

Genome assembly of the hybrid grapevine *Vitis* ‘Chambourcin’

Sagar Patel^{1,2,3,*}, Zachary N. Harris^{1,2}, Jason P. Londo⁴, Allison Miller^{1,2,*} and Anne Fennell⁵

¹ Saint Louis University, Department of Biology, 3507 Laclede Ave, St. Louis, MO 63103, USA

² Donald Danforth Plant Science Center, 975 North Warson Road, St. Louis, MO 63132, USA

³ Eastern Virginia Medical School, School of Health Professions, Norfolk, VA 23501, USA

⁴ School of Integrative Plant Science, Cornell University, 630 W. North Street, Geneva, NY 14456, USA

⁵ South Dakota State University, Agronomy, Horticulture and Plant Science Department and BioSNTR, Brookings, SD 57006, USA

ABSTRACT

‘Chambourcin’ is a French-American interspecific hybrid grape grown in the eastern and midwestern United States and used for making wine. Few genomic resources are available for hybrid grapevines like ‘Chambourcin’. Here, we assembled the genome of ‘Chambourcin’ using PacBio HiFi long-read, Bionano optical map, and Illumina short-read sequencing technologies. We generated an assembly for ‘Chambourcin’ with 26 scaffolds, with an N50 length of 23.3 Mb and an estimated BUSCO completeness of 97.9%. We predicted 33,791 gene models and identified 16,056 common orthologs between ‘Chambourcin’, *V. vinifera* ‘PN40024’ 12X.v2, VCOST.v3, Shine Muscat and *V. riparia* Gloire. We found 1,606 plant transcription factors from 58 gene families. Finally, we identified 304,571 simple sequence repeats (up to six base pairs long). Our work provides the genome assembly, annotation and the protein and coding sequences of ‘Chambourcin’. Our genome assembly is a valuable resource for genome comparisons, functional genomic analyses and genome-assisted breeding research.

Subjects Genetics and Genomics, Bioinformatics, Plant Genetics

Submitted: 18 January 2023

Accepted: 28 June 2023

Published: 03 July 2023

* Corresponding authors. E-mail: sgr308@gmail.com; allison.j.miller@slu.edu

Published by GigaScience Press.

Preprint submitted at <https://doi.org/10.1101/2023.01.18.524616>

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

Gigabyte, 2023, 1–18

BACKGROUND AND CONTEXT

Grapevines (*Vitis* species) represent the world’s most economically important berry-producing plants. Their fruits are used to make wine and other beverages, and are consumed as fresh or dried fruit. The European grapevine *Vitis vinifera* L. species *vinifera* is believed to have been domesticated approximately 8,000 years ago from wild populations of *V. vinifera* subspecies *sylvestris* growing in western Asia and Eastern Europe [1–3]. Grapevine growing (viticulture) spread rapidly through Europe and the Middle East, and eventually was introduced into North America as early as the mid 1700’s and likely earlier [4]. In addition to the introduced *V. vinifera*, North America is home to at least 20 native *Vitis* species. Although European settlers in North America cultivated native North American *Vitis* species, few native North American grapevine species are used to make wine (e.g., *Vitis labrusca*). Despite this, many native North American *Vitis* species have become critical resources for viticulture through their use in breeding programs aimed at developing disease-resistant rootstocks and hybrid scions derived from interspecific hybridizations between wild North American *Vitis* species and cultivated European *V. vinifera*. Hybrid derivatives of crosses between North American and European grapevine

species make up a large portion of the grapevines grown in eastern and midwestern North America, and hybrid rootstocks are used throughout most grape growing regions in the world.

Vitis ‘Chambourcin’ (‘Chambourcin’ from here forward) is a cultivated hybrid wine grape variety derived from crosses between North American and European *Vitis* species (NCBI:txid241073). ‘Chambourcin’ was developed by the private breeder Joannes Seyve in France. In 1985, it was introduced into the Geneva, New York, USA repository of the United States Department of Agriculture (USDA) Agricultural Research Service. A complex hybrid, ‘Chambourcin’ is the product of a cross between Joannes Seyve 11369 and ‘Plantet N’, which includes several North American *Vitis* species in its background: *V. berlandieri* Planch., *V. labrusca* L., *V. lincecumii* Buckley, *V. riparia* Michx., *V. rupestris* Scheele, and *V. vinifera*. The full pedigree of ‘Chambourcin’ is publicly available [5]. ‘Chambourcin’ produces black-skinned berries. The flavors of wines derived from ‘Chambourcin’ are described as black cherry, red fruit with herbaceous notes, black pepper, and chocolate [6]. ‘Chambourcin’ is grown in parts of France and Australia, as well as in the United States in Colorado, Missouri, Nebraska, New Jersey, New York, Pennsylvania, and Virginia, among others.

‘Chambourcin’ is increasing in importance as a cultivated hybrid wine grape in the central and eastern United States. It has been used in experimental rootstock vineyards to understand rootstock effects on shoot system phenotypes [7–11], and is the parent of the new disease-resistant cultivar ‘Regent’. The goals of this study were (1) to develop a high-quality reference genome for ‘Chambourcin’, and (2) to identify and annotate gene models for a more accurate functional genomic analysis of this disease-resistant cultivar. The work presented here advances our understanding of hybrid grapevine genomics and will facilitate the analyses of rootstock-scion interactions in ‘Chambourcin’ experimental vineyards.

METHODS

PacBio HiFi, Bionano optical map, and Illumina sequencing

‘Chambourcin’ leaf material was obtained from a 12-year-old experimental vineyard located at the University of Missouri Southwest Research Station in Mount Vernon, Missouri, USA. For PacBio HiFi sequencing, high molecular weight (HMW) DNA was isolated using the Nucleobond Kit (Macherey-Nagel, Bethlehem, PA, USA) following the manufacturer’s protocol. Approximately 20 cg DNA was sheared to a center of mass of 10–20 kilobase (kb) in a Megaruptor 3 system. Next, a HiFi sequencing library was constructed following HiFi SMRTbell protocols for the Express Template Prep Kit 2.0 according to manufacturers’ recommendations (Pacific Biosciences, California). The library was sequenced using Sequel binding and sequencing chemistry v2.0 in circular consensus sequencing (CCS) mode in a Sequel II system with a movie collection (file format of HiFi data) time of 30 h. The HiFi reads were generated with the CCS mode of pbbtools [12] using a minimum Predicted Accuracy of 0.990.

For the Bionano data, DNA was isolated from fresh young leaf tissue from a 12-year-old experimental vineyard located at the University of Missouri Southwest Research Station in Mount Vernon, Missouri, USA, using the Prep™ Plant DNA Isolation kit and labeled using the Bionano Prep™ DNA Labeling Kit Direct Label and Stain (DLS) (Bionano Genomics, San Diego, California). In total, 500 ng of ultra-high molecular weight (UHMW) DNA was used

for the DLS reaction. DNA was incubated in the presence of DLE-1 Enzyme, DL-Green, and DLE-1 Buffer for 3:20 h at 37 °C. This was followed by proteinase K digestion at 50 °C for 30 min and double cleanup of the unincorporated DL-Green label. The resulting DLS sample was combined with the Flow Buffer, dithiothreitol (DTT) and DNA stain, mixed at slow speed in a rotator mixer for an hour, and then incubated overnight at 4 °C. The labeled sample was then loaded onto a Bionano flow cell in a Saphyr System for separation, imaging, and creation of digital molecules according to the manufacturer's recommendations [13]. The raw molecule set was filtered to a molecule length of 250 kb and a minimum of nine CTTAAG labels per molecule. Bionano maps were assembled without pre-assembly using the non-haplotype parameters with no Complex Multi-Path Regions (CMPR) cut and without extend-split. Bionano software (Solve, Tools and Access, v1.5.1) [14] was used for data visualization, processing, and assembly of Bionano maps. The PacBio HiFi and Bionano sequencing were done at Corteva Agriscience, Johnston, Iowa, USA.

