Hindawi
Australian Journal of Grape and Wine Research
Volume 2023, Article ID 6686706, 14 pages
https://doi.org/10.1155/2023/6686706

AUSTRALIAN SOCIETY OF VITICULTURE AND OENOLOGY | ASVO

Research Article

The Genome Assembly of *Vitis vinifera* cv. Shiraz

Cristobal A. Onetto,[1] **Christopher M. Ward**,[1] **and Anthony R. Borneman**[1,2]

[1]*The Australian Wine Research Institute, Glen Osmond, Urrbrae, South Australia 5064, Australia*
[2]*School of Agriculture, Food and Wine, The University of Adelaide, Adelaide, South Australia 5005, Australia*

Correspondence should be addressed to Anthony R. Borneman; anthony.borneman@awri.com.au

Received 12 December 2022; Revised 15 March 2023; Accepted 29 April 2023; Published 11 May 2023

Academic Editor: Amber Parker

Background and Aims. Shiraz (Syrah) is a dark-skinned cultivar of the wine grape *Vitis vinifera* that forms the basis of some of the world's most iconic wines. Worldwide, Shiraz is the fourth most planted grapevine cultivar; however, it represents the most planted cultivar in Australia. Given the importance of Shiraz to worldwide wine production, this study aimed to produce a reference genome for this cultivar while investigating the unique genetic variants and ancestral origins of this iconic variety. *Methods and Results*. Long-read ONT data were selected to produce a highly contiguous genome assembly for Shiraz. Phylogenetic reconstruction using high-quality genome assemblies for wine grape cultivars provided further support of a kinship between Shiraz and Pinot Noir. Harnessing long-read data, transposable element insertions potentially affecting gene function were characterized in Shiraz and assessed relative to other cultivars. This revealed a heterogenous landscape of transposon insertion points across cultivars and uncovered a specific combination of allelic variants at the *VviTPS24* terpene synthase locus. *Conclusions*. This establishment of a Shiraz genome provides a detailed view of the genetics that underpin this cultivar, including the discovery of a specific combination of *VviTPS24* variants, which when combined with appropriate environmental triggers may allow Shiraz to produce high levels of rotundone, the aroma compound responsible for the distinctive peppery characteristics of this cultivar. *Significance of the Study*. The availability of a reference genome for Shiraz expands the pool of genomes available for wine grapes while providing a foundation resource for whole-genome studies involving this iconic cultivar, including intracultivar variant identification and transcriptomic studies using a matching reference genome.

1. Introduction

Shiraz (Syrah) is a dark-skinned cultivar of the wine grape *Vitis vinifera*, which is used to create some of the world's most iconic red wines. There have been many theories surrounding the history of Shiraz, including a potential origin in the city of Shiraz in ancient Persia (now a part of Iraq). However, previous DNA-based marker analysis has identified Shiraz as the offspring of the grape cultivars Dureza (dark-skinned) and Mondeuse Blanche (white-skinned), two-cultivars that are considered native to the northern Rhône in the south-east of France [1]. It is therefore likely that Shiraz also originated in this geographic location through a natural outcrossing event, which may date back to Roman times.

Globally, Shiraz is the fourth-most planted grapevine cultivar in the world. France contains the largest plantings of Shiraz, where it represents the third most commonly planted wine grape. In Australia, Shiraz is the most widely planted cultivar, with 40,000 hectares of vines, positioning the country as second only to France in worldwide plantings of Shiraz. Australia is also home to many of the oldest Shiraz vineyards in the world, with many vines that predate the devastation of grapevine phylloxera on European vineyards in the 1800s.

One of the trademark flavours of Shiraz, and especially of Shiraz grapes grown in cool climates, is black pepper [2]. These peppery notes have been attributed to the presence of the highly-potent sesquiterpene rotundone, with reported detection thresholds as low as 16 ng/L in red wine and 8 ng/L in water [3]. Although being reported in comparatively

higher concentrations in Shiraz, rotundone has also been detected in several other cultivars, with Duras, Vespolina, and Grüner Veltliner also displaying relatively high concentrations of this compound [4]. While the complete biosynthetic pathway of rotundone has not been elucidated, it has been shown that the sesquiterpene synthase *VviTPS24* is responsible for the biosynthesis of the precursor of rotundone, α-guaiene [5]. Rotundone can then be formed from α-guaiene either by simple oxidation or enzymatically through the cytochrome P450 α-guaiene 2-oxidase [6, 7].

