The assembly and annotation of two teinturier grapevine varieties, Dakapo and Rubired

Eleanore J. Ritter[1], Noé Cochetel[2], Andrea Minio[2], Peter Cousins[3], Dario Cantu[2,4], and Chad Niederhuth[1†]

[1]Department of Plant Biology, Michigan State University, East Lansing, Michigan, USA

[2]Department of Viticulture and Enology, University of California Davis, Davis, California, USA

[3]E. & J. Gallo Winery, Modesto, California, USA

[4]Genome Center, University of California Davis, Davis, California, USA

[†]Current Address: Corteva, Inc. Indianapolis, Indiana, USA

Correspondence:

Dario Cantu (dacantu@ucdavis.edu) and Chad Niederhuth (niederhu@msu.edu)

ABSTRACT

Background

Teinturier grapevine varieties were first described in the 16th century and have persisted due to their deep pigmentation. Unlike most other grapevine varieties, teinturier varieties produce berries with pigmented flesh due to anthocyanin production within the flesh. As a result, teinturier varieties are of interest not only for their ability to enhance the pigmentation of wine blends but also for their health benefits. Here, we assembled and annotated the Dakapo and Rubired genomes, two teinturier varieties.

Findings

For Dakapo, we used a combination of Nanopore sequencing, Illumina sequencing, and scaffolding to the existing grapevine genome assembly to generate a final assembly of 508.5 Mbp with an N50 scaffold length of 25.6 Mbp and a BUSCO score of 98.0%. A combination approach of *de novo* annotation and lifting over annotations from the existing grapevine reference genome resulted in the annotation of 36,940 genes in the Dakapo assembly. For Rubired, PacBio HiFi reads were assembled, scaffolded, and phased to generate a diploid assembly with two haplotypes 474.7-476.0 Mbp long. The diploid genome has an N50 scaffold length of 24.9 Mbp and a BUSCO score of 98.7%, and both haplotype-specific genomes are of similar quality. *De novo* annotation of the diploid Rubired genome yielded annotations for 56,681 genes.

Conclusions

The Dakapo and Rubired genome assemblies and annotations will provide genetic resources for future investigations into berry flesh pigmentation and other traits of interest in grapevine.

KEYWORDS

grapevine; *Vitis vinifera*; teinturier; berry color; berry flesh; anthocyanin

CONTEXT

Domesticated grapevine (*Vitis vinifera*) is the fifth most produced fruit globally [1], with 80.1 million tonnes produced in 2022 alone [2]. Various grapevine varieties have been bred since its domestication ~11,000 years ago [3], mostly for winemaking purposes. This has resulted in the selection of numerous diverse phenotypes with significant variation in traits, including berry color and aromatic compounds, as well as more utilitarian traits like yield or biotic and abiotic stress resistance. Berry color is of particular importance in wine grapes due to how it influences wine color and quality. Both consumers and experts prefer red wines with darker colorations [4,5], making strong pigmentation in berries advantageous for wine producers. Pigmentation typically occurs only in the skin of ripened grapevine berries, with most grapevine varieties having white-colored flesh. The pigmentation within the berry skin is due to the production of anthocyanins, which are colored flavonoids that also act as antioxidants [6]. As anthocyanins significantly influence both the quality of wines and their health benefits, the genetic and molecular pathways involved in anthocyanin produced in berry skin have been of high interest and well-characterized [7].

Teinturier (also known as “dyer”) varieties produce berries with pigmented skin and flesh, as well as pigmented leaves. They are highly favorable for use in red wine blends, in which they provide a deeper color. They also remain valuable resources for understanding the production of anthocyanins outside of berry skin. Dakapo and Rubired are two teinturier varieties that are widely grown, and Rubired was the 8th-most crushed grapevine variety in California in 2022 [8]. Both

varieties are descendants of Teinturier du Cher but of distinct generations. Dakapo was initially bred through a cross between Deckrot and Blauer Portugieser, with Deckrot being a direct descendant of Teinturier du Cher. Rubired is a hybrid grapevine variety bred through a cross between Tinto Cão and Alicante Ganzin, with Alicante Gazin being a fourth-generation descendent of Teinturier du Cher. While both Dakapo and Rubired are descendants of Teinturier du Cher, their Teinturier du Cher ancestors likely differ based on previous genetic work in teinturier grapes [9] (Figure 1).

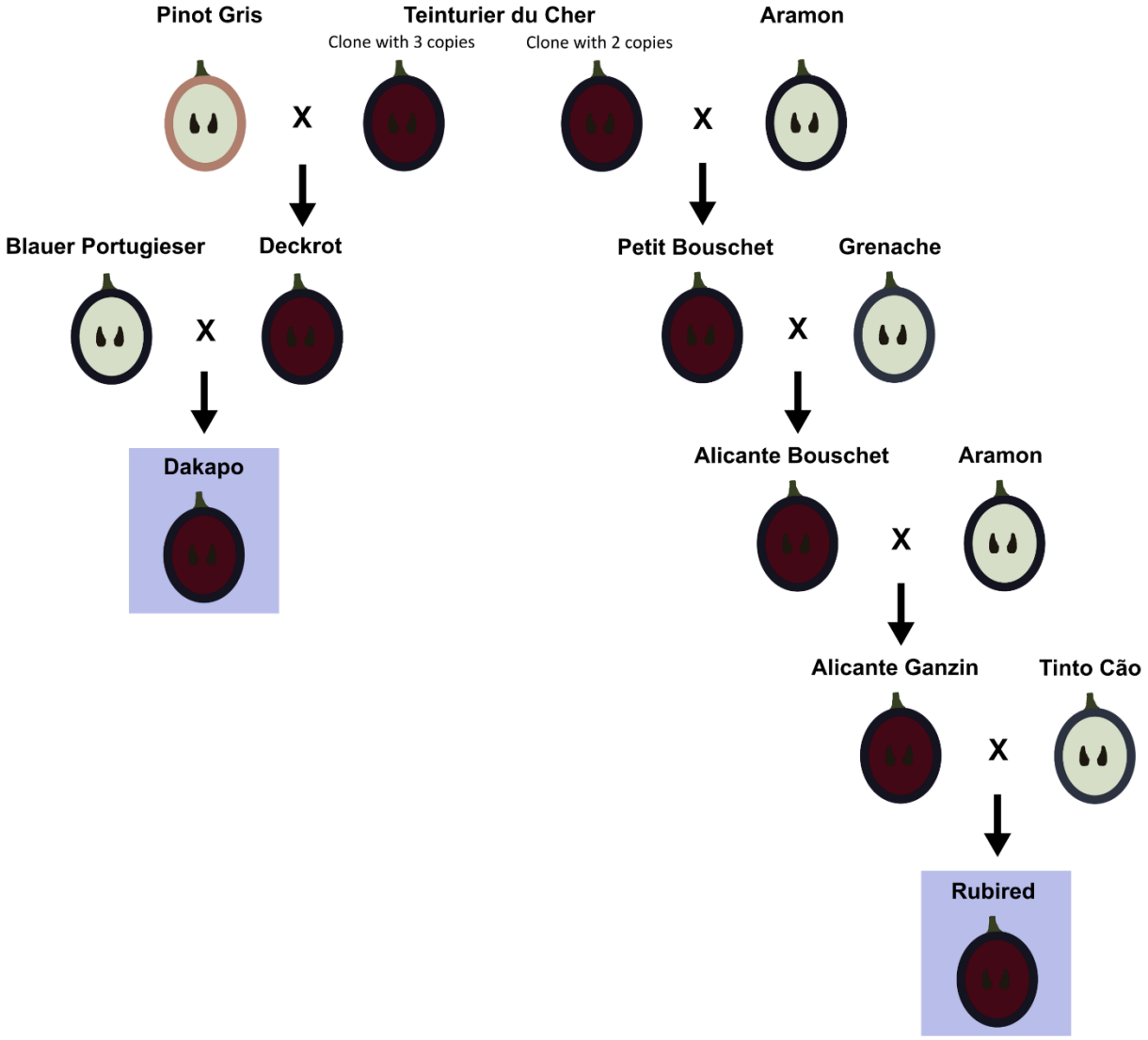

Figure 1. The ancestry of Dakapo and Rubired, with berry skin and flesh color shown. Dakapo and Rubired are thought to have been bred from Teinturier du Cher clones with differing copy numbers of a 408 bp repeat within the promoter of *VvMybA1*, which is noted in the figure.

Teinturier varieties have substantially higher anthocyanin content in their berries than non-teinturier varieties due to the accumulation of anthocyanins within their berry flesh. Previous work showed that the juice produced with Dakapo berries had 39-91 times more anthocyanin content than commercial red grape juice [10]. Teinturier varieties themselves vary in anthocyanin content and the profiles of anthocyanins present within the berry flesh [9,11]. Previous studies have made progress in investigating the genetic basis of berry flesh pigmentation and variation in overall anthocyanin production within teinturier grapes. A previous study [9] demonstrated that a 408 bp repeat in the promoter of the gene *VvMybA1* is directly linked to increased anthocyanin production in teinturier berries. They also found that the copy number of this 408 bp sequence varied between varieties and that varieties derived from Teinturier du Cher had either two, three, or five copies of this 408 bp sequence within the promoter region. Increased copies of these repeats were correlated with increased anthocyanin content within berry skin and flesh [9]. While this study greatly illuminated the genetic basis of increased anthocyanin production in Teinturier du Cher descendants, it is still unclear why different teinturier grapes have distinct anthocyanin profiles in berry flesh, regardless of the number of copies of the 408 bp sequence they have [11]. For example, the concentration of a specific type of anthocyanin, Cyanidin-3-*O*-glucoside, can vary from 1.0-41.7 mg/L when comparing the anthocyanin content within the flesh of teinturier berries from different varieties [11]. While overall anthocyanin content does correlate with the number of repeats upstream of *VvMybA1 [9]*, large differences in concentrations of specific anthocyanins exist between teinturier varieties with the same number of copies of the 408 bp repeat [11]. The

assembly and annotation of the genome of the Yan73 teinturier grapevine variety were recently generated and provided additional insight into the regulation of anthocyanin accumulation in Yan73 berry flesh [12]. However, the lack of additional genomic resources for teinturier grapes has inhibited further investigations into differences between teinturier varieties, and the genetic basis for these large differences in anthocyanin composition remains unclear as a result.

Here, we sequenced, assembled, and annotated the Dakapo and Rubired genomes to provide additional resources for understanding teinturier varieties and to further enable their use in breeding programs. These genomes will greatly facilitate future work into understanding the regulation of anthocyanins within berry flesh. Beyond anthocyanin production, Dakapo and Rubired have also been utilized in researching other traits in grapevine. A QTL mapping population of Dakapo x Cabernet Sauvignon has been established and utilized for investigating *Botrytis* bunch rot in grapevine [13]. Additionally, Rubired is notable for being highly resistant to *Xylella fastidiosa*, which causes Pierce's disease in grapevine [14–16]. As a result, we believe these high-quality reference genome assemblies and annotations will be a useful resource for the grapevine and plant science communities.

METHODS

Plant material

Vitis vinifera plants of the Dakapo variety were planted in Madera, California, USA in 2011. Young leaf tissue samples for Oxford Nanopore Technologies (ONT) long-read sequencing were collected in July 2021. The samples were frozen and shipped on dry ice overnight. Plant material used in this study was also utilized in a previous study [17] as the "Dakapo WT" samples.

For Rubired tissue, young leaves were collected from the accession Rubired FPS clone 02 maintained by the Foundation Plant Services at the University of California, Davis.

DNA extraction and sequencing of Dakapo tissue

High molecular weight DNA was extracted and a sequencing library was prepared for ONT sequencing by the Genomics Core at Michigan State University as previously described [17], using the Oxford Nanopore Technologies Ligation Sequencing Kit (SQK-LSK109). The library was sequenced on a PromethION FLO-PRO002 flow cell (R9.4.1; Oxford Nanopore Technologies) on a PromethION24 (Oxford Nanopore Technologies) running MinKNOW Release 21.11.7 (Oxford Nanopore Technologies), resulting in 61.9 Gbps of sequence (~123.8X coverage) with a read length N50 of 13.2 kbp. Base calling and demultiplexing were performed using Guppy v5.1.13 (Oxford Nanopore Technologies) with the High Accuracy base calling model. An additional 9.5 Gbps (~18.9X coverage) of ONT sequencing data and 25.6 Gbps (~51.3X coverage) of Illumina paired-end sequencing data previously published from the "Dakapo WT" sample described previously [17] were utilized for this study (see Supplementary Table S1 for full sequencing statistics).

Dakapo genome assembly

Raw ONT sequencing data from this study and previous work with the same plant material [17] were combined. Adapters were trimmed using Porechop v0.2.4 [18] with the following settings: *--min_trim_size 5, --extra_end_trim 2, --end_threshold 80, --middle_threshold 90, --extra_middle_trim_good_side 2, --extra_middle_trim_bad_side 50,* and *--min_split_read_size 300*. Reads mapping to the lambda phage genome were removed using NanoLyse v1.2.0 [19]. NanoFilt v2.8.0, with the flags *-q 0* and *-l 300*, was used to remove low-quality reads and reads

shorter than 300 base pairs (bp) [19]. The quality of the reads was analyzed using FastQC v0.11.9 [20], NanoStat v1.6.0 [19], and NanoPlot v1.38.0 [21]. ONT reads were then assembled using Flye v2.8.3-b1695 [22] for two iterations. One round of polishing was performed on the assembly using the ONT reads with Racon v1.4.20 [23] and the following settings: *--include-unpolished*, *-m 8*, *-x -6*, *-g -8*, and *-w 500*. The assembly was then scaffolded to the 12X.v2 grapevine genome assembly [24] using RagTag v2.0.1 "scaffold" [25] with the following settings: *-f 1000*, *-d 100000*, *-i 0.2*, *-a 0.0*, *-s 0.0*, *-r*, *-g 100*, and *-m 100000*.

Paired-end Illumina reads were used for the final polishing of the scaffolded assembly. These were first trimmed using Trimmomatic v0.39 [26] to remove adapters and low-quality sequences, with the following settings used: *-phred33*, *ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:4:TRUE*, *LEADING:3*, *TRAILING:3*, *SLIDINGWINDOW:4:15*, and *MINLEN:30*. The reads were then mapped to the draft genome assembly using BWA-MEM v0.7.17-r1188 [27] with the *-M* flag. PCR duplicate reads were removed using Picard MarkDuplicates v2.15.0 [28] with the *--REMOVE_DUPLICATES TRUE* flag. The mapped reads with marked duplicates were then used to polish the draft assembly using Pilon v1.24 [29] with the *--fix all* flag used to correct all errors identified and the *--diploid* flag. Two iterations of Pilon polishing were performed.

Following polishing, haplotigs were removed using Purge Haplotigs v1.1.2 [30]. To do so, all prepped ONT reads were mapped to the Dakapo draft assembly using minimap2 v2.23-r1111 [31] with the flags *-ax map-ont* and *-L*, and *purge_haplotigs hist* was then run with default settings to generate a read-depth histogram of these mapped reads. Based on the histogram generated, *purge_haplotigs cov* was run with the previous output file and the following flags: *-low 15*, *-mid 88*, and *-high 195*. Finally, *purge_haplotigs purge* was run using the previous output file to purge haplotigs from the Dakapo draft assembly [30].

Before finalizing the Dakapo genome assembly, chr00 was split apart manually at gaps since it is an artificial chromosome of unmapped contigs from the 12X.v2 grapevine genome assembly [24] used for scaffolding. The assembly was also searched for microbial contamination using the gather-by-contig.py script adapted from [32] that utilizes sourmash and its pre-built database "GTDB R06-RS202 genomic representatives" [33]. No contamination was found from this process. The chromosome names were maintained from the scaffolding to the 12X.v2 grapevine genome assembly [24]. All other contigs were sorted and renamed in order of length, including the contigs split apart from chr00, using the custom script sort_rename_fasta.sh.

To assess the quality of the Dakapo genome assembly, we used BUSCO v5.2.2 [34] to check the completeness of the assembly when compared to the eudicots_odb10 dataset at each step of genome assembly and polishing. Assembly statistics were calculated using assembly-stats v1.0.1 [35]. Finally, the quality of repetitive sequences and intergenic space was also assessed by calculating the long terminal repeats (LTR) Assembly Index (LAI) for the Dakapo assembly using LTRs annotated by the Extensive de-novo TE Annotator (EDTA) (see below for transposable element annotation methods).

Dakapo genome annotation

Transposable elements (TEs) and repeats in the Dakapo genome assembly were annotated using EDTA v1.9.4 [36] with the following flags: *--species others, --step all, --overwrite 1, --sensitive 1, --anno 1, --evaluate 0,* and *--force 1*.

MAKER was used for *de novo* annotation of genes in the Dakapo genome. Before running MAKER, RNA-seq reads from diverse tissues in grapevine and protein sequences from related species were used to provide initial support for gene models. To do so, RNA-seq samples from

previous studies [37–42] were downloaded from the NCBI Sequence Read Archive (SRA) using fasterq-dump v2.10.7 from the sra-toolkit [43] (see Supplementary Table S2 for the SRA IDs of the specific files used). Trimmomatic v0.39 [26] was used to trim adapters from Illumina RNA-seq reads with the flags: *--phred33* and *ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10*. These were then mapped to the Dakapo genome assembly using HISAT2 v2.2.1 [44] with the *--phred33* flag. PacBio RNA-seq reads were mapped to the Dakapo genome assembly using minimap2 v2.23-r1111 [31] with the flags *-ax splice:hq* and *-uf*. Transcripts from these mapped RNA-seq reads were assembled using StringTie v2.2.1 [45] with the following flags: *-c 1, -f 0.01, -m 200, -a 10, -j 1, -M 1, -s 4.75* (for mapped Illumina reads) or *1.5* (for mapped PacBio reads), and *-g 50* (for mapped Illumina reads) or *0* (for mapped PacBio reads). The output files for all RNA-seq samples were converted to gff3 files using gffread v0.12.7 [45], combined, and then sorted using gff3_sort v2.1.0 [46].

Before protein sequences were aligned to the Dakapo genome assembly, repeats in the assembly were masked using RepeatMasker v4.1.2-p1 with the TE library generated using EDTA and the following flags: *-e rmblast, -s, -norna, -xsmall, -gff, -html,* and *-source*. Protein sequences from the Arabidopsis (*Arabidopsis thaliana*) Araport11 annotation *[47]*, the *Oryza sativa* Release 7 annotation [48], and the Viridiplantae UniProtKB/Swiss-Prot reviewed protein sequence dataset from UniProt release 2023_05 [49] were aligned to the masked Dakapo genome assembly using exonerate v2.4.0 [50] with the following flags: *--model protein2genome, --bestn 5, --minintron 10, --maxintron 5000, --querychunktotal 5, --targetchunktotal 10, --showtargetgff yes, --showalignment no, --showvulgar no, --ryo ">%qi length=%ql alnlen=%qal\n>%ti length=%tl alnlen=%tal\n"*. The outputs for each dataset were combined, reformatted using the custom script reformat_exonerate_protein_gff.pl, and sorted using gff3_sort v2.1.0 [46].

MAKER v3.01.04 [51] was initially run on the Dakapo genome assembly with the gff files generated through transcript assembly and protein sequence alignment. These initial annotations generated by MAKER were then used to train SNAP and AUGUSTUS. To train SNAP, maker2zff from MAKER was first used to convert genes to the ZFF format with the flag *-x 0.1*. This input was used with SNAP v2013_11_29 [52] to first categorize genes by running the command *fathom* to produce reformatted files, followed by the command *forge* to estimate parameters. Hidden Markov Models (HMMs) were created using *hmm-assembler.pl* from SNAP [52]. To train AUGUSTUS, maker2zff from MAKER was first used to convert genes to the ZFF format with the following flags: *-c 0.5, -e 0.5, -o 0.5, -a 0, -t 0, -l 200,* and *-x 0.2*. The *fathom* command from SNAP [52] followed by the custom script fathom_to_genbank.pl were then run to reformat the files and keep only 600 randomly sampled annotations. Fasta files of the subsetted genes were then generated using the custom script get_subset_of_fastas.pl. These subsetted genes were split into training and test files, and then *autoAug.pl* from AUGUSTUS v3.4.0 [53] was run to produce batch scripts that were then run. This step was repeated, using the following flags with *autoAug.pl*: *-useexisting* and *--index=1*. The sensitivity and specificity of the AUGUSTUS HMMs were evaluated by running the *augustus* command. A second round of MAKER v3.01.04 [51] was then run using the HMMs from SNAP and AUGUSTUS to produce gene annotations.

We then filtered annotations and flagged genes that may have been misannotated as transposons using methods described previously [54]. To ensure that our annotations were as complete as possible, we used Liftoff v1.6.2 to transfer annotations from the PN40024.v4 grapevine genome assembly [55] to the Dakapo genome. We then used the methods described previously [54] to assign “pseudogene” and “gene” labels to the lifted genes based on the confidence of the lifted gene model.

Finally, gene functions were assigned to each annotated gene by first using InterProScan v5.66-98.0 [56] to assign Pfam domains and corresponding gene ontology (GO) terms using the following flags: *-appl pfam, -goterms, -pa, -dp, -iprlookup, -t p,* and *-f TSV*. Then, Arabidopsis orthologs were identified by running DIAMOND v2.0.15.153 [57] with protein sequences from Dakapo and the TAIR10 Arabidopsis annotation [58] and the following flags: *--evalue 1e-6, --max-hsps 1, --max-target-seqs 5,* and *--outfmt 0*. The results from InterProScan, DIAMOND, and Arabidopsis gene functions and GO terms of orthologs [59] were all combined to generate a file with functional descriptions for each gene using the custom script create_functional_annotation_file.pl.

DNA extraction and sequencing of Rubired tissue

High molecular weight genomic DNA was extracted using the method previously described [60]. The HiFi library preparation and sequencing were performed as previously described [61]. The HiFi library fraction with a length >15 kbp was sequenced in two SMRT cells on a PacBio Sequel IIe platform at the DNA Technology Core Facility, University of California, Davis. The sequencing generated 31.3 Gbp sequences corresponding to 62.6X coverage with an N50 of 11.5 kbp.

Rubired genome assembly

The pseudomolecules of the Rubired genome were assembled, phased, and scaffolded as described previously [61]. Briefly, after testing multiple Hifiasm v.0.16.1-r374 [62] parameters, the best assembly obtained with the configuration '*-a 4 -k 41 -w 71 -f 25 -r 4 -s 0.7 -D 3 -N 100 -n 25 -z 20*' was selected consisting of 273 contigs with an N50 = 12.9 Mb. An integrated phasing and scaffolding procedure further led to the construction of chromosome-scale pseudomolecules

using HaploSync [63] combined with a high-density consensus map [64]. Quality and completeness of the assembly were assessed as described above for the Dakapo genome.

Rubired genome annotation

The structural and functional annotation of the Rubired genome followed the exhaustive annotation pipeline described [65]. Briefly, high-quality Iso-Seq data from *V. vinifera* Cabernet Sauvignon [39], quality-based filtered RNA-Seq data from *V. rupestris* [66], and external databases were used to generate a collection of assemblies, alignments, *ab initio* predictions, and transcript/protein evidence. The resulting consensus gene models were then polished, filtered, and functionally annotated following the annotation workflow established previously [65].

Investigating the *VvMybA1* sequences in the Dakapo and Rubired genomes

Teinturier grape varieties are known to have tandem copies of a 408 bp sequence within the promoter region of *VvMybA1*, a key gene in anthocyanin biosynthesis, which leads to increased anthocyanin accumulation within their berries. Dakapo and Rubired both contain 3 and 2 copies of this repeat, respectively [9]. To investigate the sequence similarities of these repeat sequences in our assembled genome, we first used blastn from BLAST v2.10.0+ [67] to search for the locations of *VvMybA1* within the Dakapo genome and both Rubired haplotypes. We then extracted the sequences of *VvMybA1* and the 10 kbp surrounding the gene from the genomes using bedtools v2.27.1 *getfasta* [68]. We searched these sequences for the 408 bp repeat identified previously [9] both manually and using blastn from BLAST v2.10.0+ [67].

Exploring synteny among various grapevine genomes

GENESPACE v1.3.1 [69] was used with MCScanX [70] and Orthofinder v2.5.5 [71] to align and plot protein sequences from the following grapevine genomes: Dakapo, Rubired, Cabernet Franc [63], Cabernet Sauvignon [63], Chardonnay [72], and Pinot Noir [73]. Individual chromosomes from the Dakapo and Rubired assemblies were aligned using MUMmer v4.0.0rc1 [74]. To do so, first, the *nucmer* command was run with default settings. The command *delta-filter* was then run with the following flags: *-i 90 -l 5000*. Finally, plots were generated using the *mummerplot* command.

DATA DESCRIPTION AND QUALITY CONTROL

The Dakapo genome assembly and annotation

The Dakapo genome was assembled using ONT reads representing 142.4X coverage (based on a genome size of 500 Mbps). This draft assembly was then scaffolded to the 12X.v2 grapevine reference genome [24] and polished using both ONT reads (142.4X coverage) and Illumina reads (51.3X coverage) to produce the final genome assembly of 508.5 Mbp. It is comprised of 19 chromosomes and 542 unplaced contigs, with 96.3% of the Dakapo assembly sequence located on the chromosomes and 2,644 gaps of unknown sequence. The final genome assembly is highly contiguous, with an N50 of 25.6 Mbp, slightly higher than the most recent PN40024.v4 assembly [55]. The Dakapo assembly has a high BUSCO (Benchmarking Universal Single-Copy Orthologue) score of 98% complete BUSCOs (94.7% single-copy BUSCOs and 3.3% duplicated BUSCOs), similar to prior PN40024 reference assemblies (97.7% for 12X.v2 [24] and 98.3% for PN40024.v4 [55]). In addition, the Dakapo genome received a raw LAI score of 12.22 and thus contains a reference-quality assembly of repetitive/intergenic sequences [75] (Table 1).

	Dakapo	Rubired whole assembly	Rubired haplotype 1	Rubired haplotype 2	12X.v2	PN40024.v4
Assembly size (Mbp)	508.5	983.8	476.0	474.7	486.2	475.6
Number of contigs	561	185	19	19	20	22
N50 (Mbp)	25.6	24.9	24.7	24.9	24.3	24.4
Number of gaps	2,644	97	38	59	15,325	4,019
Total complete BUSCO	98.0%	98.7%	98.3%	97.3%	97.7%	98.3%
Raw LAI	12.22	N/A*	15.22	15.62	9.4**	13.97

Table 1. Assembly statistics of the Dakapo and Rubired genome assemblies presented here, as well as assembly statistics of the 12X.v2 [24] and PN40024.v4 [55] grapevine reference genome assemblies. The Rubired whole assembly contains both haplotypes and unplaced sequences.
*The raw LAI score was not calculated for the whole assembly due to high sequence similarity between haplotypes which would prevent an accurate calculation
**Previously calculated [75]

The Dakapo genome was annotated using a combination of *de novo* annotations using MAKER [51] and annotations lifted from PN40024.v4 [55] using Liftoff [76]. This resulted in 36,940 genes being annotated. We also annotated both TEs and repeat sequences and found that these comprised 45.38% of the genome, similar to what has previously been reported in grapevine (41.4-51.1% [77–79]). LTRs make up a majority of the repetitive sequences annotated in the Dakapo genome and comprise 30.48% of the genome, with *Gypsy* LTRs specifically being the

most abundant type, comprising 12.88% of the Dakapo genome sequence (Supplementary Table S3).

The Rubired genome assembly and annotation

The Rubired genome was sequenced with highly accurate long-read sequencing generating 62.6X HiFi coverage (using a haploid genome size of 500 Mbp as reference). Pseudomolecules were constructed by scaffolding and phasing the assembly using HaploSync [63] generating two haplotypes comprising 19 chromosomes and averaging to a total length of ~475 Mbp. With complete BUSCO scores of 98.3% and 97.3% for haplotype 1 and haplotype 2, respectively (between 95.4-96.3% single copy BUSCOs and between 1.9-2.0% duplicated BUSCOs), and only 33 Mbp of unplaced sequences in the diploid assembly of the Rubired genome, the Rubired assembly is highly complete. Both genomes for the two Rubired haplotypes also have high raw LAI scores (15.22 for haplotype 1 and 15.62 for haplotype 2), demonstrating that the diploid Rubired genome contains a reference-quality assembly of repetitive/intergenic sequences that are likely more complete than the 12X.v2 [24]3] and PN40024.v4 [55] assemblies (Table 1). The gene annotation resulted in 56,681 genes showing a chromosome anchoring of 97.7% further supporting the reference quality of the assembly. Overall, the genome was composed of 50.46% of repetitive sequences with a clear accumulation in the unplaced sequences with 74.34% of its sequences annotated as repeats. The repeat distribution was similar to the Dakapo genome with *Gypsy* LTRs as the predominant repeats type corresponding to 13.91% of the genome sequence (Supplementary Table S4).

RE-USE POTENTIAL

Grapevine varieties have been bred to produce berries in a variety of colors, commonly divided into red-, black-, and white-skinned berries that typically have white flesh. However, there are several varieties of teinturier grapes, which contain pigmented skin and pigmented flesh, including the Dakapo and Rubired varieties sequenced here. By assembling these genomes, we fully assembled *VvMybA1* and the tandem repeat associated with anthocyanin content in teinturier grapes [9]. As expected, we found three tandem copies of this repeat within the promoter region of *VvMybA1* (the *VvMybA1t3* allele) in the Dakapo genome, exactly as described previously [9]. All three repeats contain identical 408 bp repeat sequences (Figure 2A). In addition, the Rubired haplotype 2 assembly contained two tandem copies of this repeat at the exact location (the *VvMybA1t2* allele) as expected [9], with both copies containing the same 408 bp sequence as those in Dakapo (Figure 2B). The Rubired haplotype 1 assembly did not contain teinturier-associated alleles but instead contained the *VvMybA1a* allele responsible for white berry skin color [80], as expected based on previous findings [9]. The *VvMybA1a* allele is distinct from teinturier alleles and other functional *VvMybA1* alleles due to the presence of the *Gret1* retrotransposon upstream of coding sequences [80]. However, it does contain the 408 bp repeat upstream of the *Gret1* retrotransposon [9]. This repeat sequence is not perfectly identical to the repeat in Dakapo or Rubired haplotype 2 and instead contains three single base pair mutations within the sequence (Figure 2C).

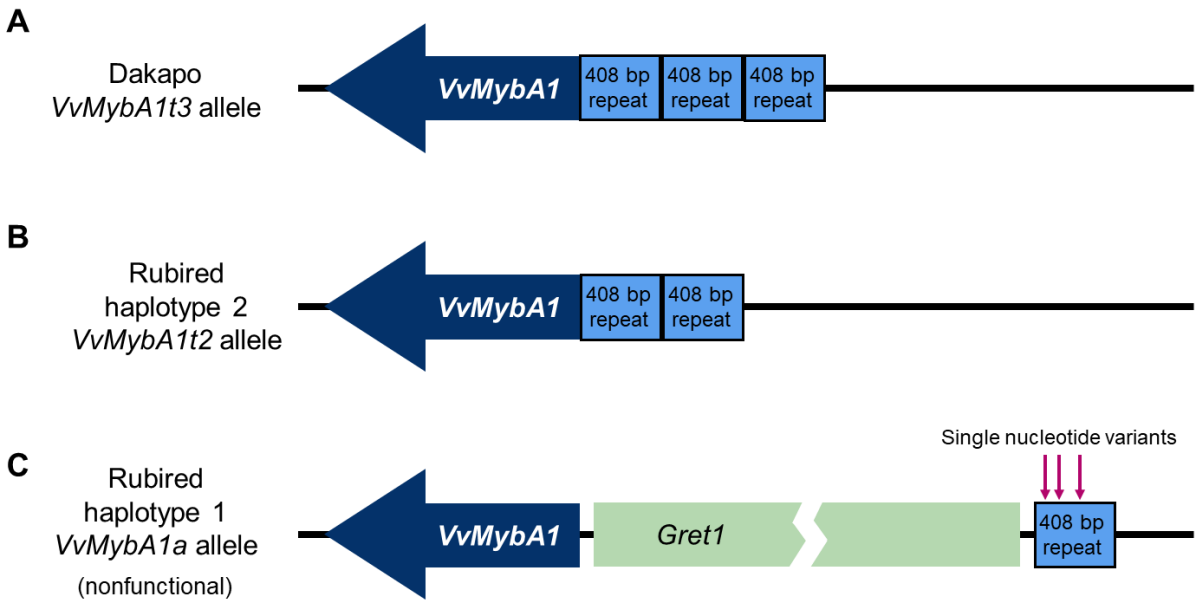

Figure 2. Diagrams of the *VvMybA1* alleles in **A)** the Dakapo assembly, **B)** the Rubired haplotype 2 assembly, and **C)** the Rubired haplotype 1 assembly. *VvMybA1* is represented by the dark blue arrow and the 408 bp repeats are shown in light blue boxes. Dakapo contains three tandem copies of the 408 bp repeat, while the Rubired haplotype 2 assembly contains two tandem copies. The Rubired haplotype 1 assembly contains the nonfunctional *VvMybA1a* allele with the *Gret1* retrotransposon shown upstream of *VvMybA1* in a light green box, truncated to fit in the figure. The three single nucleotide variants within the 408 bp repeat of the nonfunctional allele in Rubired haplotype 1 are indicated by arrows.

Beyond fully sequencing the *VvMybA1* alleles of Dakapo and Rubired, these genomes will enable more insight into grapevine berry color by providing two high-quality teinturier grapevine genomes for future studies. As previously mentioned, teinturier grapes differ in the composition of total anthocyanins produced, and this phenomenon does not seem to be driven by differences in *VvMybA1* alleles [11]. These genomes will provide resources for investigating the genetic mechanisms driving this phenomenon. Focusing on the berry color locus on chromosome 2 [81–

83] and the anthocyanin locus on chromosome 14 [84], in particular, may provide insight into the regulation of specific anthocyanin molecules within the flesh of teinturier berries.

The Dakapo and Rubired genomes and annotations will also offer additional resources for future work in grapevine. Beyond berry flesh color, the Dakapo and Rubired genomes will also provide resources for investigating additional traits. For example, Dakapo is both frost-susceptible [85] and *Botrytis*-susceptible [13], while Rubired is notably highly mildew-resistant [86]. A QTL-mapping population generated through a cross between Dakapo x Cabernet Sauvignon has also been previously established [13], so the availability of this reference genome will greatly aid future studies with this population. These genomes will ultimately provide new resources for investigating a variety of grapevine traits, enabling advances in grapevine breeding and agriculture and allowing for comparisons between grapevine genomes (Figure 3).

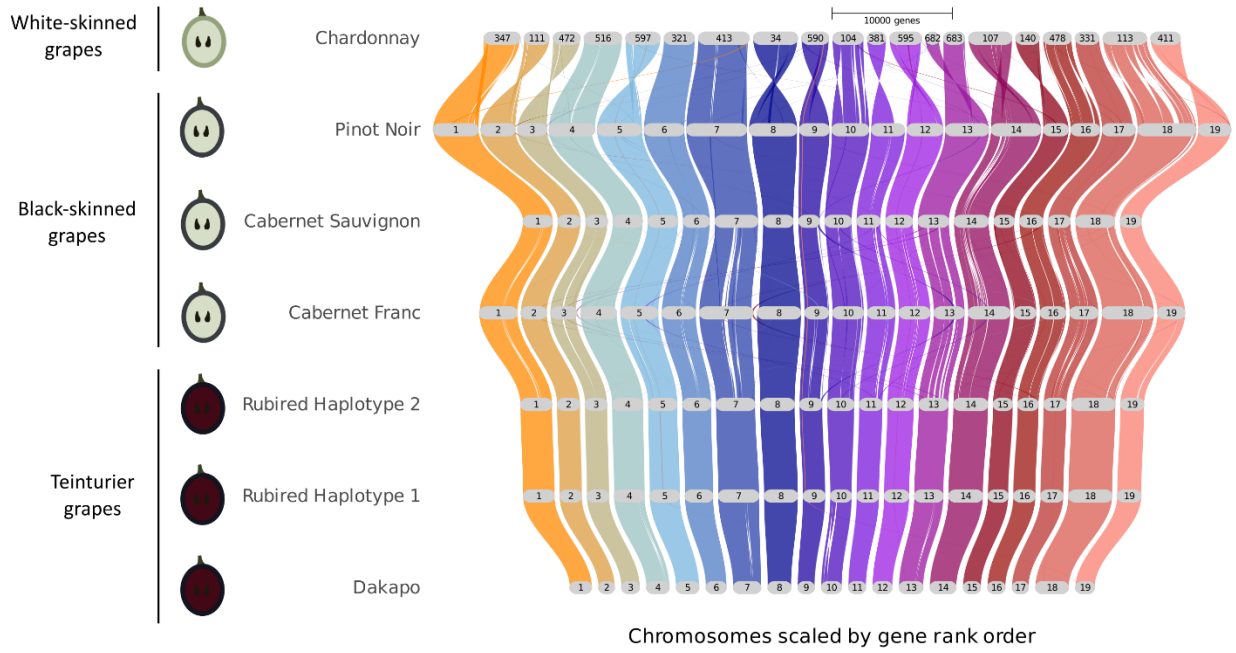

Figure 3. Synteny of several grapevine genomes with chromosome-scale assemblies, organized by berry and flesh color.

Initial comparisons of these assemblies to other grapevine genomes even revealed a large (1.82 Mbp) inversion within Dakapo (Figure 4) that contains 274 genes (Supplementary Table S5).

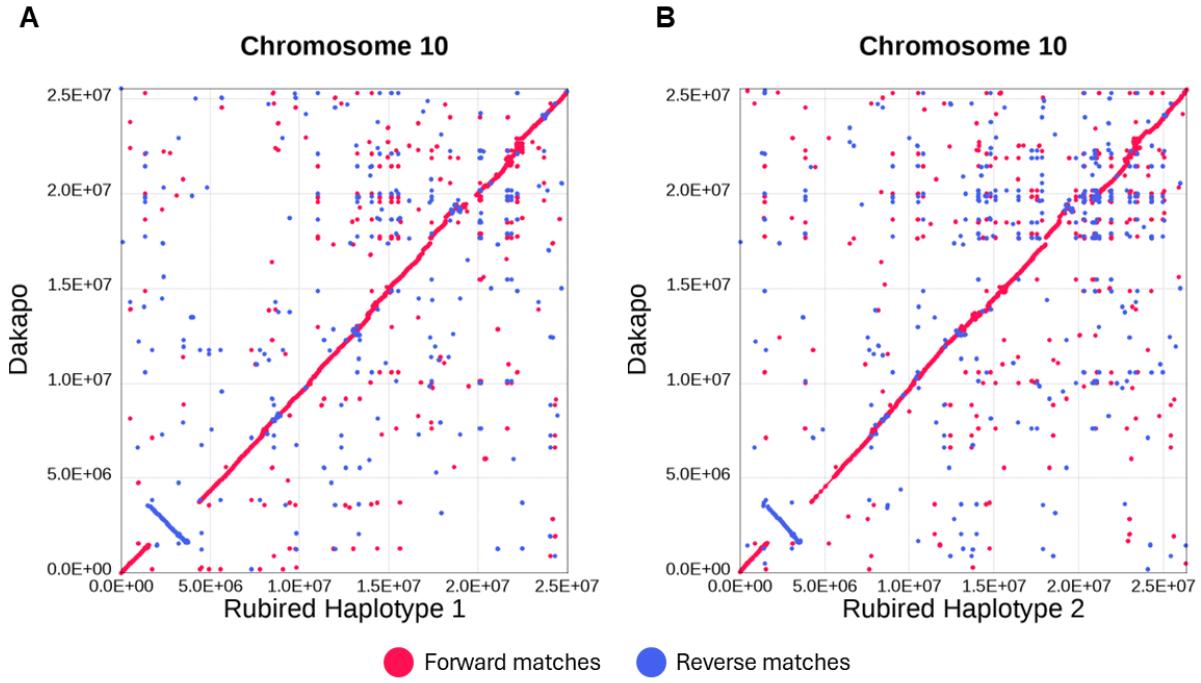

Figure 4. Alignment of chromosome 10 in Dakapo versus chromosome 10 in **A)** the Rubired haplotype 1 assembly and **B)** the Rubired haplotype 2 assembly, showing the 1.82 Mbp inversion present in Dakapo. The dotplots show forward matches in red and reverse matches in blue.

Inversions can cause changes in gene expression depending on various genetic factors [54,87,88], so we were interested in whether the Dakapo chromosome 10 inversion could contribute to Dakapo's increased cold susceptibility and/or increased pathogen susceptibility. Several genes within the inversion do appear to be involved in cold- and/or pathogen-responsive pathways, including VvDak_v1.10g0003381, whose *Arabidopsis* ortholog (AT3G07650) regulates the

expression of genes within the cold acclimation pathway [89], and VvDak_v1.10g0003951, whose *Arabidopsis* and rice orthologs (AT4G03960 and *OsPFA-DSP2*, respectively) negatively regulate pathogen response pathways [90]. The implications of this inversion remain unclear; however, future research could unveil the potential phenotypic impacts of this inversion.

Grapevine is a useful model system due to its unique life and domestication history and is one of few lianas (woody vines) with robust genomic resources. In addition, grapevine breeding and propagation have been ongoing for millennia, resulting in a fascinating array of phenotypes and an abundance of accumulated somatic variants. The assemblies and annotations of the Dakapo and Rubired genomes add to a growing number of grapevine genomes that will provide valuable tools for both grapevine breeders and geneticists.

AVAILABILITY OF SOURCE CODE AND REQUIREMENTS

- **Project name:** Dakapo and Rubired genomes
- **Project home page:** https://github.com/eleanore-ritter/teinturier-grapevine-genomes
- **Operating system:** Platform-independent
- **Programming languages:** bash, perl, python, and R
- **Other requirements:** packages described in methods
- **License:** Apache-2.0

DATA AVAILABILITY

The genomes and annotations for both Dakapo and Rubired are available on grapegenomics.com and Zenodo (10.5281/zenodo.10806332 and 10.5281/zenodo.10543934). Sequencing data from this study are provided on the NCBI Sequence Read Archive under BioProjects PRJNA1094988 and PRJNA1085245.

ADDITIONAL FILES

Supplementary Table S1: Summary of sequencing statistics for reads used to assemble and polish the Dakapo and Rubired genomes.

Supplementary Table S2: SRA IDs of RNA-seq files used for annotating the Dakapo genome.

Supplementary Table S3: Repeat classes of transposable elements and repetitive sequences annotated in the Dakapo genome.

Supplementary Table S4: Repeat classes of transposable elements and repetitive sequences annotated in the diploid Rubired whole genome assembly.

Supplementary Table S5: Genes (including pseudogenes) within the chromosome 10 inversion of the Dakapo grapevine variety.

DECLARATIONS

List of abbreviations

BUSCO: Benchmarking Universal Single-Copy Orthologue

Gbp: Gigabase pairs

GO: Gene ontology

HMMs: Hidden Markov Models

kbp: Kilobase pairs

LAI: LTR Assembly Index

LTRs: Long terminal repeats

Mbp: Megabase pairs

ONT: Oxford Nanopore Technologies

PE: Paired-end

TE: Transposable element

Competing interests

All authors declare no conflict of interest.

Funding

The Dakapo genome work was supported by Michigan State University and the USDA National Institute of Food and Agriculture MICL02572. Oxford Nanopore Technologies provided sequencing for the Dakapo assembly. The Rubired genome was supported by E. & J. Gallo Winery and the NSF grant #1741627.

Authors' contributions

CN envisioned the project, secured the funding, and supervised research on Dakapo. DC envisioned the project, secured the funding, and supervised research on Rubired. EJR assembled and annotated the Dakapo genome, designed and executed comparative analyses within the study, and wrote the first draft of the manuscript. NC and AM assembled and annotated the Rubired genome. PC led Dakapo vine cultivation and tissue sample collection and helped conceptualize comparative analyses within the study. All authors assisted with final drafts of the manuscript.

Acknowledgments

We are grateful to Dan Chitwood, Emily Josephs, and Robin Buell for helpful discussions on this work and feedback on this manuscript. We are grateful for the Genomics Core at Michigan State University, the Institute for Cyber-Enabled Research at Michigan State University, and the UC Davis DNA Technology Core for their services. We would like to thank Kevin Childs for providing guidance and custom scripts for genome annotation. We acknowledge Rosa Figueroa-Balderas for processing the samples, extracting the nucleic acids, and preparing the sequencing libraries for the Rubired genome.

REFERENCES

1. FAO. Agricultural production statistics 2000–2022. FAOSTAT Analytical Briefs, No. 79; 2023.

2. Statistics Department of the International Organisation of Vine and Wine. Annual Assessment of the World Vine and Wine Sector in 2022. International Organisation of Vine and Wine; 2022.

3. Dong Y, Duan S, Xia Q, Liang Z, Dong X, Margaryan K, et al. Dual domestications and origin of traits in grapevine evolution. Science. 2023;379:892–901.

4. Parpinello GP, Versari A, Chinnici F, Galassi S. Relationship among sensory descriptors, consumer preference and color parameters of Italian Novello red wines. Food Res Int. 2009;42:1389–95.

5. Sáenz-Navajas M-P, Echavarri F, Ferreira V, Fernández-Zurbano P. Pigment composition and color parameters of commercial Spanish red wine samples: linkage to quality perception. Eur Food Res Technol. 2011;232:877–87.

6. Flamini R, Mattivi F, De Rosso M, Arapitsas P, Bavaresco L. Advanced knowledge of three important classes of grape phenolics: anthocyanins, stilbenes and flavonols. Int J Mol Sci. 2013;14:19651–69.

7. He F, Mu L, Yan G-L, Liang N-N, Pan Q-H, Wang J, et al. Biosynthesis of anthocyanins and their regulation in colored grapes. Molecules. 2010;15:9057–91.

8. California Department of Food and Agriculture. California Grape Crush Final Report March 10, 2023. 2023.

9. Röckel F, Moock C, Braun U, Schwander F, Cousins P, Maul E, et al. Color Intensity of the Red-Fleshed Berry Phenotype of *Vitis vinifera* Teinturier Grapes Varies Due to a 408 bp Duplication in the Promoter of *VvmybA1*. Genes. 2020;11.

10. Fröhling B, Patz CD, Dietrich H, Will F. Anthocyanins, total phenolics and antioxidant capacities of commercial red grape juices, black currant and sour cherry nectars. Fruit Process. 2012;3:100–4.

11. Kőrösi L, Molnár S, Teszlák P, Dörnyei Á, Maul E, Töpfer R, et al. Comparative Study on Grape Berry Anthocyanins of Various Teinturier Varieties. Foods. 2022;11.

12. Zhang K, Du M, Zhang H, Zhang X, Cao S, Wang X, et al. The haplotype-resolved T2T genome of teinturier cultivar Yan73 reveals the genetic basis of anthocyanin biosynthesis in grapes. Hortic Res. 2023;10:uhad205.

13. Herzog K, Schwander F, Kassemeyer H-H, Bieler E, Dürrenberger M, Trapp O, et al. Towards Sensor-Based Phenotyping of Physical Barriers of Grapes to Improve Resilience to *Botrytis* Bunch Rot. Front Plant Sci. 2021;12:808365.

14. Rashed A, Daugherty MP, Almeida RPP. Grapevine genotype susceptibility to *Xylella fastidiosa* does not predict vector transmission success. Environ Entomol. 2011;40:1192–9.

15. Rashed A, Kwan J, Baraff B, Ling D, Daugherty MP, Killiny N, et al. Relative susceptibility of *Vitis vinifera* cultivars to vector-borne *Xylella fastidiosa* through time. PLoS One. 2013;8:e55326.

16. Wallis CM, Wallingford AK, Chen J. Effects of cultivar, phenology, and *Xylella fastidiosa* infection on grapevine xylem sap and tissue phenolic content. Physiol Mol Plant Pathol. 2013;84:28–35.

17. Ritter EJ, Cousins P, Quigley M, Kile A, Kenchanmane Raju SK, Chitwood DH, et al. From buds to shoots: insights into grapevine development from the Witch's Broom bud sport. BMC Plant Biol. 2024; 24(283).

18. Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genom. 2017;3:e000132.

19. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018;34:2666–9.

20. Andrews S (2010). FastQC: a quality control tool for high throughput sequence data (Version 0.11.9) http://www.bioinformatics.babraham.ac.uk/projects/fastqc

21. De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read sequencing data. Bioinformatics. 2023;39.

22. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37:540–6.

23. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017;27:737–46.

24. Canaguier A, Grimplet J, Di Gaspero G, Scalabrin S, Duchêne E, Choisne N, 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.

25. Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, et al. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biol. 2022;23:258.

26. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.

27. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM . arXiv. 2013.

28. Broad Institute (2017). Picard Toolkit (Version 2.15.0) https://github.com/broadinstitute/picard/releases/tag/2.15.0

29. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9:e112963.

30. Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics. 2018;19:460.

31. Li H. New strategies to improve minimap2 alignment accuracy. Bioinformatics. 2021;37:4572–4.

32. Titus Brown C. Detecting microbial contamination in long-read assemblies (from known microbes). 2018. http://ivory.idyll.org/blog/2018-detecting-contamination-in-long-read-assemblies.html. Accessed 2 May 2024.

33. Titus Brown C, Irber L. sourmash: a library for MinHash sketching of DNA. J Open Source Softw. 2016;1:27.

34. Manni M, Berkeley MR, Seppey M, Simao FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. arXiv. 2021.

35. Pathogen Informatics, Wellcome Sanger Institute (2020). assembly-stats (Version 1.0.1) https://github.com/sanger-pathogens/assembly-stats/releases/tag/v1.0.1-docker1

36. Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 2019;20:275.

37. Perazzolli M, Moretto M, Fontana P, Ferrarini A, Velasco R, Moser C, et al. Downy mildew resistance induced by *Trichoderma harzianum* T39 in susceptible grapevines partially mimics transcriptional changes of resistant genotypes. BMC Genomics. 2012;13:660.

38. Da Silva C, Zamperin G, Ferrarini A, Minio A, Dal Molin A, Venturini L, et al. The high polyphenol content of grapevine cultivar tannat berries is conferred primarily by genes that are not shared with the reference genome. Plant Cell. 2013;25:4777–88.

39. Minio A, Massonnet M, Figueroa-Balderas R, Vondras AM, Blanco-Ulate B, Cantu D. Iso-Seq Allows Genome-Independent Transcriptome Profiling of Grape Berry Development. G3 . 2019;9:755–67.

40. Vannozzi A, Palumbo F, Magon G, Lucchin M, Barcaccia G. The grapevine (*Vitis vinifera* L.) floral transcriptome in Pinot noir variety: identification of tissue-related gene networks and whorl-specific markers in pre- and post-anthesis phases. Hortic Res. 2021;8:200.

41. Daldoul S, Hanzouli F, Hamdi Z, Chenenaoui S, Wetzel T, Nick P, et al. The root transcriptome dynamics reveals new valuable insights in the salt-resilience mechanism of wild grapevine (*Vitis vinifera* subsp. *sylvestris*). Front Plant Sci. 2022;13:1077710.

42. Ma S-H, He G-Q, Navarro-Payá D, Santiago A, Cheng Y-Z, Jiao J-B, et al. Global analysis of alternative splicing events based on long- and short-read RNA sequencing during grape berry development. Gene. 2023;852:147056.

43. SRA Toolkit Development Team (2020). sra-tools (Version 2.10.7) https://github.com/ncbi/sra-tools/releases/tag/2.10.7

44. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37:907–15.

45. Shumate A, Wong B, Pertea G, Pertea M. Improved transcriptome assembly using a hybrid of long and short reads with StringTie. PLoS Comput Biol. 2022;18:e1009730.

46. Chen M-JM, Lin H, Chiang L-M, Childers CP, Poelchau MF. The GFF3toolkit: QC and Merge Pipeline for Genome Annotation. Methods Mol Biol. 2019;1858:75–87.

47. Cheng C-Y, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD. Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. Plant J. 2017;89:789–804.

48. Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. Rice. 2013;6:4.

49. UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023;51:D523–31.

50. Slater GSC, Birney E. Automated generation of heuristics for biological sequence

comparison. BMC Bioinformatics. 2005;6:31.

51. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 2011;12:491.

52. Korf I. Gene finding in novel genomes. BMC Bioinformatics. 2004;5:59.

53. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 2008;24:637–44.

54. Kollar LM, Stanley LE, Raju SKK, Lowry DB, Niederhuth CE. The role of breakpoint mutations, supergene effects, and ancient nested rearrangements in the evolution of adaptive chromosome inversions in the yellow monkey flower, *Mimulus guttatus*. bioRxiv. 2023; doi:10.1101/2023.12.06.570460.

55. Velt A, Frommer B, Blanc S, Holtgräwe D, Duchêne É, Dumas V, et al. An improved reference of the grapevine genome reasserts the origin of the PN40024 highly homozygous genotype. G3. 2023;13.

56. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.

57. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12:59–60.

58. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. 2012;40:D1202–10.

59. Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, et al. The Arabidopsis information resource: Making and mining the “gold standard” annotated reference plant genome. Genesis. 2015;53:474–85.

60. Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, et al. Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods. 2016;13:1050–4.

61. Minio A, Cochetel N, Massonnet M, Figueroa-Balderas R, Cantu D. HiFi chromosome-scale diploid assemblies of the grape rootstocks 110R, Kober 5BB, and 101–14 Mgt. Scientific Data. 2022;9:1–8.

62. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021;18:170–5.

63. Minio A, Cochetel N, Vondras AM, Massonnet M, Cantu D. Assembly of complete diploid-phased chromosomes from draft genome sequences. G3. 2022;12.

64. Zou C, Karn A, Reisch B, Nguyen A, Sun Y, Bao Y, et al. Haplotyping the *Vitis* collinear

core genome with rhAmpSeq improves marker transferability in a diverse genus. Nat Commun. 2020;11:413.

65. Cochetel N, Minio A, Massonnet M, Vondras AM, Figueroa-Balderas R, Cantu D. Diploid chromosome-scale assembly of the *Muscadinia rotundifolia* genome supports chromosome fusion and disease resistance gene expansion during *Vitis* and *Muscadinia* divergence. G3. 2021;11.

66. Cochetel N, Minio A, Guarracino A, Garcia JF, Figueroa-Balderas R, Massonnet M, et al. A super-pangenome of the North American wild grape species. Genome Biol. 2023;24:290.

67. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.

68. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.

69. Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, et al. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. Elife. 2022;11.

70. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40:e49.

71. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:238.

72. Minio A, Cochetel N, Figueroa-Balderas R, Cantu D. Grapegenomics.com - Genome release: *Vitis vinifera* cv. Chardonnay cl. 04 v2.0. 2024. Zenodo. doi:10.5281/zenodo.10578344.

73. Cantu Lab. Vitis vinifera cv. Pinot Noir cl. FPS123. grapegenomics.com. 2024. https://www.grapegenomics.com/pages/VvPinNoir/VvPinNoir123/.

74. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A fast and versatile genome alignment system. PLoS Comput Biol. 2018;14:e1005944.

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

76. Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. Bioinformatics. 2021;37:1639–43.

77. Zhou Y, Minio A, Massonnet M, Solares E, Lv Y, Beridze T, et al. The population genetics of structural variants in grapevine domestication. Nat Plants. 2019;5:965–79.

78. Minio A, Massonnet M, Figueroa-Balderas R, Castro A, Cantu D. Diploid Genome Assembly of the Wine Grape Carménère. G3 . 2019;9:1331–7.

79. Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A, et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature. 2007;449:463–7.

80. Kobayashi S, Goto-Yamamoto N, Hirochika H. Retrotransposon-induced mutations in grape skin color. Science. 2004;304:982.

81. Doligez A, Bouquet A, Danglot Y, Lahogue F, Riaz S, Meredith C, et al. Genetic mapping of grapevine (*Vitis vinifera* L.) applied to the detection of QTLs for seedlessness and berry weight. Theor Appl Genet. 2002;105:780–95.

82. Walker AR, Lee E, Bogs J, McDavid DAJ, Thomas MR, Robinson SP. White grapes arose through the mutation of two similar and adjacent regulatory genes. Plant J. 2007;49:772–85.

83. Azuma A, Kobayashi S, Goto-Yamamoto N, Shiraishi M, Mitani N, Yakushiji H, et al. Color recovery in berries of grape (*Vitis vinifera* L.) “Benitaka”, a bud sport of “Italia”, is caused by a novel allele at the *VvmybA1* locus. Plant Sci. 2009;176:470–8.

84. Matus JT, Cavallini E, Loyola R, Höll J, Finezzo L, Dal Santo S, et al. A group of grapevine MYBA transcription factors located in chromosome 14 control anthocyanin synthesis in vegetative organs with different specificities compared with the berry color locus. Plant J. 2017;91:220–36.

85. Lisek J. Winter frost injury of buds on one-year-old grapevine shoots of cultivars and interspecific hybrids in Poland. Folia Horticulturae. 2012;24:97–103.

86. Doster MA. Effects of leaf maturity and cultivar resistance on development of the powdery mildew fungus on grapevines. Phytopathology. 1985;75:318.

87. Puig M, Casillas S, Villatoro S, Cáceres M. Human inversions and their functional consequences. Brief Funct Genomics. 2015;14:369–79.

88. Loveland JL, Lank DB, Küpper C. Gene Expression Modification by an Autosomal Inversion Associated With Three Male Mating Morphs. Front Genet. 2021;12:641620.

89. Li Y, Shi Y, Li M, Fu D, Wu S, Li J, et al. The CRY2-COP1-HY5-BBX7/8 module regulates blue light-dependent cold acclimation in Arabidopsis. Plant Cell. 2021;33:3555–73.

90. He H, Su J, Shu S, Zhang Y, Ao Y, Liu B, et al. Two homologous putative protein tyrosine phosphatases, OsPFA-DSP2 and AtPFA-DSP4, negatively regulate the pathogen response in transgenic plants. PLoS One. 2012;7:e34995.