For the Illumina whole genome data, DNA was extracted from 'Chambourcin' leaf tissue collected from the USDA Grape Germplasm Collection located in Geneva, New York. DNA was extracted using Qiagen DNeasy Plant Mini Kits (Qiagen, Valencia, California, USA) and assessed for purity and concentration using a NanoDrop spectrophotometer and Qubit fluorometer. DNA was cleaned using the Qiagen Dneasy PowerClean Pro Cleanup Kit. DNA libraries were prepared and shotgun Illumina sequenced at Novogene (San Diego, California, USA) with paired-end 150 nt reads with 40X coverage. The raw Illumina reads were trimmed with Trimmomatic (v0.39; RRID:SCR_011848) [15] using HEADCROP:4 MINLEN:70 parameters.

Genome size estimation

The PacBio HiFi reads and 19 nt k-mers were used to estimate the genome heterozygosity using jellyfish (v2.3.0; RRID:SCR_005491) [16]. The resulting ".histo" file was visualized with GenomeScope (RRID:SCR_017014) [17].

Genome assembly

The PacBio HiFi assembly was generated using the Hifiasm assembler (RRID:SCR_021069; v0.13-r308) [18] with default parameters. To reduce the number of small, low-coverage artifactual contigs often generated by Hifiasm [18], the assembly was filtered to exclude less than 70,000 bp contigs. The resulting HiFi contigs were merged to the DLS Bionano maps with Bionano Solve (v3.5.1) [14] using the hybridscaffold.pl script of Bionano Solve (v3.5.1) [14] to get a hybrid assembly. Each scaffold of the hybrid assembly was then checked, and small overlapping contigs were curated and removed to make a contiguous sequence. This curated diploid assembly was examined to identify alternative contigs using Purge_Haplotigs (v1.1.1; RRID:SCR_017616) [19], and the primary assembly and haplotig assemblies were created. We mapped trimmed Illumina whole genome sequences to both assemblies separately with bowtie2 (v2.3.4; RRID:SCR_016368) [20] and samtools (v1.9; RRID:SCR_002105) [21]. The resulting .bam files were used for polishing both assemblies using Pilon (v1.23; RRID:SCR_014731) [22] with one round, and the final assembly (primary assembly) and haplotig assemblies were prepared. In this study, we used only the primary assembly for all downstream analysis, but the haplotigs are maintained to cover the total heterozygous genome. Scaffolds were aligned to the *V. vinifera* 'PN40024' 12X.v2 [23]

reference genome using minimap2 (v2.17; RRID:SCR_018550) [24] and renamed based on the longest alignment with the reference genome *V. vinifera* ‘PN40024’ 12X.v2 chromosomes. We mapped two thousand ‘Chambourcin’ rhAmpSeq marker sequences [25] to the ‘Chambourcin’ genome assembly using the BWA aligner (v0.7.17; RRID:SCR_010910) [26]. The rhAmpSeq markers were designed to target the core *Vitis* genome and were developed from gene-rich collinear regions of 10 *Vitis* genomes [25]. These markers aid in mapping contigs on chromosomes and checking their orientation.

Genome assembly assessment and dot plot

All assemblies generated by PacBio HiFi and Bionano data were assessed by Benchmarking Universal Single-Copy Orthologs (BUSCO) (v5.4.2; RRID:SCR_015008) [27] with genome mode and the embryophyta_odb10 dataset. The alignment of the two genomes was obtained using minimap2 (v2.17) [24] with default parameters, where the ‘Chambourcin’ primary assembly was considered as query while *V. vinifera* ‘PN40024’ 12X.v2, Shine Muscat [28], and *V. riparia* Gloire [29] were considered as the reference genome. A dot plot was obtained using the R (RRID:SCR_001905) script pafCoordsDotPlotly.R [30].

Genome assembly analysis using k-mer spectra

The trimmed Illumina whole genome sequences were mapped separately to diploid, primary, and haplotig ‘Chambourcin’ assemblies using KAT (v2.4.2; RRID:SCR_016741) [31]; specifically, we used kat comp and kat plot commands.

De novo gene prediction, functional annotation, and orthologous genes

De novo repeats were identified with RepeatModeler2 (v2.0.2a) [32], and repeats were masked by RepeatMasker (v4.1.1; RRID:SCR_012954) [33]. ‘Chambourcin’ RNA-seq data were downloaded from a previously published study [7] and trimmed using Trimmomatic (v0.39) [15] with HEADCROP:15 LEADING:30 TRAILING:30 MINLEN:20 parameters. The trimmed ‘Chambourcin’ RNA-seq reads were then mapped to the masked ‘Chambourcin’ primary genome assembly using HISAT2 (v2.1.0; RRID:SCR_015530) [34] and samtools (v1.9) [21] with default parameters. The resulting alignments (.bam files) and protein sequences of the *V. vinifera* ‘PN40024’ 12X.v2, VCost.v3 were used for gene prediction using BRAKER2 (v2.1.6; RRID:SCR_018964) [35] with `-prg=gth -gth2traingenes -gff3` parameters. The resulting gene predictions (proteins, coding sequences, and annotations) were completed separately for the ‘Chambourcin’ primary assembly and the ‘Chambourcin’ haplotig assembly. The quality of the predicted proteins was assessed using BUSCO (v5.4.2) [27] with protein mode and the embryophyta_odb10 dataset. The predicted proteins of the *Vitis* ‘Chambourcin’ primary assembly were then functionally annotated using eggNOG-mapper (v2) (RRID:SCR_021165) [36] and related to Gene Ontology (GO), KEGG pathway, and other functional information. The GO plot was developed using the WEGO tool [37]. For the analysis of orthologous gene models, the sequences of ‘Chambourcin’ primary gene models, *V. vinifera* PN40024 12X.v2, VCost.v3, Shine Muscat, and *V. riparia* Gloire were analyzed using OrthoVenn2 [38] with default settings, *E*-value: 1×10^{-5} , and inflation value: 1.5.

Plant transcription factors prediction, phylogenetic tree, and WRKY classification

The plant transcription factors for the gene models of the ‘Chambourcin’ primary assembly, *V. vinifera* PN40024 12X.v2, and VCost.v3 were identified using the Plant Transcription Factor Database (PlantTFDB v5.0; RRID:SCR_003362) [39]. The identified transcription factors were divided into subfamilies according to their sequence relationship with *V. vinifera*. For the circular phylogenetic tree and WRKY sequences of ‘Chambourcin’ primary gene models and *V. vinifera* PN40024 12X.v2, VCost.v3 gene models retrieved from PlantTFDB (5.0) [39] and aligned using ClustalW method in MEGA7 [40]. A phylogenetic analysis was carried out using the neighbor-joining method with 1,000 bootstrap replications, and the evolutionary distances were computed using the Poisson correction method with the Pairwise Deletion option. The WRKY classification of ‘Chambourcin’ primary gene models was carried out using the same method described in a previous study [41].

Synteny, Simple Sequence Repeats (SSRs), and Circos plot

The ‘Chambourcin’ masked primary genome assembly and gene annotations were aligned to *V. vinifera* ‘PN40024’ 12X.v2, Shine Muscat, and *V. riparia* ‘Gloire’ genomes and gene annotations separately using the ‘promer’ option of the MUMmer program in SyMAP (v4.2) [42]. We used MicroSatellite [43] to find SSRs in the unmasked ‘Chambourcin’ primary genome assembly. The Circos plots were generated using circos (v0.69.6; RRID:SCR_011798) [44] with the ‘Chambourcin’ primary genome assembly, SSRs, and the ‘Chambourcin’ primary gene annotations.

Mapping of Illumina whole genome reads, and RNA-seq reads to the genome assembly

The trimmed Illumina whole genome sequences were mapped to the diploid, primary, and haplotig ‘Chambourcin’ assemblies separately with bowtie2 (v2.3.4) [20] and samtools (v1.9) [21]. The resulting .bam files were used to obtain mapping results using the samtools flagstat [21] command. We also mapped the trimmed ‘Chambourcin’ RNA-seq reads [7] to diploid, primary, and haplotig ‘Chambourcin’ assemblies separately using HISAT2 (v2.1.0) [34] and samtools (v1.9) [21]. The resulting .bam files were used to obtain mapping results using the samtools flagstat [21] command.

RESULTS AND DISCUSSION

Genome sequencing and assembly of ‘Chambourcin’

We generated a high-quality and contiguous genome sequence of ‘Chambourcin’ using PacBio HiFi Sequencing, Bionano third-generation DNA sequencing, and Illumina short-read sequencing. A total of 1,634,814 PacBio HiFi filtered reads was produced with an average length of 16,148 bp and genome coverage of 28X. The filtered Bionano data resulted in a subset of 1,243,428 molecules with a total length of 429,808.857 Mbp and coverage of 188.70X. In total, 124 Bionano maps, with a total length of 962.964 Mbp and an N50 of 13,725 bp, were assembled, corresponding to the diploid complement. A total of 154,152,068 filtered Illumina short reads and genome coverage of 40X were generated for genome polishing. We estimated heterozygosity to be 2.28% in the ‘Chambourcin’ genome (Figure 1), which is higher than estimates for heterozygosity in any of the other *Vitis* genomes

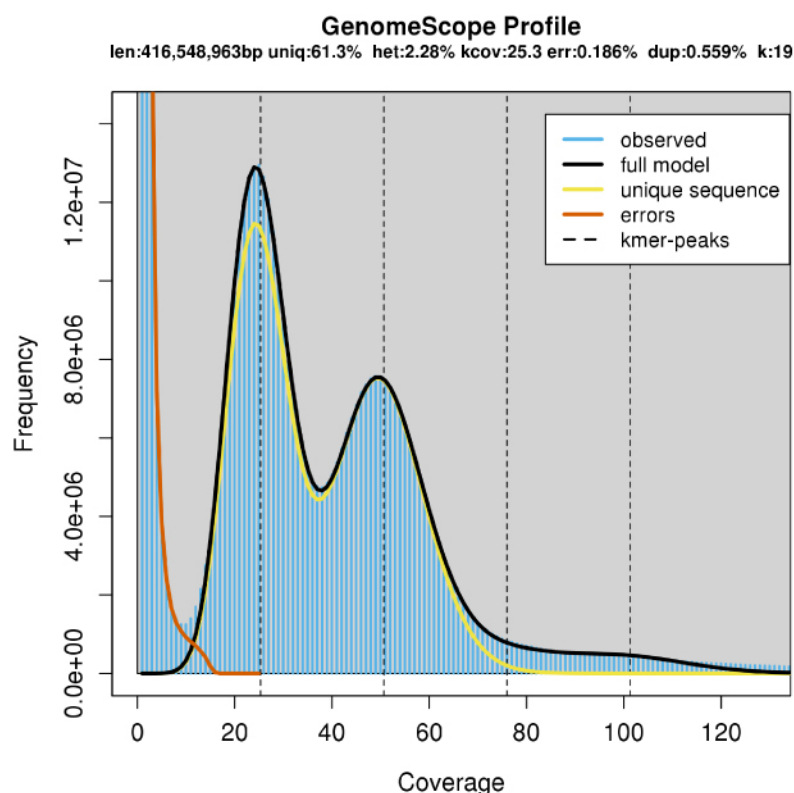


Figure 1. GenomeScope plot estimating the heterozygosity of *Vitis* ‘Chambourcin’.

sequenced to date [23, 28, 29]. Relatively higher levels of heterozygosity in the ‘Chambourcin’ genome compared to other *Vitis* species are expected, given the complex interspecific pedigree of this cultivar. A GenomeScope plot of clean reads demonstrated two peaks of coverage; the first peak located at 25X coverage corresponds to the heterozygous portion of the genome, and the second peak at 52X coverage corresponds to the homozygous portion of the genome (Figure 1).

A *de novo* ‘Chambourcin’ genome was assembled using HiFi, Bionano, and Illumina data. First, a contig assembly of the PacBio HiFi reads resolved the reads into 196 contigs with an N50 of 12,215,205 bp and a total length of 949,347,381 bp (Table 1). The PacBio HiFi contig assembly was then merged with the Bionano maps to get an initial hybrid assembly comprising 67 scaffolds with an N50 length of 16,400,326 bp, a maximum scaffold length of 39,458,994 bp, and a total scaffold length of 903,810,753 bp (Table 1). After manual curation, the hybrid assembly included 64 scaffolds with an N50 length of 16,278,793 bp, a maximum length of 39,458,994 bp, and a total length of 869,222,201 bp (Table 1). The hybrid assembly was partitioned into a final primary assembly (493,554,689 bp) and a haplotig assembly (375,458,233 bp) (Table 1). Pilon (v1.23) corrected 19,771 single nucleotide polymorphisms (SNPs), 636 ambiguous bases, 16,469 small insertions totaling 121,315 bases, and 21,889 small deletions totaling 129,590 bases in the primary assembly; Pilon (v1.23) also corrected 21,394 SNPs, 606 ambiguous bases, 14,999 small insertions totaling 102,126 bases, and 18,929 small deletions totaling 105,957 bases in the haplotig assembly. After polishing for ‘Chambourcin’, the final primary assembly contained 26 scaffolds with an N50 length of

Table 1. Descriptive statistics and BUSCO results for the ‘Chambourcin’ genome assembly.

Details	PacBio HiFi assembly	Hybrid assembly (PacBio HiFi + Bionano)	Hybrid assembly (curated)	Final primary assembly (after purge haplotigs)	Final haplotig assembly
Genome assembly results					
Number of scaffolds	196	67	64	26	38
Total size of scaffolds	949,347,381	903,810,753	869,222,201	493,554,689	375,458,233
Longest scaffold	47,120,234	39,458,994	39,458,994	39,456,434	28,439,729
Shortest scaffold	70,924	1,571,397	1,571,397	5,519,669	1,571,459
Number of scaffolds > 1M nt	131	67	64	26	38
Number of scaffolds > 10M nt	27	40	38	22	16
N50 scaffold length	12,215,205	16,400,326	16,278,793	23,325,629	12,462,019
scaffold %N	0	0.83	0.76	0.02	1.71
Number of contigs	196	1,365	1,194	44	136
Total size of contigs	949,347,381	896,337,746	862,647,977	493,450,270	369,041,914
Longest contigs	47,120,234	32,988,011	32,988,011	32,983,396	22,801,634
Shortest contigs	70,924	6	6	6	6
Number of contigs > 1M nt	131	100	95	36	59
Number of contigs > 10M nt	27	33	31	22	9
N50 contigs length	12,215,205	14,022,606	13,473,621	16,713,841	7,864,215
contig %N	0	0	0	0	0
BUSCO results					
Complete BUSCOs (C) = (S) + (D)	1,593 (98.7%)	1,594 (98.7%)	1,592 (98.6%)	1,580 (97.9%)	1,180 (73.1%)
Complete and single-copy BUSCOs (S)	121 (7.5%)	367 (22.7%)	480 (29.7%)	1,546 (95.8%)	1,140 (70.6%)
Complete and duplicated BUSCOs (D)	1,472 (91.2%)	1,227 (76%)	1,112 (68.9%)	34 (2.1%)	40 (2.5%)
Fragmented BUSCOs (F)	13 (0.8%)	13 (0.8%)	14 (0.9%)	17 (1.1%)	23 (1.4%)
Missing BUSCOs (M)	8 (0.5%)	7 (0.5%)	8 (0.5%)	17 (1%)	411 (25.5%)
Total BUSCOs	1,614	1,614	1,614	1,614	1,614

23,325,629 bp, and the longest scaffold measured 39,456,434 bp (Table 1). The secondary haplotig assembly after polishing contained 38 haplotig scaffolds with an N50 length of 1,2462,019 bp, and the longest scaffold measured 28,439,729 bp (Table 1). We identified 97.9% complete BUSCOs for primary genome assembly and 73.1% complete BUSCOs for haplotig genome assembly (Table 1).

The ‘Chambourcin’ primary genome assembly was aligned to the reference genomes *V. vinifera* ‘PN40024’ 12X.v2 [23] (see Table 1 in GigaDB [45]), Shine Muscat [28], and *V. riparia* ‘Gloire’ [29]. A dot plot was generated to facilitate the comparisons among genomes. Collinearity between ‘Chambourcin’ and *V. vinifera* ‘PN40024’ 12X.v2, Shine Muscat, and *V. riparia* ‘Gloire’ was observed as a straight diagonal line without large gaps in the dot plot, confirming the high synteny of the ‘Chambourcin’ genome with *V. vinifera* ‘PN40024’ 12X.v2 (Figure 2A), Shine Muscat (Figure 2B), and *V. riparia* ‘Gloire’ (Figure 2C). To further validate the ‘Chambourcin’ genome assembly, we mapped ‘Chambourcin’ rhAmpSeq markers [25] to the ‘Chambourcin’ genome assembly. We found 99% of rhAmpSeq markers mapped to ‘Chambourcin’ scaffolds and mapped to the same chromosomes and positions the markers were derived from in the collinear *Vitis* core genome (see Table 2 in GigaDB [45]).

Synteny analyses of the ‘Chambourcin’ primary genome assembly with *V. vinifera* ‘PN40024’ 12X.v2, Shine Muscat, and *V. riparia* ‘Gloire’ genomes were used to identify syntenic blocks between species. The ‘Chambourcin’ primary assembly scaffolds aligned with larger syntenic blocks and covered the whole chromosomes of *V. vinifera* PN40024 12X.v2 (Figure 2D), Shine Muscat (Figure 2D), and *V. riparia* ‘Gloire’ (Figure 2E). This alignment of the primary genome with *V. vinifera* PN40024 12X.v2, Shine Muscat, and

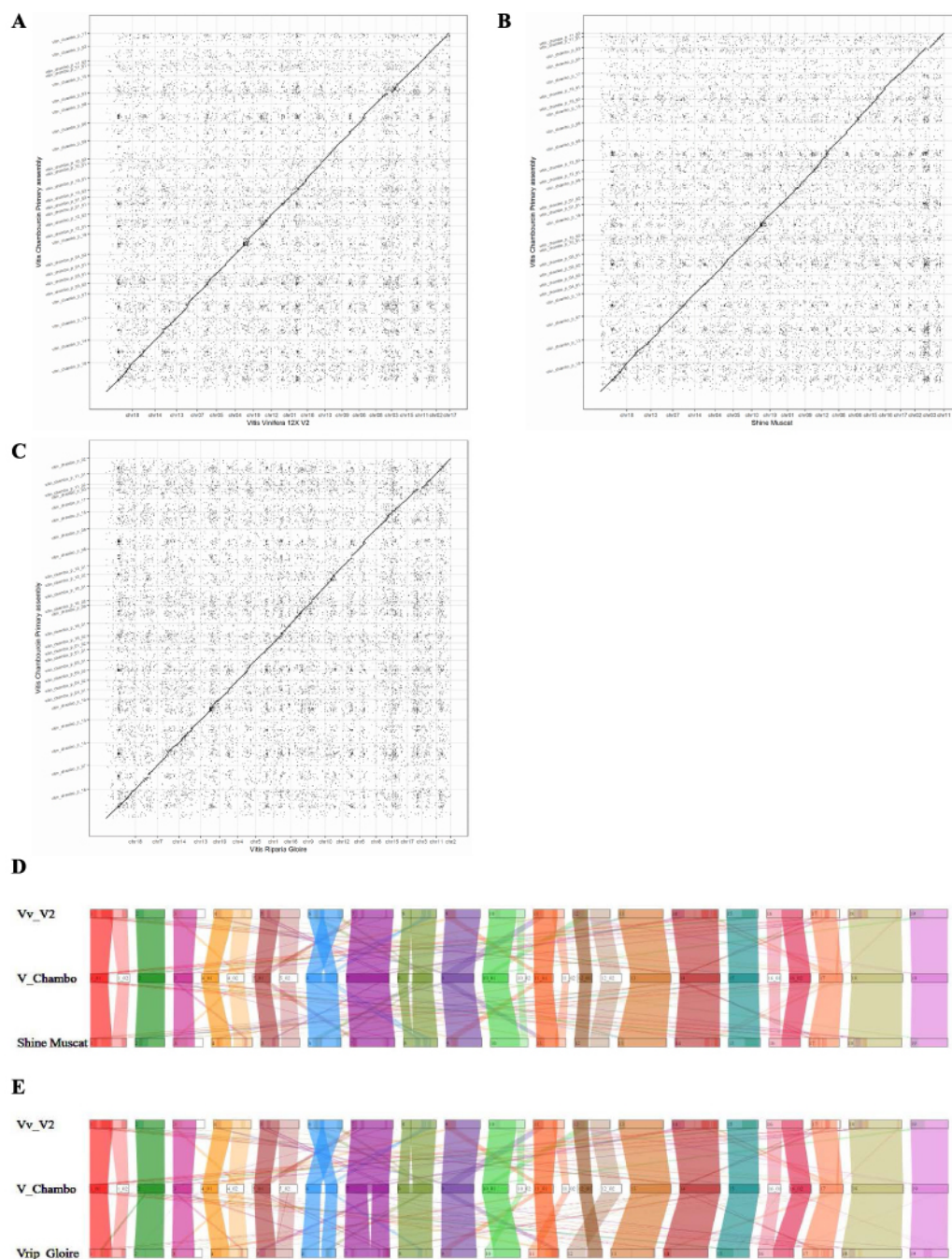


Figure 2. Comparative study of the 'Chambourcin' genome assembly. (A) Dotplot of the 'Chambourcin' primary genome assembly and *V. vinifera* 'PN40024' 12X.v2. (B) Dotplot of the 'Chambourcin' primary genome assembly and Shine Muscat. (C) Dotplot of the 'Chambourcin' primary genome assembly and *V. riparia* Gloire. (D) Synteny between 'Chambourcin' primary genome assembly, *V. vinifera* PN40024 12X.v2 genome, and Shine Muscat. (E) Synteny between 'Chambourcin' primary genome assembly, *V. vinifera* PN40024 12X.v2 genome, and *V. riparia* Gloire genome.

V. riparia ‘Gloire’ indicated highly contiguous ‘Chambourcin’ scaffolds useful for comparative genomic analyses.

Genome assembly analysis using K-mer spectra plot

The genome quality was assessed with Illumina whole genome reads separately for diploid, primary, and haplotig genome assemblies using KAT tool [31] and K-mer spectra plots were generated. A K-mer spectra is a graphical representation showing how many k-mers appear a certain number of times. The frequency of occurrence is plotted on the x-axis and the number of k-mers is plotted on the y-axis. All K-mer spectra plots for ‘Chambourcin’ diploid, primary and haplotig assemblies were identified with an error distribution under 10X, a heterozygous peak at 35X, and a homozygous peak at 67X (Figure 3A–C). The different colors in the K-mer spectra plot shows the different occurrences of k-mers. The black color represents read content occurs at zero time (0X), the red color represents unique content occurs at one time (1X), the purple color represents content occurs at two times (2X) and the green color represents content occurs at three times (3X). The K-mer spectra plot of the diploid genome assembly shows that the read content shown in black color is absent from the assembly, and red peak occurs once, showing most of the heterozygous content. At the same time, purple peak indicates more duplications on homozygous content (Figure 3A). The K-mer spectra plot of the primary genome is more collapsed, including mostly a single copy of the homozygous content and less of the heterozygous content (Figure 3B). The K-mer spectra plot for the haplotig genome assembly identified two black peaks representing read content in both the heterozygous and homozygous regions (Figure 3C). These K-mer spectra plots provides useful information for genome assembly assessment using whole genome short reads to identify duplicate regions in the assembly and visualize the genome assembly. This visualization is useful for genome assembly curation steps to identify accurate primary and haplotig assembly from a diploid genome assembly.

Repeat sequence annotation

Repeated regions were binned into seven different classes: long interspersed nuclear elements (LINEs) (4.43%), long terminal repeats (LTRs) (15.66%), DNA transposons (2.03%), rolling-circles (0.58%), low complexity repeats (0.37%), simple repeats (1.21%), and unclassified repeats (31.95%) (Figure 4; Table 2). The repetitive sequence content in the ‘Chambourcin’ primary genome assembly (56.23%) was higher than previously reported for *V. riparia* ‘Manitoba 37’ (46%) [41], *V. vinifera* ‘PN40024’ 12X.v2 (35.12%) [23], Shine Muscat (48%) [28], and *V. riparia* ‘Gloire’ (33.94%) [29]. SSRs are tandem repeats of DNA that have been used to develop robust genetic markers. We identified 304,571 SSRs, repeating units of 1–6 base pairs in length, in the ‘Chambourcin’ primary genome assembly (Figure 4; and see GigaDB Table 3 [45]).

Gene annotation and orthologous genes

A total of 33,791 gene models were predicted for the ‘Chambourcin’ primary genome assembly (Figure 4). We identified 94.6% complete BUSCOs (C); of these, 86.9% were designated single-copy BUSCOs (S), and 7.7% were designated duplicated BUSCOs (D) (Table 3). As evidenced by the high number of complete single-copy genes identified, the BUSCO results indicate that the ‘Chambourcin’ primary genome assembly offers

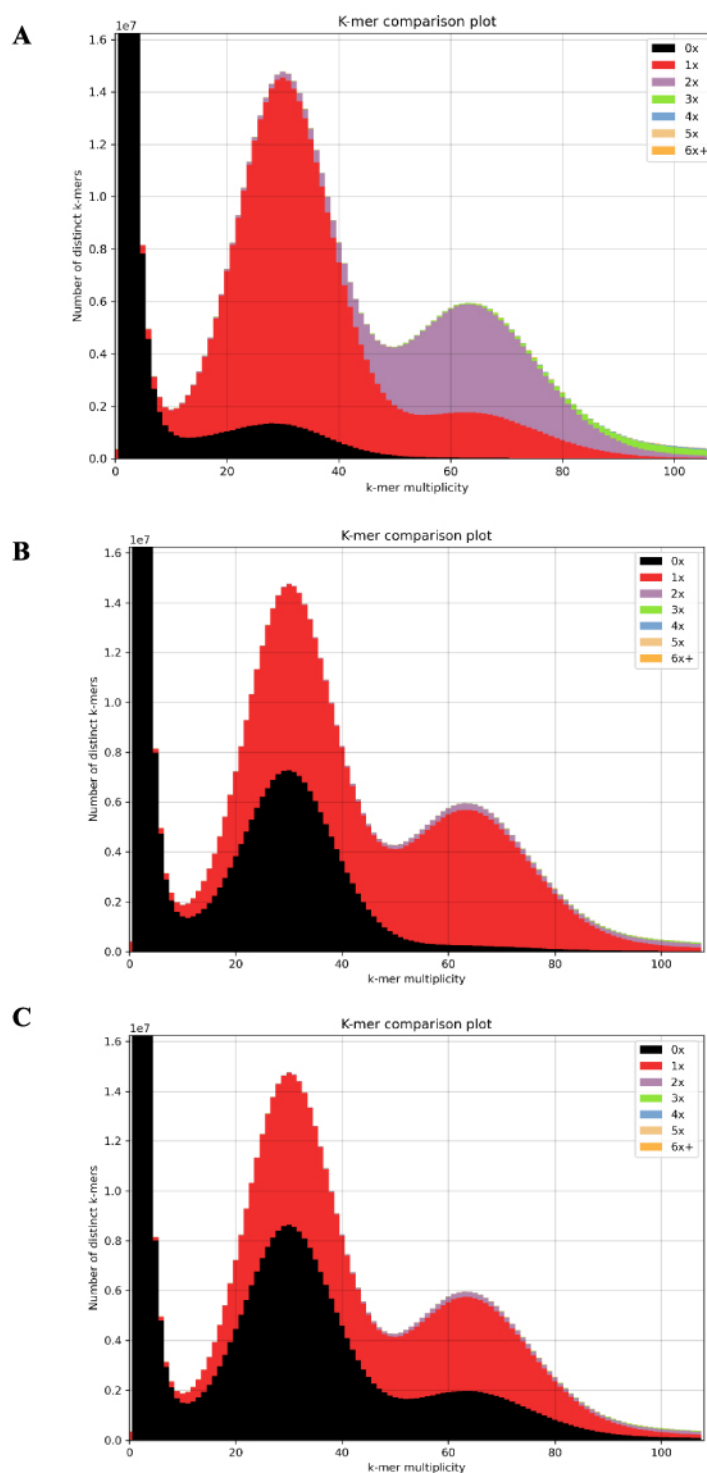


Figure 3. KAT k-mer spectra plot. (A) k-mer spectra plot for the diploid genome assembly. (B) k-mer spectra plot for the primary genome assembly. (C) k-mer spectra plot for the haplotig genome assembly.

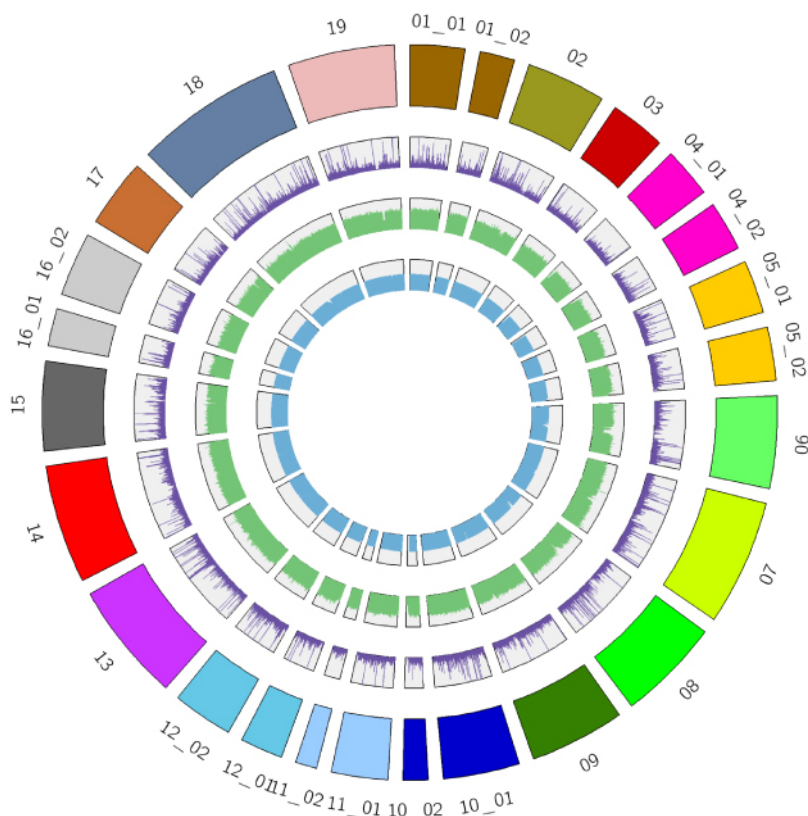


Figure 4. Circos plot. The outer ring represents all scaffolds of ‘Chambourcin’ primary genome assembly in different colors. The second ring (purple) represents the SSRs. The third ring (green) represents the repetitive sequences. The fourth ring (blue) represents the gene annotations.

Table 2. Repetitive sequences in the ‘Chambourcin’ genome assembly.

Details	Primary assembly	Haplotig assembly
Total length	493,554,689 bp	375,458,233 bp
Bases masked	277,503,542 bp (56.23%)	211,395,579 bp (56.30%)
Retroelements:	99,122,846 bp (20.08%)	82,725,835 bp (22.03%)
LINES:	21,852,466 bp (4.43%)	15,553,322 bp (4.14%)
RTE/Bov-B	180,809 bp (0.04%)	133,053 bp (0.04%)
L1/CIN4	21,671,657 bp (4.39%)	15,420,269 bp (4.11%)
LTR elements:	77,270,380 bp (15.66%)	67,172,513 bp (17.89%)
Ty1/Copia	39,408,750 bp (7.98%)	32,756,463 bp (8.72%)
Gypsy/DIRS1	33,756,474 bp (6.84%)	31,170,356 bp (8.30%)
DNA transposons:	9,995,420 bp (2.03%)	7,711,872 bp (2.05%)
hobo-Activator	3,826,157 bp (0.78%)	2,619,712 bp (0.70%)
Tourist/Harbinger	640,576 bp (0.13%)	454,161 bp (0.12%)
Rolling-circles	2,864,922 bp (0.58%)	2,386,306 bp (0.64%)
Unclassified	157,704,064 bp (31.95%)	112,960,798 bp (30.09%)
Total interspersed repeats	266,822,330 bp (54.06%)	203,398,505 bp (54.17%)
Simple repeats	5,980,823 bp (1.21%)	4,443,253 bp (1.18%)
Low complexity	1,835,467 bp (0.37%)	1,167,515 bp (0.31%)

comprehensive coverage of the expected gene space. Functional annotation of the ‘Chambourcin’ primary gene models (33,791) was done using the EggNOG database

Table 3. ‘Chambourcin’ gene prediction (Coding Sequences (CDS) and protein sequences) and BUSCO results of protein sequences.

Details	Primary assembly	Haplotig assembly
Gene prediction results		
Total CDS and protein	33,791	24,018
Total CDS (bp)	36,761,139	25,814,082
Mean CDS (bp)	1,087.9	1,074.8
Longest CDS (bp)	15,867	21,819
Total protein (bp)	12,219,929	8,580,681
Mean protein (bp)	361.6	357.3
Longest protein (bp)	5,288	7,272
BUSCO results		
Complete BUSCOs (C) = (S) + (D)	1,528 (94.6%)	1,136 (70.3%)
Complete and single-copy BUSCOs (S)	1,403 (86.9%)	1,032 (63.9%)
Complete and duplicated BUSCOs (D)	125 (7.7%)	104 (6.4%)
Fragmented BUSCOs (F)	50 (3.1%)	31 (1.9%)
Missing BUSCOs (M)	36 (2.3%)	447 (27.8%)
Total BUSCOs	1,614	1,614

(see GigaDB Table 4 [45]). A total of 27,075 ‘Chambourcin’ primary proteins were annotated, and 86% (22,977) of these proteins were annotated with *V. vinifera* V1 gene models (see GigaDB Table 4 [45]). Out of the 27,075 ‘Chambourcin’ annotated primary proteins, 13,311 gene models were identified with GO accessions and further classified into three sub-ontologies: biological process (11,399), cellular component (11,472), and molecular function (9,977) (Figure 5F) (GigaDB Table 4 [45]). A total of 8,460 ‘Chambourcin’ primary proteins were annotated with KEGG pathways (GigaDB Table 4 [45]). Using OrthoVenn2, we identified 16,056 common orthologs between ‘Chambourcin’ primary gene models, *V. vinifera* PN40024 12X.v2 annotation, Shine Muscat, and *V. riparia* ‘Gloire’ (Figure 5A). In total, 16,476 orthologous gene models were found between the ‘Chambourcin’, Shine Muscat, and *V. riparia* ‘Gloire’ (Figure 5B). Finally, 19,477 gene models were orthologous with *V. vinifera* PN40024 12X.v2 VCost.v3 proteins (Figure 5C), 18,669 gene models were orthologous with Shine Muscat (Figure 5D), and 18,183 gene models were orthologous with *V. riparia* ‘Gloire’ (Figure 5E).

Plant transcription factors and ‘Chambourcin’ WRKY transcription factor classification

Using PlantTFDB 5.0, 1,606 plant transcription factors representing 58 gene families were identified from ‘Chambourcin’ primary proteins (see GigaDB Table 5 [45]). A similar number of transcription factors was identified for the AP2, NAC, RAV, and WRKY gene families, as found in *V. vinifera* ‘PN40024’ 12X.v2, VCost.v3. We identified 65 WRKY sequences in ‘Chambourcin’ and 62 in *V. vinifera* PN40024 12X.v2, VCost.v3 (Figure 6) (Table 4; GigaDB Table 5 [45]). WRKY transcription factors regulate many processes in plants and algae, such as the responses to biotic and abiotic stresses and seed dormancy. The ‘Chambourcin’ WRKY subfamily classification was similar to *V. vinifera* ‘PN40024’ 12X.v2 and *V. riparia* ‘Manitoba 37’ (Table 4). These results show the high coverage of ‘Chambourcin’ primary proteins.

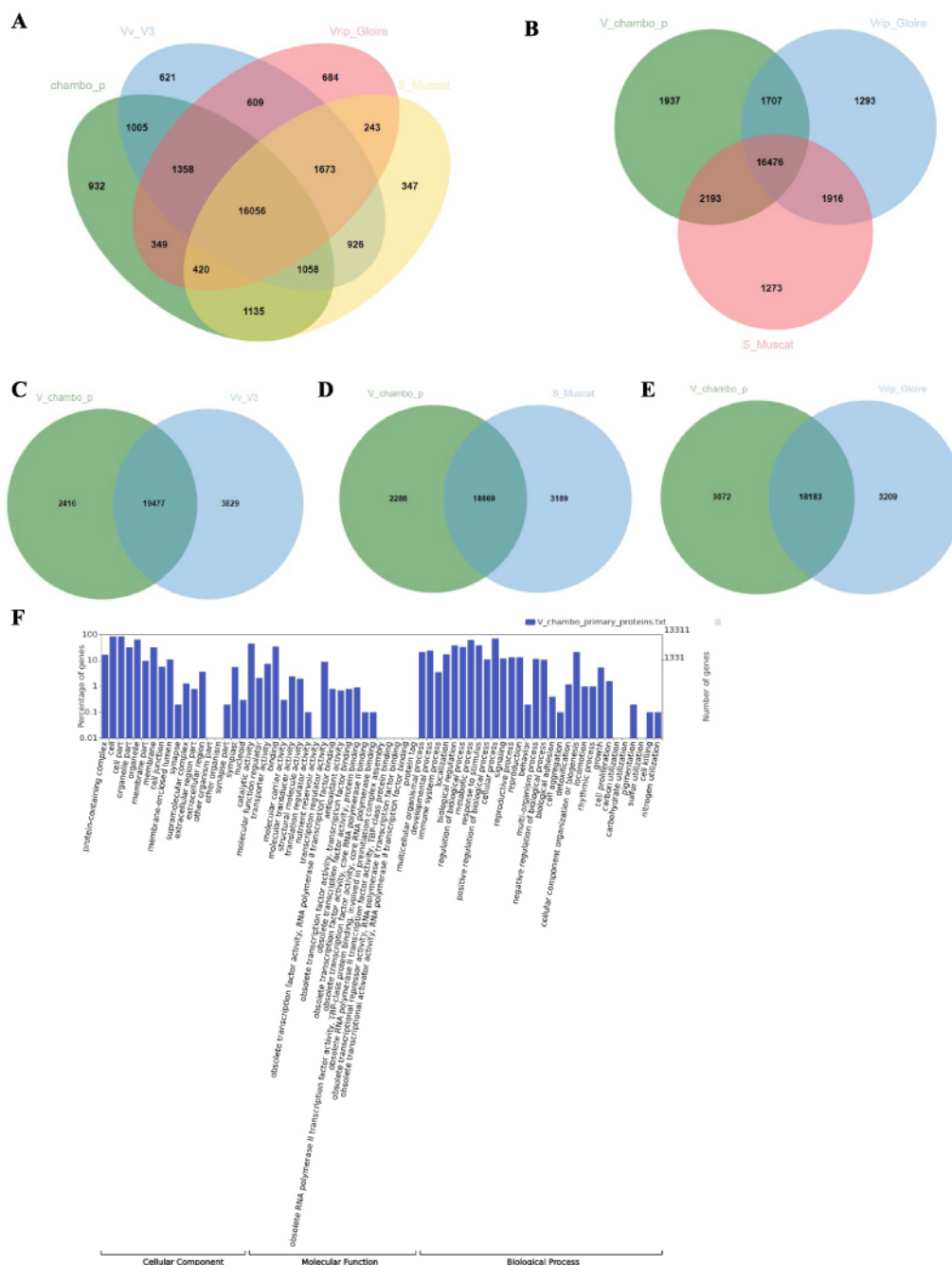


Figure 5. Venn diagram of 'Chambourcin' primary proteins with other grapevine species. (A) Venn diagram of orthologous genes in the 'Chambourcin' primary proteins, *V. vinifera* PN40024 12X.v2, VCost.v3, Shine Muscat, and *V. riparia* Gloire. (B) Orthologous genes in the 'Chambourcin' primary proteins, Shine Muscat, and *V. riparia* Gloire. (C) Orthologous genes in 'Chambourcin' primary proteins and *V. vinifera* PN40024 12X.V3 proteins. (D) Orthologous genes in 'Chambourcin' primary proteins and Shine Muscat. (E) Orthologous genes in 'Chambourcin' primary proteins and *V. riparia* Gloire species. (F) GO results for 'Chambourcin' primary proteins.

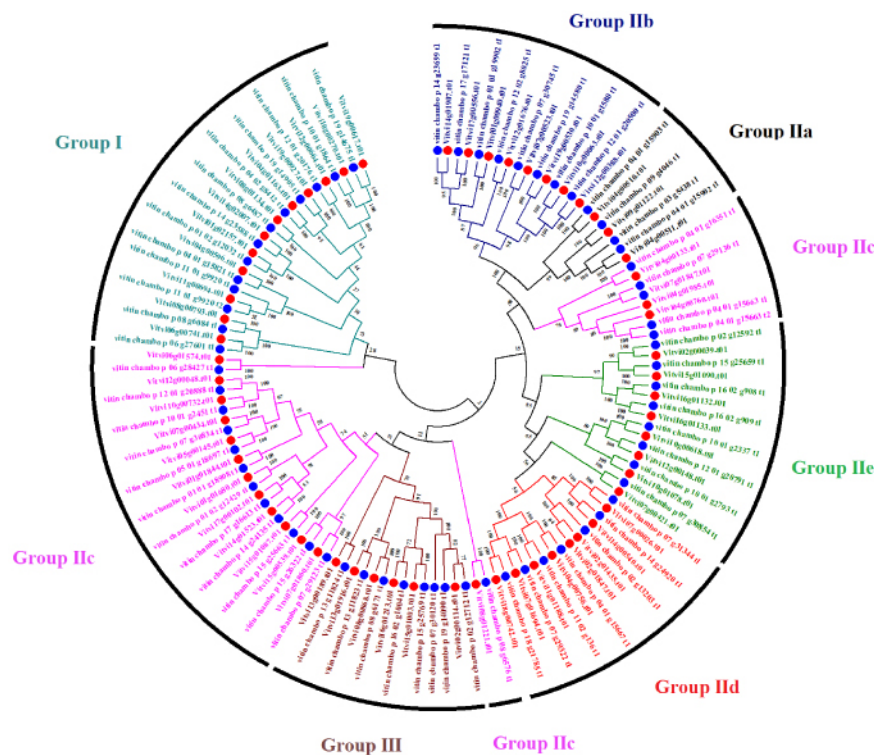


Figure 6. ‘Chambourcin’ and *V. vinifera* PN40024 12X.v2, VCost.v3 WRKY transcription factors. The blue dots represent ‘Chambourcin’ and the red dots represent *V. vinifera* PN40024 12X.v2, VCost.v3. The bootstrap values displayed are at the nodes.

Table 4. Comparison of the WRKY transcription factor classification for ‘Chambourcin’ with other grape species.

Species	Group I		Group II			Group III	Total
	IIa	IIb	IIc	IId	IIe		
Chambourcin	13	4	8	17	7	8	65
<i>V. vinifera</i> V3	12	3	8	17	8	8	62
<i>V. riparia</i> *	13	3	8	19	8	9	67

* (Patel et al., 2020) [41].

Mapping of Illumina whole genome reads, and RNA-seq reads to the genome assembly

We aligned trimmed Illumina whole genome reads to ‘Chambourcin’ diploid, primary, and haplotig assemblies separately and obtained an average of 95.08%, 90.17%, and 75.88% mapping results, respectively (Table 5). We also separately mapped trimmed RNA-seq reads to ‘Chambourcin’ diploid, primary, and haplotig assemblies and obtained 87%, 80.81%, and 60.77% mapping results, respectively. The mapping results of both trimmed Illumina whole genome and trimmed RNA-seq reads to genome assemblies show that most reads were mapped to the diploid assembly, followed by primary and haplotig assemblies. The mapping results for the primary genome assembly retained most of the genome from the diploid assembly, while the smallest number of mapped reads belonged to the haplotig assembly. These results suggest that the smallest genome portion is missing in the haplotig assembly.

Table 5. Mapping of Illumina whole genome reads to the genome assembly.

Illumina reads	Diploid assembly	Primary assembly	Haplotig assembly
HF3YFDSXX Total reads: 105,912,898	101,506,499 (mapped 95.84%) 99,659,968 (properly paired 94.10%)	96,261,693 (mapped 90.89%) 92,943,104 (properly paired 87.75%)	81,153,020 (mapped 76.62%) 77,352,356 (properly paired 73.03%)
HJNMYDSXX Total reads: 18,428,676	17,632,525 (mapped 95.68%) 17,318,958 (properly paired 93.98%)	16,719,389 (mapped 90.72%) 16,148,280 (properly paired 87.63%)	14,095,158 (mapped 76.48%) 13,438,716 (properly paired 72.92%)
HL7HHDSXX Total reads: 12,705,214	12,193,364 (mapped 95.97%) 11,974,672 (properly paired 94.25%)	11,559,561 (mapped 90.98%) 11,164,510 (properly paired 87.87%)	9,734,853 (mapped 76.62%) 9,285,522 (properly paired 73.08%)
HMMHLDSXX Total reads: 171,257,348	159,008,113 (mapped 92.85%) 147,418,324 (properly paired 86.08%)	150,914,465 (mapped 88.12%) 136,316,148 (properly paired 79.60%)	126,442,365 (mapped 73.83%) 112,044,416 (properly paired 65.42%)

CONCLUSION

In this study, we presented the genome assembly of ‘Chambourcin’, a complex interspecific hybrid grape cultivar, using PacBio HiFi long read sequencing, Bionano third-generation sequencing data, and Illumina short read data. The comparative genomic analyses of ‘Chambourcin’ with the reference genome of *V. vinifera* ‘PN40024’ 12X.v2, Shine Muscat, and *V. riparia* ‘Gloire’ indicated that the ‘Chambourcin’ genome aligns well with other grape genomes without any large structural variation. Ortholog analyses of the ‘Chambourcin’ primary gene models, *V. vinifera* ‘PN40024’ 12X.v2, VCost.v3, Shine Muscat, and *V. riparia* ‘Gloire’, revealed that our ‘Chambourcin’ genome assembly and gene annotations are a high-quality grapevine resource for the research community.

Interspecific hybrids derived from two or more *Vitis* species are common in nature [1]. They are the cornerstone of grapevine rootstocks grown worldwide, cultivars that predominate in eastern and midwestern North America, and new disease-resistant genotypes currently in development [46]. The sequencing data, scaffold assemblies, and gene annotations of the ‘Chambourcin’ genome assembly described here provide a valuable resource for genome comparisons, functional genomic analyses, and genome-assisted breeding research.

DATA AVAILABILITY

The PacBio HiFi and Illumina whole genome reads are deposited in the NCBI BioProject with accession [PRJNA754438](#). The Sequence Read Archive (SRA) accession of the PacBio HiFi reads is [SRR15530464](#), and the SRA accession of the Illumina whole genome reads are [SRR24093946](#), [SRR24093988](#), [SRR24095403](#), and [SRR24097763](#). The Bionano maps, genome assembly, gene annotation, proteins, and other data are available on figshare [47]. Supplementary tables and additional data is in the GigaDB repository [45].

LIST OF ABBREVIATIONS

BUSCO: Benchmarking Universal Single-Copy Orthologs; CCS: circular consensus sequencing; DLS: Direct Label and Stain; GO: Gene Ontology; HMW: high molecular weight; LINEs: long interspersed nuclear elements; LTRs: long terminal repeats; CDS: Coding Sequences; PlantTFDB: Plant Transcription Factor Database; SNPs: single nucleotide polymorphisms; SSRs: Simple Sequence Repeats; USDA: United States Department of Agriculture.

DECLARATIONS

Ethics approval

The authors declare that ethical approval was not required for this type of research.

Competing Interests

The authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

Authors' contributions

SP, ZH, AF, and AM conceived and designed this study. AF provided computational resources and guidance, and JPL provided 'Chambourcin' samples for the Illumina whole genome sequencing and rhAmpseq marker haplotype sequences. SP processed the DNA sequences for 'Chambourcin', assembled the genome, and conducted the synteny analysis. SP processed the RNA-Seq data for gene prediction, and conducted gene prediction and annotation. SP conducted comparative genomics analyses and uploaded the sequences to NCBI and figshare. SP wrote the first draft of the manuscript. SP, AM, ZH, AF, and JPL reviewed and finalized the manuscript.

Funding

This project was funded by NSF Plant Genome Research Program 1546869 to AM, AF, and JPL.

Acknowledgements

We acknowledge Laszlo Kovacs for collecting 'Chambourcin' samples for sequencing and Alex Harkess for assistance in developing protocols for DNA extractions from 'Chambourcin'. Roberto Villegas-Diaz, Chad Julius, Luke Grassman, and Rachael Auch assisted with installing and debugging tools in the South Dakota University Research Cyberinfrastructure High Performance Computing Cluster Roaring Thunder.

REFERENCES

- 1 **Morales-Cruz A, Aguirre-Liguori JA, Zhou Y et al.** Introgression among North American wild grapes (*Vitis*) fuels biotic and abiotic adaptation. *Genome Biol.*, 2021; **22**: 254. doi:10.1186/s13059-021-02467-z.
- 2 **Myles S, Boyko AR, Owens CL et al.** Genetic structure and domestication history of the grape. *Proc. Natl. Acad. Sci. USA*, 2011; **108**(9): 3530–3535. doi:10.1073/pnas.1009363108.
- 3 **Dong Y, Duan S, Xia Q et al.** Dual domestications and origin of traits in grapevine evolution. *Science*, 2023; **379**(6635): 892–901.
- 4 **Pinney T.** A History of Wine in America. Volume 1. From the Beginnings to Prohibition. Berkeley, CA: University of California Press, 1989.
- 5 **Maul E et al.** Vitis International Variety Catalogue. 2023; <https://www.vivc.de/>.
- 6 **Winetraveller Website.** <https://www.winetraveler.com>.
- 7 **Migicovsky Z, Harris ZN, Klein LL et al.** Rootstock effects on scion phenotypes in a 'Chambourcin' experimental vineyard. *Hortic. Res.*, 2019; **6**: 64. doi:10.1038/s41438-019-0146-2.
- 8 **Maimaitiyiming M, Sagan V, Sidike P et al.** Leveraging very-high spatial resolution hyperspectral and thermal UAV imageries for characterizing diurnal indicators of grapevine physiology. *Remote Sens.*, 2020; **12**: 3216. doi:10.3390/rs12193216.
- 9 **Awale M, Liu C, Kwasniewski M.** A metabolomics-based approach to differentiate volatiles between a European and Hybrid Grapes and Wines. In: Hortscience. vol. 56, USA: Amer Soc Horticultural Science, 2021.

- 10 Harris ZN, Pratt JE, Bhakta N et al. Temporal and environmental factors interact with rootstock genotype to shape leaf elemental composition in grafted grapevines. *Plant Direct*, 2022; 6(8): e440. doi:10.1002/pld3.440.
- 11 Harris ZN, Awale M, Bhakta N et al. Multi-dimensional leaf phenotypes reflect root system genotype in grafted grapevine over the growing season. *GigaScience*, 2021; 10(12): giab087. doi:10.1093/gigascience/giab087.
- 12 SMRT Analysis software. <https://www.pacb.com/products-and-services/analytical-software/smrt-analysis/>. Accessed 1st January 2023.
- 13 Bionano Support Documentation. <https://bionano.com/support-documentation/>. Accessed 1st January 2023.
- 14 Bionano Software and Data Analysis Support Materials. <https://bionano.com/software-and-data-analysis-support-materials/>. Accessed 1st January 2023.
- 15 Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics*, 2014; 30(15): 2114–2120.
- 16 Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 2011; 27: 764–770.
- 17 Vurtture GW, Sedlazeck FJ, Nattestad M et al. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*, 2017; 33: 2202–2204.
- 18 Cheng H, Concepcion GT, Feng X et al. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods*, 2021; 18: 170–175.
- 19 Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinform.*, 2018; 19: 460. doi:10.1186/s12859-018-2485-7.
- 20 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, 2012; 9(4): 357–359. doi:10.1038/nmeth.1923.
- 21 Danecek P, Bonfield JK, Liddle J et al. Twelve years of SAMtools and BCFtools. *GigaScience*, 2021; 10(2): giab008. doi:10.1093/gigascience/giab008.
- 22 Walker BJ, Abeel T, Shea T et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*, 2014; 9(11): e112963.
- 23 Canaguier A, Grimplet J, Di Gaspero G et al. A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3). *Genom. Data*, 2017; 14: 56–62.
- 24 Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 2018; 34(18): 3094–3100. doi:10.1093/bioinformatics/bty191.
- 25 Zou C, Karn A, Reisch B et al. Haplotyping the Vitis collinear core genome with rhAmpSeq improves marker transferability in a diverse genus. *Nat. Commun.*, 2020; 11: 413.
- 26 Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 2009; 25(14): 1754–1760.
- 27 Simão FA, Waterhouse RM, Ioannidis P et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 2015; 31(19): 3210–3212. doi:10.1093/bioinformatics/btv351.
- 28 Shirasawa K, Hirakawa H, Azuma A et al. De novo whole-genome assembly in an interspecific hybrid table grape, ‘Shine Muscat’. *DNA Res.*, 2022; 29(6): dsac040. doi:10.1093/dnares/dsac040.
- 29 Girollet N, Rubio B, Lopez-Roques C et al. De novo phased assembly of *Vitis riparia* grape genome. *Sci. Data*, 2019; 6: 127.
- 30 dotPlotly GitHub. 2018; <https://github.com/tpoorten/dotPlotly>.
- 31 Mapleson D, Accinelli GG, Kettleborough G et al. KAT: A K-mer Analysis Toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics*, 2016; 33(4): 574–576. doi:10.1093/bioinformatics/btw663.
- 32 Flynn JM, Hubley R, Goubert C et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA*, 2020; 117(17): 9451–9457. doi:10.1073/pnas.1921046117.
- 33 Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013–2015; <http://www.repeatmasker.org>.

- 34 Kim D, Paggi JM, Park C et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.*, 2019; 37: 907–915.
- 35 Brůna T, Hoff KJ, Lomsadze A et al. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.*, 2021; 3(1): lqaa108.
- 36 Cantalapiedra CP, Hernández-Plaza A, Letunic I et al. eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.*, 2021; 38(12): 5825–5829. doi:10.1093/molbev/msab293.
- 37 Ye J, Zhang Y, Cui H et al. WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Res.*, 2018; 46(W1): W71–W75. doi:10.1093/nar/gky400.
- 38 Xu L, Dong Z, Fang L et al. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.*, 2019; 47(W1): W52–W58.
- 39 Jin JP, Tian F, Yang DC et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.*, 2017; 45(D1): D1040–D1045.
- 40 Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.*, 2016; 33: 1870–1874.
- 41 Patel S, Robben M, Fennell A et al. Draft genome of the Native American cold hardy grapevine *Vitis riparia* Michx. ‘Manitoba 37’. *Hortic Res.*, 2020; 7: 92. doi:10.1038/s41438-020-0316-2.
- 42 Soderlund C, Bomhoff M, Nelson W. SyMAP: A turnkey synteny system with application to plant genomes. *Nucleic Acids Res.*, 2010; 39(10): e68.
- 43 Beier S, Thiel T, Münch T et al. MISA-web: a web server for microsatellite prediction. *Bioinformatics*, 2017; 33(16): 2583–2585.
- 44 Krzywinski M, Schein J, Birol I et al. Circos: an information aesthetic for comparative genomics. *Genome Res.*, 2009; 19(9): 1639–1645. doi:10.1101/gr.092759.109.
- 45 Patel S, Harris ZN, Londo JP et al. Supporting data for “Genome assembly of the hybrid grapevine *Vitis* ‘Chambourcin’”. *GigaScience Database*, 2023; <http://dx.doi.org/10.5524/102415>.
- 46 Migicovsky Z, Sawler J, Money D et al. Genomic ancestry estimation quantifies use of wild species in grape breeding. *BMC Genom.*, 2016; 17: 478. doi:10.1186/s12864-016-2834-8.
- 47 Patel S, Harris ZN, Londo JP et al. Genome assembly of the hybrid grapevine *Vitis* ‘Chambourcin’. *Figshare Dataset*, 2023; <https://doi.org/10.6084/m9.figshare.15505788.v1>.