There are thousands of distinct cultivars of *V. vinifera* that are used for wine production, which display extensive phenotypic diversity. Given the economic importance of this species, genome sequencing is being used to determine the genetic differences that separate the various types of wine grapes. Early efforts in the production of reference genomes for *V. vinifera* were confounded by high levels of heterozygosity and hemizygosity [8], such that inbreeding was used to produce a homozygous line derived from Pinot Noir for initial attempts at assembling a complete grapevine genome [9]. Advances in "long-read" sequencing and phased genome assembly algorithms have now allowed for the production of highly-contiguous assemblies for the grapevine cultivars Chardonnay [10, 11], Cabernet Sauvignon [12], Carménère [13], Zinfandel (syn Primativo) [14], Nebbiolo [15], Cabernet Franc [16], Riesling [17], and Merlot [18]. These studies have expanded the knowledge on the mechanisms of genome evolution in this species, highlighting the importance of structural variants and repetitive elements as drivers of cultivar and clonal phenotypic diversity.

Given the importance of Shiraz to worldwide wine production, a reference genome assembly for this cultivar is required. Long-read data were selected to produce a highly contiguous diploid genome assembly for Shiraz, which could provide the basis for detailed phylogenetic investigations and to compare structural variations across Shiraz and other cultivars for which high-quality, phased genomes were available. Overall, this study aims to provide a resource for future comparative genomics of grapes with implications for diploid genome evolution, the determination of clonal variation, and historical provenance of *V. vinifera* L. varieties.

2. Materials and Methods

2.1. Sampling, DNA Extraction, and Whole Genome Sequencing. DNA was extracted from early-season *V. vinifera* Shiraz clone 1654 leaves taken from field-grown plants at the Coombe Vineyard (Waite Campus, University of Adelaide, Adelaide, Australia). Samples were immediately frozen and ground to powder in liquid nitrogen. Approximately 100 mg of plant material was used for DNA extraction using a DNeasy Plant Mini Kit (QIAGEN, Australia), following the manufacturer's instructions. Prior to Oxford Nanopore Technologies (ONT) sequencing library preparation, high-molecular weight DNA >10 kb was enriched using a Short Read Elimination Kit SRE XS (Pacific Biosciences, CA, USA). Sequencing libraries were prepared using the SQK-LSK110 kit and loaded into two FLO-MIN106 and one FLO-MIN111 flow cells. Fast5 files were base-called using Guppy v. 5.0.16 (Oxford Nanopore Technologies, Oxford, UK) with the "sup" model and a minimum quality score filtering of 7. A total sequencing yield of 30,663 Mb was obtained (63-fold coverage) with an N_{50} length of 21.8 kb. For short-read sequencing, genomic libraries were prepared using the Illumina DNA Prep library kit and sequenced on a Novaseq 6000 instrument using an S4 flow cell and 2×150 bp chemistry (Ramaciotti Centre for Genomics, NSW, Australia). A total of 97 million read-pairs were obtained (64-fold coverage).

For cultivar Sauvignon Blanc, DNA was extracted from early-season leaves taken from a field-grown plant of clone F4V6 located at the South Australian Research and Development Institute Nuriootpa Research Centre (Nuriootpa, Barossa Valley, Australia). Samples were ground to powder in liquid nitrogen, and nuclei were isolated following protocol 102-574-800 (Pacific Biosciences, CA, USA). DNA was extracted from nuclei using the Nanobind plant nuclei kit (Pacific Biosciences, CA, USA), following the manufacturer's instructions. High-molecular weight DNA >10 kb was enriched using a Short Read Elimination Kit SRE XS (Pacific Biosciences, CA, USA). Sequencing libraries were prepared using the SQK-LSK112 kit and loaded into two FLO-MIN112 flow cells. Fast5 files were base-called using Guppy v. 6.4.2 (Oxford Nanopore Technologies, Oxford, UK) with the "sup" model and a minimum quality score filtering of 7. A total sequencing yield of 19,655 Mb was obtained (43-fold coverage). For short-read sequencing, genomic libraries were prepared using DNA from clone F4V6 with the Illumina DNA Prep library kit and sequenced on a Novaseq 6000 instrument using an S4 flow cell and 2×150 bp chemistry (Ramaciotti Centre for Genomics, NSW, Australia).

2.2. Genome Assembly and Annotation. Preliminary assemblies for Shiraz were performed using Canu v. 2.1.1 [19] and Flye v. 2.8.3 [20] and then polished with ONT reads using Medaka v. 1.5.0 (https://github.com/nanoporetech/medaka). Both assemblies were then combined using quickmerge v. 0.3 [21] with the Canu assembly as reference and a minimum overlap of 20 kb, and polished twice using short-reads and Pilon v. 1.24 [22]. Lastly, allelic contig reassignment was performed using Purge Haplotigs v. 1.1.2 [23] and assessed with BUSCO v. 5.3.2 [24] using the embryophyta ODB v10 database. The same methodology was applied for the genome assembly of the cultivar Sauvignon Blanc using the assemblers Canu v. 2.1.1 [19] and SMARTdenovo [25]. A scaffolded version of the Shiraz primary assembly was created for visualization purposes using the *V. vinifera* reference genome (accession GCA_000003745.2) and RagTag v. 2.1.0 [26].

A custom repeat library was built for Shiraz using RepeatModeler v. 2.0.3 [27] and the LTR pipeline extension, which applies LtrHarvest and Ltr_retriever [28] during *de novo* repeat identification. Identification of miniature inverted-repeat transposable elements (MITEs) was performed using MITE-Tracker [29]. The custom repeat

sequences were combined into a single library and used for repeat annotation using RepeatMasker v. 4.1.2 [30]. Gene prediction was performed following the funannotate pipeline v. 1.8.13 [31], including Genemark-ES v. 4.68 [32], SNAP [33], Augustus v. 3.3.3 [34], and Glimmerhmm v. 3.0.4 [35] annotations, allowing a maximum intron length of 10 kb. Previously published RNA-seq data for Shiraz ([36], Table S1) and the protein data of the *V. vinifera* reference genome (accession GCA_000003745.2) were provided as evidence for gene model prediction.

Homo and hemizygous regions were investigated by mapping short-read data to the primary assembly. Heterozygous SNPs were called using VarScan v. 2.3 [37] and read-depth and SNP density calculated in 50 kb windows (25 kb steps) using BEDTools v. 2.30.0 [38].

2.3. Phylogenetics and Identity by Descent (IBD). Single-copy orthologs (SCOs) were identified in Nebbiolo, Chardonnay, Carménère, Zinfandel, Cabernet Sauvignon, Cabernet Franc, Riesling, and Merlot along with *V. vinifera sylvestris* and *V. rotundifolia*, using the BUSCO eudicot dataset. Alignment was then carried out using MUSCLE [39]. To ensure that errors in annotation do not bias phylogenetic reconstruction, each alignment was manually investigated and trimmed to remove mis-annotated exons between transcripts through the excision of regions representing insertions and/or deletions that were not present in any other sample or sequences that represented unaligned extensions to the start or end of the gene. Each gene alignment was imported into BEAST2 [40] with unlinked Hasegawa-Kishino-Yano site and relaxed log normal clock [41] model priors. A MCMC chain was then run across 1×10^8 samples with a yule tree prior. Tracer was used to identify when the model began to mix and select an appropriate burn-in (48%).

Previously published RAD-Seq lllumina paired-end reads from cultivars Dureza and Mondeuse Blanche [42] were mapped to the Shiraz primary assembly using Minimap2 v. 2.24 [43]. Mapped reads had their variants called (MQ > 20) and were filtered (DP > 5, GQ > 20, F_MISS = 0) using BCFtools v. 1.16 [44]. Heterozygous sites in Shiraz were further filtered to remove any sites where allele depth did not fit a binomial distribution, thereby removing somatic mutations developed after the cross event between Dureza and Mondeuse blanche. Putative alleles that suffered from dropout due to mutations in the RADtag cut site were identified by inspecting the first four base pairs of both the forward and reverse reads for mutations within Shiraz with geaR v. 0.1 [45]. Any tags that contained mutations were removed from the analysis space. Filtered variants were converted to the GDS format and IBD calculated using SNPrelate v. 1.32 [46].

2.4. Characterization of TE Content and Comparative Genomics. Read depth and structural variant information were leveraged using plyranges v. 1.18 [47] and geaR v. 0.1 [45] to collapse fragmented transposable element annotations into a single record. First, read depth was calculated against the Shiraz primary assembly across the middle 10 bp of a TE annotation using SAMtools v. 1.16.1 [44]. This was then compared to the median read depth observed across surrounding coding regions and overlapped with structural variants called from long-read nanopore data using Sniffles v. 2.0.2 [48] to determine zygosity. Homozygous annotations were then conditionally merged if adjacent annotations from the same class were both contained within the same read and were also homozygous. Heterozygous annotations were then compared to overlapping heterozygous structural variants. