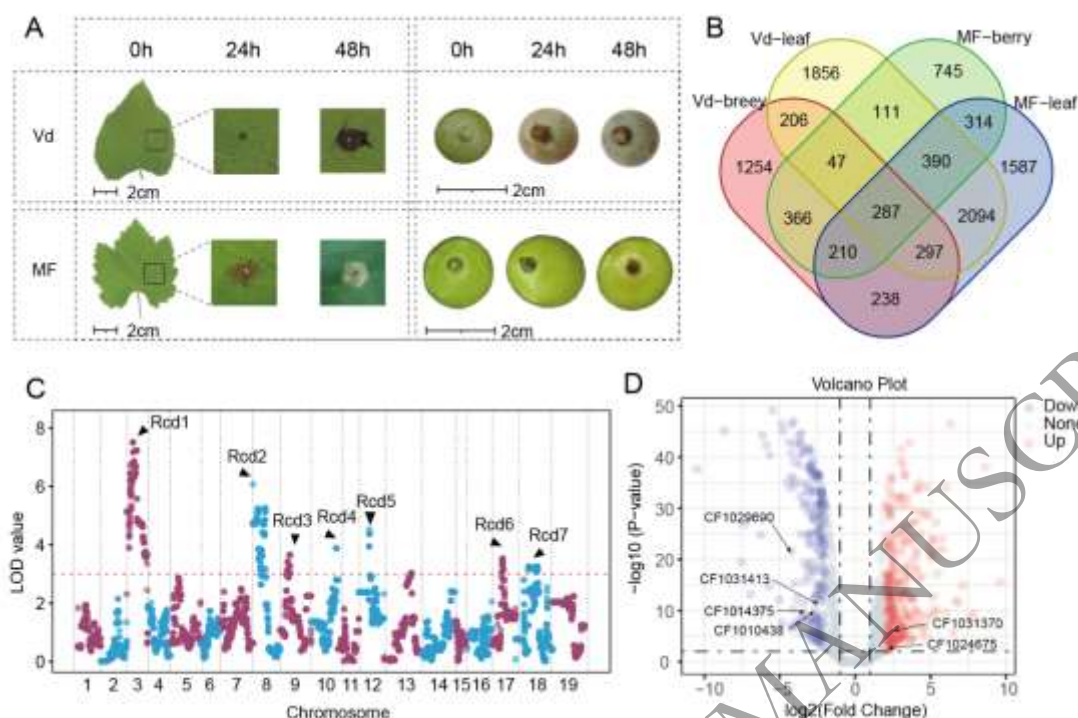


Figure 6. Genetic mapping of white rot resistance in the *Vd* genome. A. Schematic diagram of transcriptome period, organization, and variety relationships in the of *C. diploidie* inoculation experiment. B. Venn diagram of genes with elevated expression in different varieties and tissues after inoculation. C. QTL mapping for the Rcd profiles of F_1 progeny derived from the genetic population. D. Volcano plot of the significant differential expression of *V. davidii* leaves after inoculation with *C. diploidie*.

Discussion

The grape genome has undergone significant improvement, achieving gapless assembly spanning from telomere to telomere in the PNT2T version⁵⁰, marking a major advance in grape breeding. The availability of high-quality genome assemblies is instrumental in expanding the genetic diversity of grape breeding programs. By incorporating assemblies from diverse grapevine cultivars and wild relatives, breeders can access a wider pool of genetic variation. One of the most promising applications of wild grape genome assemblies lies in the identification of disease-resistant loci. The utilization of wild germplasm resources is conducive to improving the effectiveness of grape breeding. Researchers are paying great attention to the genetic information of wild grapes. Noé Cochetel et al. used Illumina HiSeq 4000 sequencing technology to assemble the genomes of 9 American wild grape species, including *V. acerifolia*, *V. aestivalis*, and *V. arizonica*, but did not achieve T2T gap-free assembly quality²⁸. The assembly of four Chinese wild grape genomes has been completed, however, genotyping was not fully achieved, and gaps were present in the genomes of *V. amurensis*²⁹, *V. adenoclada*³⁰, and *V. heyneana*³¹, except the genome of *V. retordii*¹⁶. *V. davidii* is a typical representative of disease-resistant wild grape resources in southeastern China, characterized by excellent disease resistance⁵⁵. Therefore, the assembly of wild grape *V. davidii* will be beneficial for exploring the genetic basis of resistance and further guiding grape breeding.

Our assembly of the wild grape *V. davidii* provide insights into the landscape and genomic patterns of resistance loci in the grape genome. R genes play a crucial role in the immune defense responses of plants by directly recognizing pathogen effectors⁵⁶. The contraction and expansion of R genes are influenced, in part, by selection pressures to adapt to complex environments^{57,58}. Through comparisons of multiple haplotypes from wild and cultivated grape genomes, we found a similar number of R genes. Interestingly, our examination showed that 60% of R genes in the grape genome exhibited variability across grape genomes, even when comparing paired haplotypes within a single sample. The dynamic nature of R genes, contributes to the gene pool essential for constructing intricate resistance networks during plant

457 evolution⁵⁹. In our exploration, we traced the enrichment of TEs in regions proximal
458 to R genes.

459 SVs have the potential to drive phenotypic evolution, with one possible
460 mechanism acting as regulators in promoters, influencing gene expression⁶⁰.
461 Previously, Kim et al.⁶¹ and Maestri et al.²⁵ revealed the evolution of plant disease-
462 resistance genes in pepper plants and their expansion was associated with
463 retroduplication. The association between variable genes in the grape genome and
464 distinct SVs in different species or haplotypes also highlights the impact of SVs on
465 gene copy variation. SVs may insert into the R gene structure, thereby disrupting the
466 R gene. Large SVs covering the entire structure of the R gene have been reported in
467 crop genomes⁶²⁻⁶⁴. Gene duplication leads to the production of redundant copies of the
468 R gene in the grape genomes⁶³. The duplicated R genes can be retained or lost during
469 evolution, thus facilitating rapid evolution against strains with pathogenic factors.
470 Similarly, research by Guan et al.⁶⁵ suggested that genomic SVs drive functional
471 discoveries in peach genomes. Moreover, the role of LTR-type TEs, as indicated by
472 Rispe et al.⁶⁶, significantly impacts R genes evolution. This finding aligns with the
473 study conducted by Lu et al.⁶⁷, which investigated the evolution of wheat sharp
474 eyespot *Rhizoctonia cerealis*.

475 The arrangement of R genes primarily in clusters, exhibiting exceptionally high
476 protein homology within each cluster, marks expansion events in gene family
477 evolution⁶⁸. This clustering suggests potential collaborative interactions among the
478 gene products of paired and larger clusters during immune activation. These five
479 specific NBS-type R gene clusters enriched in the genome of Vd, which might be
480 related to synergistic modulation of immune responses. It is a common for different
481 genes within these clusters to undergo sub-functionalization in response to specific
482 pathogenic factors or even lose function over time.

483 The phase-resolved genomics has opened new possibilities for identifying
484 founder haplotypes and tracing them in breeding populations, as demonstrated in the
485 study by Minamikawa et al.⁶⁹. These studies collectively demonstrate the pivotal role
486 of the T2T reference genome in advancing molecular breeding applications, gene

function discovery, and understanding genome evolution, thus contributing to the improvement of crop breeding and agriculture. After replacing the reference genome with the newly generated assemblies of *V. davidii*, we performed genetic mapping using previously published populations⁴⁸ to investigate novel disease-resistant loci. Our investigation newly identify three QTLs situated on chromosomes 9, 10, and 17 (Rcd3, Rcd4 and Rcd6). Several similar genomics-driven projects have been established. For example, the FruitBreedomics project has highlighted the importance of bridging the gap between molecular genetics research and its application in breeding, thereby increasing breeding efficiency in fruit tree crops⁷⁰. Going forward, combining these new QTLs from *V. davidii* with molecular marker-assisted breeding may significantly accelerate grape resistance breeding.

Acknowledgements.

This study was funded by the National Key Research and Development Program of China (2021YFD1200200), the Agricultural Science and Technology Innovation Program (CAAS-ASTIP-2021-ZFRI), the National Natural Science Foundation of China (31872057), the China Agriculture Research System (CARS-29).

Author contributions

Y.Z conceived and designed the project. Y.L., S.C., X.W., Z.L., X.S., and S.H. performed the genome assembly and gene annotation. P.L. and X.T. performed transcriptome analysis. C.L., X.F., J.J., L.S., and G.L. assisted in grape culture and sample collection. Z.L., Z.M., X.X., Z.J. and Y.Z. assisted in bioinformatics analyses and data visualization. Y.L. and Y.Z. wrote and purified the manuscript.

Data availability

All the raw sequencing data generated for this project have been deposited in the National Genomics Data Center (NGDC) Genome Sequence Archive (GSA) (<https://ngdc.cnrb.ac.cn/gsa/>) with the number of CRA017609. The assembly and annotation as well as the sequences of centromeres and heterozygous regions have been deposited in Zenodo <https://zenodo.org/records/12671741>.

Conflict of interest

The authors declare no conflict of interests.

Materials and Methods

Plant materials and genome sequencing

A wild grape *V. davidii* Föex (0940) and a cultivated grape *V. vinifera* L. Manicure Finger, materials conserved at the National Repository for Grapevine and Peach (Zhengzhou, China). Fresh leaves served as sequencing materials. We collected these leaves and subsequently frozen in liquid nitrogen. These leaves are then ground in a mortar at a low temperature for genomic DNA extraction. We extracted genomic DNA based on the improved CTAB method. Two genomic sequencing methods, HiFi and Hi-C sequencing, were employed in this study. For the Hi-C libraries, we constructed by digestion with the four-cutter restriction enzyme MboI. The prepared libraries were sequenced on the Illumina HiSeq X Ten platform (San Diego, CA, USA) using 150-bp paired mode. For HiFi sequencing, high-fidelity reads were sequenced on the PacBio Sequel II platform using the circular consensus sequencing (CCS) mode. The raw sequencing data was preprocessed using SMRTlink software, which included removing adapters and correcting subreads to generate high-quality HiFi sequences for further analysis.

The PNT2T genome⁵⁰ as a reference, which was sequenced using PacBio HiFi and fully annotated. Additionally, collected RNA-seq data from NCBI for 0h, 24h, and 48h after inoculation with live white rot bacteria on leaves⁴⁸ and berries (SRR shown in Table S12).

Genome assembly

We combined HiFi and Hi-C data to assemble the genomes of two grape accessions using the Hifiasm⁷¹ program with default parameters. For each accession, we obtain two separated contig-level haplotype assemblies. The heterozygosity was investigated using a k-mer based approach integrated in GenomeScope⁷² v2.0. To improve the assembly, the Hi-C data were utilized to anchor and remove certain short contigs (< 100 kb). The program RagTag⁷³ was

employed to determine the approximate order of contigs on chromosomes, with PNT2T serving as the reference genome. The Hi-C reads were further used to anchor all contigs using Juicer⁷⁴ v1.5, followed by a 3D-DNA⁷⁵ scaffolding pipeline. Subsequently, we manually adjusted these genome assemblies using Juicebox^{76,77} before rerunning 3D-DNA to obtain the genome at the chromosome-level. The completeness of four haplotypes was evaluated using BUSCOs⁷⁸ with the *embryophyta_odb10* database, while the genome continuity was assessed by calculating the length of contig N50.

Identification of telomeres and centromeres

We used the telomere pipeline developed by TIDK⁷⁹ v.0.2.0 to identify the repeated unit of telomeres. As a result, we identify telomeric sequences characterized by the unit 5'-CCCATTT and 3'-TTTAGGG. The telomere peak line was visualized using R scripts based on rapid statistics of telomeres. Tandem repeat annotation was performed using TRF⁸⁰ v4.09, and the results were merged using the TRF2GFF script. Data statistics was completed by extracting information using the `awk` command in the Linux system and visualizing the results in Integrative Genomics Viewer⁸¹ (IGV, v2.12). The TRF program was used to scan tandem repeats in the genome. According to *V. vinifera* centromeric repeat sequences, we confirmed a 107-nucleotide monomer (AGTACCGAAAAAGGGTCGAATCAGTGTGAGTACCGAAAAATGGTAGAA TCCGGGCGAGTACCGGGAAAAGGTAGAATCCGTGCGAGTATCGAAAAAC TGTCCGGGCG) in the centromere region of each chromosome of two assemblies.

Genome annotation

The gene structure was annotated based on evidence from expression data and homologous protein. These expression data were aligned to the softmasked assembly using Hisat2⁸² v2.10.2 and assembled into transcripts with StringTie⁸³ v1.3.0. To identify genes, transcripts were initially searched using UniProt⁸⁴ as supporting evidence. For genome annotation, we primarily utilized the multiple programs including MAKER⁸⁵, SNAP⁸⁶ and PASA⁸⁷. We constructed a preliminary gene model, followed by further searching with AUGUSTUS⁸⁸ v3.4.0. Genes

located in duplicated regions, with CDS regions shorter than 90, or lacking evidence were filtered out. Additionally, we checked missing genes and refined gene models based on alternative splicing analysis. Consequently, we validated these gene structures using the Hidden Markov Models based on the Pfam⁸⁹ database and filtered the low confidence gene structure.

To identify genomic TEs, we annotated repeated sequences by the Extensive de novo TE Annotator⁹⁰ (EDTA v2.1.0) and categorized into four major clades including: long terminal repeat (LTR), terminal inverted repeat (TIR), long interspersed nuclear element (LINE) repeats and Helitron-like DNA transposons. For the downstream annotation, we masked repeat element sequences using the program RepeatMasker⁹¹ v4.1.2.

Resistance genes identification and validation

We detected genome-wide R genes based on the domain and motif structures using four haplotypes from two grape accessions. The RGAugury⁹² v.2.2 pipeline and Resistify⁹³ were used to identify the putative nucleotide-binding site (NB-ARC) domain-encoding genes into eight subgroups: TN (Toll/interleukin-1 receptor (TIR) and NB-ARC), CN (coiled-coil (CC) and NB-ARC), NL (NB-ARC and leucine rich repeat (LRR)), CNL (CC, NB-ARC and LRR), NB (NB-ARC), TNL (TIR, NB-ARC and LRR), RN (RPW8 and NB-ARC) and RNL (RPW8, NB-ARC and LRR). To verify the trustworthiness of R gene, we generated features of R protein sequences using the program iFeatureOmega⁹⁴. To the best of our knowledge, this program expands its functionality by integrating 15 feature analysis algorithms. Collected high dimension features from composition of three collections: k-spaced amino acid pairs (CKSAAP), Dipeptide composition (DPC) and Conjoint k-spaced triad (KSCTriad). The high dimension matrix was introduced to the program PHATE⁹⁵ for dimension reduction and used for validating of R genes.

We analyzed TE enrichment surrounding each R gene and classified the TEs according to their respective families. Search intervals were defined as 50 kb upstream of the gene start site and 50 kb downstream of the gene end site. Overlapping TE structures within these intervals were included in the enrichment

analysis. To provide a genomic comparison, we randomly selected an equivalent number of non-R genes and analyzed their corresponding intervals of identical length, repeating this process 100 times. To evaluate the robustness of our findings, we employed a jackknife subsampling test ($\alpha = 0.05$) and performed a Student's t-test for statistical significance, which yielded the final P value.

Analysis of genomic collinearity

We used Minimap2⁹⁶ v2.24 to align the haplotype sequences to the reference genome, and the resulting alignment BAM file was indexed and sorted using SAMtools⁹⁷ v1.11. Then, we deployed the Synteny and Rearrangement Identifier⁹⁸ (SyRI) to identify collinear orthologs and genomic large rearrangements between the two haplotypes of the Vd and MF genome. For visual interpretation of the data, the Plotsr⁹⁹ tool was used. Additionally, MUMmer¹⁰⁰ v4.0 was used for an in-depth comparison of the Vd genome with the PNT2T reference genome. We aligned genome sequences using nucmer (nucmer --mum) and filtered the results (delta-filter -i 95 -l 10000). Subsequently, we visualized the comparison results in dot plots. These readable comparison data were extracted using the 'show-coords -T -q -H'.

The analysis of SVs and hemizygous genes

Using SVIM-asm¹⁰¹ v1.0.3 to detect chromosomal SVs in Vd and MF, with PNT2T as the reference genome. First, the four haplotypes were compared to PNT2T using minimap2 to obtain bam files; Second, SVs were detected using SVIM-asm. For each identified SV, we viewed it in IGV to confirm its authenticity. Survivors¹⁰² v1.0.7 was used to integrate the SVs on the four haplotypes, obtaining the SV datasets of Vd and MF relative to the PNT2T reference genome. Only SVs supported by multiple genomes were included as potential SVs here. We combined the SV map with gene annotation to identify hemizygous genes in the Vd and MF genomes. As a result, we filtered heterozygous SVs and identified hemizygous genes according to whether there is overlap ($> 10\%$) between SV and gene coding regions.

Phylogenetic tree and gene clusters

The phylogenetic tree was constructed using two methods. First, using MEGA¹⁰³ software (bootstrap=1000) to construct a phylogenetic tree for all NBS sequences. The second phylogenetic tree was drawn in Python using the ETE Toolkit¹⁰⁴ for the NBS cluster sequences. Downloaded a gene family with an hmm structure from the Pfam database and used hmmsearch software to predict conserved domains. A cluster was defined as a 1.0 Mb genomic region containing at least three homologous proteins.

Inoculation treatment

We perform inoculation treatment at the National Repository for Grapevine in Zhengzhou, China, using *V. davidii* (0940) and Manicure Finger grape varieties. We punctured fruits pedicels from both grape varieties with a sterile needle. Fungal mycelium was manually exposed to the wound mediated by a potato dextrose agar (PGA) block. The treated fruits were incubated at 28°C. Samples were collected at 24h, 48h post-inoculation, and the uninfected fruits serving as the negative control. We also conduct three biological replicates, and all samples were utilized for subsequent expression analysis. The RNA-seq data of leaves inoculation treatment was collected from BioSample accession number PRJNA938012.

Analysis of RNA-seq datasets

We filtered and trimmed the raw RNA-seq data using the FastQC¹⁰⁵ program. These filtered RNA-seq reads were aligned to the reference genome using the hisat2 program. Using the featureCounts¹⁰⁶ -p --countReadPairs function, the sam files of all periods were counted according to the gene ID order in the gtf file, and the results were converted to Transcripts per million (TPM) for data normalization. We estimated the allelic expression of genes by counting the reads at the heterozygous SNPs. For a gene, the allelic imbalance level was calculated as the mean of the reads count. We profiled significant allele-specific expression using 'Fold Change' $\geq$ 2 between alleles and FDR corrected p-values < 0.05 calculated based on the EdgeR¹⁰⁷ package in R.

Genetic map construction and QTL mapping

The Vd genome was indexed using the BWA¹⁰⁸ tool and the population-level whole genome resequencing data was aligned to the Vd genome. The BWA-generated BAM files were examined using GATK¹⁰⁹ markdropped function, with the 'remove_dropped' option set to 'true'. To enhance precision, the BAM files were sorted by coordinates using the 'sort' command and then merged the sorted BAM files to extract SNPs for each sample using the GTX¹¹⁰ tool. The final SNP dataset was generated using the 'GTX -joint' mode. Quality control was conducted using the PLINK¹¹¹ tool with the '-geno 0.2 --maf 0.05' parameters.

We filtered the mismatch variants and collected available marker datasets for the fine mapping in the F₁ population. To increase the accuracy of the association, we selected markers that covered all progeny. These markers showing significant separation distortion (P value < 0.001) were also filtered. The high-quality genetic markers were assigned to linkage groups integrating chromosome information. Each linkage group was organized using LepMap3¹¹² software. We calculated the genetic distance follow the Kosambi function.

R/qt12¹¹³ was used for QTL analysis. We choose the method of Haley-Knott regression to detect potential QTLs. To make the results more reliable, we also performed a thousand permutations to determine the statistics. We identify a threshold value ($LOD = 3.0$) to be considered significant. Phenotypic variation explained (PVE) was estimated based on candidate loci, and confidence intervals were defined using 95% Bayesian credible intervals.

688 **References**

- 689 1 Alston, J. M. & Sambucci, O. in *The Grape Genome* (eds Dario Cantu & M. Andrew
690 Walker) 1-24 (Springer International Publishing, 2019).
- 691 2 Dong, Y. *et al.* Dual domestications and origin of traits in grapevine evolution.
692 *Science* **379**, 892-901, doi:10.1126/science.add8655 (2023).
- 693 3 Zhou, Y., Massonnet, M., Sanjak, J. S., Cantu, D. & Gaut, B. S. Evolutionary
694 genomics of grape (*Vitis vinifera* ssp. *vinifera*) domestication. *Proc Natl Acad Sci U S*
695 *A* **114**, 11715-11720, doi:10.1073/pnas.1709257114 (2017).
- 696 4 Xiao, H. *et al.* Adaptive and maladaptive introgression in grapevine domestication.
697 *Proc Natl Acad Sci U S A* **120**, e2222041120, doi:10.1073/pnas.2222041120 (2023).
- 698 5 Myles, S. *et al.* Genetic structure and domestication history of the grape. *Proc Natl*
699 *Acad Sci U S A* **108**, 3530-3535, doi:10.1073/pnas.1009363108 (2011).
- 700 6 Allaby, R. Clonal crops show structural variation role in domestication. *Nature Plants*
701 **5**, 915-916, doi:10.1038/s41477-019-0515-8 (2019).
- 702 7 Zhang, G. *et al.* The jasmonate-ZIM domain gene VqJAZ4 from the Chinese wild
703 grape *Vitis quinquangularis* improves resistance to powdery mildew in *Arabidopsis*
704 *thaliana*. *Plant Physiol Biochem* **143**, 329-339, doi:10.1016/j.plaphy.2019.09.018
705 (2019).
- 706 8 Cadle-Davidson, L. Variation Within and Between *Vitis* spp. for Foliar Resistance to
707 the Downy Mildew Pathogen *Plasmopara viticola*. *Plant Dis* **92**, 1577-1584,
708 doi:10.1094/pdis-92-11-1577 (2008).
- 709 9 Cadle-Davidson, L., Chicoine, D. R. & Consolie, N. H. Variation Within and Among

- 710 *Vitis* spp. for Foliar Resistance to the Powdery Mildew Pathogen *Erysiphe necator*.
711 *Plant Dis* **95**, 202-211, doi:10.1094/pdis-02-10-0092 (2011).
- 712 10 Alleweldt, G. & Possingham, J. V. Progress in grapevine breeding. *Theoretical and*
713 *Applied Genetics* **75**, 669-673, doi:10.1007/BF00265585 (1988).
- 714 11 Reynolds, A. G. in *Grapevine Viruses: Molecular Biology, Diagnostics and*
715 *Management* (eds Baozhong Meng, Giovanni P. Martelli, Deborah A. Golino, & Marc
716 Fuchs) 3-29 (Springer International Publishing, 2017).
- 717 12 Hedrick, U. P. & Booth, N. O.
- 718 13 Daldoul, S., Boubakri, H., Gargouri, M. & Mliki, A. Recent advances in
719 biotechnological studies on wild grapevines as valuable resistance sources for smart
720 viticulture. *Molecular Biology Reports* **47**, 3141-3153, doi:10.1007/s11033-020-
721 05363-0 (2020).
- 722 14 Foria, S. *et al.* Extent of wild-to-crop interspecific introgression in grapevine (*Vitis*
723 *vinifera*) as a consequence of resistance breeding and implications for the crop
724 species definition. *Hortic Res* **9**, doi:10.1093/hr/uhab010 (2022).
- 725 15 Migicovsky, Z. *et al.* Genomic ancestry estimation quantifies use of wild species in
726 grape breeding. *BMC Genomics* **17**, 478, doi:10.1186/s12864-016-2834-8 (2016).
- 727 16 Zhang, T. *et al.* Population genomics highlights structural variations in local
728 adaptation to saline coastal environments in woolly grape. *J Integr Plant Biol*,
729 doi:10.1111/jipb.13653 (2024).
- 730 17 Zhang, K. *et al.* The haplotype-resolved T2T genome of teinturier cultivar Yan73
731 reveals the genetic basis of anthocyanin biosynthesis in grapes. *Horticulture*

- 732 *Research* **10**, doi:10.1093/hr/uhad205 (2023).
- 733 18 Wang, X. *et al.* Telomere-to-telomere and gap-free genome assembly of a
734 susceptible grapevine species (Thompson Seedless) to facilitate grape functional
735 genomics. *Horticulture Research* **11**, doi:10.1093/hr/uhad260 (2023).
- 736 19 Vondras, A. M. *et al.* Rootstock influences the effect of grapevine leafroll-associated
737 viruses on berry development and metabolism via abscisic acid signalling. *Mol Plant*
738 *Pathol* **22**, 984-1005, doi:10.1111/mpp.13077 (2021).
- 739 20 Chin, C. S. *et al.* Phased diploid genome assembly with single-molecule real-time
740 sequencing. *Nature methods* **13**, 1050-1054, doi:10.1038/nmeth.4035 (2016).
- 741 21 Minio, A., Massonnet, M., Figueroa-Balderas, R., Castro, A. & Cantu, D. Diploid
742 Genome Assembly of the Wine Grape Carmenère. *G3 (Bethesda)* **9**, 1331-1337,
743 doi:10.1534/g3.119.400030 (2019).
- 744 22 Minio, A. *et al.* Iso-Seq Allows Genome-Independent Transcriptome Profiling of
745 Grape Berry Development. *G3 (Bethesda)* **9**, 755-767, doi:10.1534/g3.118.201008
746 (2019).
- 747 23 Zhou, Y. *et al.* The population genetics of structural variants in grapevine
748 domestication. *Nature Plants* **5**, 965-979, doi:10.1038/s41477-019-0507-8 (2019).
- 749 24 Massonnet, M. *et al.* The genetic basis of sex determination in grapes. *Nat Commun*
750 **11**, 2902, doi:10.1038/s41467-020-16700-z (2020).
- 751 25 Maestri, S. *et al.* 'Nebbiolo' genome assembly allows surveying the occurrence and
752 functional implications of genomic structural variations in grapevines (*Vitis vinifera* L.).
753 *BMC Genomics* **23**, 159, doi:10.1186/s12864-022-08389-9 (2022).

- 754 26 Shirasawa, K. *et al.* De novo whole-genome assembly in an interspecific hybrid table
755 grape, 'Shine Muscat'. *DNA Res* **29**, doi:10.1093/dnares/dsac040 (2022).
- 756 27 Girollet, N. *et al.* De novo phased assembly of the *Vitis riparia* grape genome.
757 *Scientific Data* **6**, 127, doi:10.1038/s41597-019-0133-3 (2019).
- 758 28 Cochetel, N. *et al.* A super-pangenome of the North American wild grape species.
759 *Genome Biol* **24**, 290, doi:10.1186/s13059-023-03133-2 (2023).
- 760 29 Wang, Y. *et al.* The genome of Shanputao (*Vitis amurensis*) provides a new insight
761 into cold tolerance of grapevine. *Plant J* **105**, 1495-1506, doi:10.1111/tpj.15127
762 (2021).
- 763 30 Cheng, G. *et al.* Chromosome-scale genomics, metabolomics, and transcriptomics
764 provide insight into the synthesis and regulation of phenols in *Vitis adenoclada*
765 grapes. *Frontiers in Plant Science* **14**, doi:10.3389/fpls.2023.1124046 (2023).
- 766 31 Qin, B. *et al.* The complete chloroplast genome of *Vitis heyneana* Roem. et Schult.,
767 an economic plant to China. *Mitochondrial DNA Part B* **5**, 562-563,
768 doi:10.1080/23802359.2019.1710289 (2020).
- 769 32 Ijaz, S., Haq, I. U., Babar, M. & Nasir, B. in *Cereal Diseases: Nanobiotechnological*
770 *Approaches for Diagnosis and Management* (eds Kamel A. Abd-Elsalam & Heba I.
771 Mohamed) 249-269 (Springer Nature Singapore, 2022).
- 772 33 Porter, B. W. *et al.* Genome-wide analysis of *Carica papaya* reveals a small NBS
773 resistance gene family. *Mol Genet Genomics* **281**, 609-626, doi:10.1007/s00438-009-
774 0434-x (2009).
- 775 34 Wan, H. *et al.* Genome-wide analysis of NBS-encoding disease resistance genes in

776 Cucumis sativus and phylogenetic study of NBS-encoding genes in Cucurbitaceae
 777 crops. *BMC Genomics* **14**, 109, doi:10.1186/1471-2164-14-109 (2013).

778 35 Guo, S. *et al.* The draft genome of watermelon (*Citrullus lanatus*) and resequencing
 779 of 20 diverse accessions. *Nat Genet* **45**, 51-58, doi:10.1038/ng.2470 (2013).

780 36 Luo, S. *et al.* Dynamic nucleotide-binding site and leucine-rich repeat-encoding genes
 781 in the grass family. *Plant physiology* **159**, 197-210 (2012).

782 37 Shao, Z.-Q. *et al.* Long-term evolution of nucleotide-binding site-leucine-rich repeat
 783 genes: understanding gained from and beyond the legume family. *Plant physiology*
 784 **166**, 217-234 (2014).

785 38 Zhang, Y. M. *et al.* Uncovering the dynamic evolution of nucleotide-binding site-
 786 leucine-rich repeat (NBS-LRR) genes in Brassicaceae. *Journal of Integrative Plant*
 787 *Biology* **58**, 165-177 (2016).

788 39 Qian, L.-H. *et al.* Distinct patterns of gene gain and loss: diverse evolutionary modes
 789 of NBS-encoding genes in three Solanaceae crop species. *G3: Genes, Genomes,*
 790 *Genetics* **7**, 1577-1585 (2017).

791 40 Tirnaz, S. *et al.* Resistance gene analogs in the Brassicaceae: Identification,
 792 characterization, distribution, and evolution. *Plant physiology* **184**, 909-922 (2020).

793 41 Liu, Y. *et al.* An angiosperm NLR Atlas reveals that NLR gene reduction is associated
 794 with ecological specialization and signal transduction component deletion. *Molecular*
 795 *Plant* **14**, 2015-2031, doi:<https://doi.org/10.1016/j.molp.2021.08.001> (2021).

796 42 Gu, L., Si, W., Zhao, L., Yang, S. & Zhang, X. Dynamic evolution of NBS-LRR genes
 797 in bread wheat and its progenitors. *Mol Genet Genomics* **290**, 727-738,

798 doi:10.1007/s00438-014-0948-8 (2015).

799 43 Li, K. *et al.* Gapless indica rice genome reveals synergistic contributions of active
800 transposable elements and segmental duplications to rice genome evolution. *Mol*
801 *Plant* **14**, 1745-1756, doi:10.1016/j.molp.2021.06.017 (2021).

802 44 Bao, Y. *et al.* A gap-free and haplotype-resolved lemon genome provides insights into
803 flavor synthesis and huanglongbing (HLB) tolerance. *Hortic Res* **10**, uhad020,
804 doi:10.1093/hr/uhad020 (2023).

805 45 Zhang, Y. *et al.* Genome-wide assessment of population structure, linkage
806 disequilibrium and resistant QTLs in Chinese wild grapevine. *Scientia Horticulturae*
807 **215**, 59-64, doi:<https://doi.org/10.1016/j.scienta.2016.12.014> (2017).

808 46 Chen, H. *et al.* The class B heat shock factor HSFB1 regulates heat tolerance in
809 grapevine. *Hortic Res* **10**, uhad001, doi:10.1093/hr/uhad001 (2023).

810 47 Liu, X. *et al.* Natural variations of HSFA2 enhance thermotolerance in grapevine.
811 *Hortic Res* **10**, uhac250, doi:10.1093/hr/uhac250 (2023).

812 48 Li, P. *et al.* QTL detection and candidate gene analysis of grape white rot resistance
813 by interspecific grape (*Vitis vinifera* L. × *Vitis davidii* Foex.) crossing. *Hortic Res* **10**,
814 uhad063, doi:10.1093/hr/uhad063 (2023).

815 49 Li, B. & Gschwend, A. R. *Vitis labrusca* genome assembly reveals diversification
816 between wild and cultivated grapevine genomes. *Front Plant Sci* **14**, 1234130,
817 doi:10.3389/fpls.2023.1234130 (2023).

818 50 Shi, X. *et al.* The complete reference genome for grapevine (*Vitis vinifera* L.) genetics
819 and breeding. *Horticulture Research* **10**, doi:10.1093/hr/uhad061 (2023).

- 820 51 He, Q. *et al.* A graph-based genome and pan-genome variation of the model plant
821 *Setaria*. *Nature Genetics* **55**, 1232-1242, doi:10.1038/s41588-023-01423-w (2023).
- 822 52 Spielmann, M., Lupiáñez, D. G. & Mundlos, S. Structural variation in the 3D genome.
823 *Nature Reviews Genetics* **19**, 453-467, doi:10.1038/s41576-018-0007-0 (2018).
- 824 53 Joshi, R. K. & Nayak, S. Perspectives of genomic diversification and molecular
825 recombination towards R-gene evolution in plants. *Physiol Mol Biol Plants* **19**, 1-9,
826 doi:10.1007/s12298-012-0138-2 (2013).
- 827 54 Chu, C. *et al.* Comprehensive identification of transposable element insertions using
828 multiple sequencing technologies. *Nature Communications* **12**, 3836,
829 doi:10.1038/s41467-021-24041-8 (2021).
- 830 55 Niu, S. *et al.* Examination of molecular mechanism for the color mutation in Chinese
831 wild grapevine (*Vitis davidii*). *Acta Physiologiae Plantarum* **39**, 171,
832 doi:10.1007/s11738-017-2473-y (2017).
- 833 56 Venkatesh, J. & Kang, B.-C. Current views on temperature-modulated R gene-
834 mediated plant defense responses and tradeoffs between plant growth and immunity.
835 *Current Opinion in Plant Biology* **50**, 9-17,
836 doi:<https://doi.org/10.1016/j.pbi.2019.02.002> (2019).
- 837 57 Freitas, L. & Nery, M. F. Expansions and contractions in gene families of
838 independently-evolved blood-feeding insects. *BMC Evolutionary Biology* **20**, 87,
839 doi:10.1186/s12862-020-01650-3 (2020).
- 840 58 Wu, J.-Y., Xue, J.-Y. & Van de Peer, Y. Evolution of NLR Resistance Genes in
841 Magnoliids: Dramatic Expansions of CNLs and Multiple Losses of TNLs. *Frontiers in*

842 *Plant Science* **12**, doi:10.3389/fpls.2021.777157 (2021).

843 59 Märkle, H., Saur, I. M. L. & Stam, R. Evolution of resistance (R) gene specificity.

844 *Essays Biochem* **66**, 551-560, doi:10.1042/ebc20210077 (2022).

845 60 Xu, Z., Li, Q., Marchionni, L. & Wang, K. PhenoSV: interpretable phenotype-aware

846 model for the prioritization of genes affected by structural variants. *Nature*

847 *Communications* **14**, 7805, doi:10.1038/s41467-023-43651-y (2023).

848 61 Kim, S. *et al.* New reference genome sequences of hot pepper reveal the massive

849 evolution of plant disease-resistance genes by retroduplication. *Genome Biol* **18**, 210,

850 doi:10.1186/s13059-017-1341-9 (2017).

851 62 Su, Y. *et al.* Phased telomere-to-telomere reference genome and pangenome reveal

852 an expansion of resistance genes during apple domestication. *Plant Physiology*,

853 doi:10.1093/plphys/kiae258 (2024).

854 63 Long, Q. *et al.* Population comparative genomics discovers gene gain and loss during

855 grapevine domestication. *Plant Physiology* **195**, 1401-1413,

856 doi:10.1093/plphys/kiae039 (2024).

857 64 Tang, D. *et al.* Genome evolution and diversity of wild and cultivated potatoes. *Nature*

858 **606**, 535-541, doi:10.1038/s41586-022-04822-x (2022).

859 65 Guan, J. *et al.* Genome structure variation analyses of peach reveal population

860 dynamics and a 1.67 Mb causal inversion for fruit shape. *Genome Biol* **22**, 13,

861 doi:10.1186/s13059-020-02239-1 (2021).

862 66 Rispe, C. *et al.* The genome sequence of the grape phylloxera provides insights into

863 the evolution, adaptation, and invasion routes of an iconic pest. *BMC Biol* **18**, 90,

doi:10.1186/s12915-020-00820-5 (2020).

67 Hessenauer, P. *et al.* Hybridization and introgression drive genome evolution of
 Dutch elm disease pathogens. *Nat Ecol Evol* **4**, 626-638, doi:10.1038/s41559-020-
 1133-6 (2020).

68 Mizuno, H. *et al.* Evolutionary dynamics and impacts of chromosome regions carrying
 R-gene clusters in rice. *Scientific Reports* **10**, 872, doi:10.1038/s41598-020-57729-w
 (2020).

69 Minamikawa, M. F. *et al.* Tracing founder haplotypes of Japanese apple varieties:
 application in genomic prediction and genome-wide association study. *Hortic Res* **8**,
 49, doi:10.1038/s41438-021-00485-3 (2021).

70 Van de Weg, E. *et al.* *The New EU Project FruitBreedomics: an Integrated Approach
 for Increasing Breeding Efficiency in Fruit Tree Crops.* (2012).

71 Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de
 novo assembly using phased assembly graphs with hifiasm. *Nature methods* **18**, 170-
 175, doi:10.1038/s41592-020-01056-5 (2021).

72 Vurture, G. W. *et al.* GenomeScope: fast reference-free genome profiling from short
 reads. *Bioinformatics* **33**, 2202-2204, doi:10.1093/bioinformatics/btx153 (2017).

73 Alonge, M. *et al.* Automated assembly scaffolding using RagTag elevates a new
 tomato system for high-throughput genome editing. *Genome Biol* **23**, 258,
 doi:10.1186/s13059-022-02823-7 (2022).

74 Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-
 Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98, doi:10.1016/j.cels.2016.07.002

886 (2016).

887 75 Dudchenko, O. *et al.* De novo assembly of the *Aedes aegypti* genome using
888 Hi-C yields chromosome-length scaffolds. *Science* **356**, 92-95,
889 doi:doi:10.1126/science.aal3327 (2017).

890 76 Robinson, J. *et al.* Juicebox.js Provides a Cloud-Based Visualization System for Hi-C
891 Data. *Cell Systems* **6**, doi:10.1016/j.cels.2018.01.001 (2018).

892 77 Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-
893 320.e324, doi:10.1016/j.cell.2017.09.026 (2017).

894 78 Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M.
895 BUSCO: assessing genome assembly and annotation completeness with single-copy
896 orthologs. *Bioinformatics* **31**, 3210-3212, doi:10.1093/bioinformatics/btv351 (2015).

897 79 Deng, Y. *et al.* A telomere-to-telomere gap-free reference genome of watermelon and
898 its mutation library provide important resources for gene discovery and breeding. *Mol*
899 *Plant*, doi:10.1016/j.molp.2022.06.010 (2022).

900 80 Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic*
901 *Acids Res* **27**, 573-580, doi:10.1093/nar/27.2.573 (1999).

902 81 Robinson, J. T., Thorvaldsdottir, H., Turner, D. & Mesirov, J. P. igv.js: an embeddable
903 JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics*
904 **39**, doi:10.1093/bioinformatics/btac830 (2022).

905 82 Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome
906 alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907-
907 915, doi:10.1038/s41587-019-0201-4 (2019).

908 83 Perte, M. *et al.* StringTie enables improved reconstruction of a transcriptome from
909 RNA-seq reads. *Nat Biotechnol* **33**, 290-295, doi:10.1038/nbt.3122 (2015).

910 84 Consortium, T. U. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic*
911 *Acids Research* **51**, D523-D531, doi:10.1093/nar/gkac1052 (2022).

912 85 Campbell, M. S. *et al.* MAKER-P: A Tool Kit for the Rapid Creation, Management,
913 and Quality Control of Plant Genome Annotations *Plant Physiology* **164**, 513-524,
914 doi:10.1104/pp.113.230144 (2013).

915 86 Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59, doi:10.1186/1471-
916 2105-5-59 (2004).

917 87 Haas, B. J., Zeng, Q., Pearson, M. D., Cuomo, C. A. & Wortman, J. R. Approaches to
918 Fungal Genome Annotation. *Mycology* **2**, 118-141,
919 doi:10.1080/21501203.2011.606851 (2011).

920 88 Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically
921 mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637-
922 644, doi:10.1093/bioinformatics/btn013 (2008).

923 89 Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**,
924 D412-d419, doi:10.1093/nar/gkaa913 (2021).

925 90 Ou, S. *et al.* Benchmarking transposable element annotation methods for creation of
926 a streamlined, comprehensive pipeline. *Genome Biol* **20**, 275, doi:10.1186/s13059-
927 019-1905-y (2019).

928 91 Saha, S., Bridges, S., Magbanua, Z. V. & Peterson, D. G. Empirical comparison of ab
929 initio repeat finding programs. *Nucleic Acids Res* **36**, 2284-2294,

doi:10.1093/nar/gkn064 (2008).

Li, P. *et al.* RGAugury: a pipeline for genome-wide prediction of resistance gene
analogs (RGAs) in plants. *BMC Genomics* **17**, 852, doi:10.1186/s12864-016-3197-x
(2016).

Smith, M., Jones, J. & Hein, I. (bioRxiv, 2024).

Chen, Z. *et al.* iFeatureOmega: an integrative platform for engineering, visualization
and analysis of features from molecular sequences, structural and ligand data sets.
Nucleic Acids Research **50**, W434-W447, doi:10.1093/nar/gkac351 (2022).

Moon, K. R. *et al.* Visualizing structure and transitions in high-dimensional biological
data. *Nature Biotechnology* **37**, 1482-1492, doi:10.1038/s41587-019-0336-3 (2019).

Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**,
3094-3100, doi:10.1093/bioinformatics/bty191 (2018).

Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

Goel, M., Sun, H., Jiao, W. B. & Schneeberger, K. SyRI: finding genomic
rearrangements and local sequence differences from whole-genome assemblies.
Genome Biol **20**, 277, doi:10.1186/s13059-019-1911-0 (2019).

Goel, M. & Schneeberger, K. plotsr: visualizing structural similarities and
rearrangements between multiple genomes. *Bioinformatics* **38**, 2922-2926,
doi:10.1093/bioinformatics/btac196 (2022).

Marçais, G. *et al.* MUMmer4: A fast and versatile genome alignment system. *PLoS*
Comput Biol **14**, e1005944, doi:10.1371/journal.pcbi.1005944 (2018).

- 952 101 Heller, D. & Vingron, M. SVIM-asm: structural variant detection from haploid and
953 diploid genome assemblies. *Bioinformatics* **36**, 5519-5521,
954 doi:10.1093/bioinformatics/btaa1034 (2020).
- 955 102 Jeffares, D. C. *et al.* Transient structural variations have strong effects on quantitative
956 traits and reproductive isolation in fission yeast. *Nat Commun* **8**, 14061,
957 doi:10.1038/ncomms14061 (2017).
- 958 103 Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics
959 Analysis Version 11. *Molecular Biology and Evolution* **38**, 3022-3027,
960 doi:10.1093/molbev/msab120 (2021).
- 961 104 Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: Reconstruction, Analysis, and
962 Visualization of Phylogenomic Data. *Molecular Biology and Evolution* **33**, 1635-1638,
963 doi:10.1093/molbev/msw046 (2016).
- 964 105 Wingett, S. W. & Andrews, S. FastQ Screen: A tool for multi-genome mapping and
965 quality control. *F1000Res* **7**, 1338, doi:10.12688/f1000research.15931.2 (2018).
- 966 106 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program
967 for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930,
968 doi:10.1093/bioinformatics/btt656 (2014).
- 969 107 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for
970 differential expression analysis of digital gene expression data. *Bioinformatics* **26**,
971 139-140, doi:10.1093/bioinformatics/btp616 (2010).
- 972 108 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler
973 transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).

974 109 Brouard, J. S. & Bissonnette, N. Variant Calling from RNA-seq Data Using the GATK
975 Joint Genotyping Workflow. *Methods Mol Biol* **2493**, 205-233, doi:10.1007/978-1-
976 0716-2293-3_13 (2022).

977 110 Jiang, Y. *et al.* GTX.Digest.VCF: an online NGS data interpretation system based on
978 intelligent gene ranking and large-scale text mining. *BMC Med Genomics* **12**, 193,
979 doi:10.1186/s12920-019-0637-x (2019).

980 111 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-
981 based linkage analyses. *Am J Hum Genet* **81**, 559-575, doi:10.1086/519795 (2007).

982 112 Rastas, P. Lep-MAP3: robust linkage mapping even for low-coverage whole genome
983 sequencing data. *Bioinformatics* **33**, 3726-3732, doi:10.1093/bioinformatics/btx494
984 (2017).

985 113 Broman, K. W. *et al.* R/qt12: Software for Mapping Quantitative Trait Loci with High-
986 Dimensional Data and Multiparent Populations. *Genetics* **211**, 495-502,
987 doi:10.1534/genetics.118.301595 (2019).
988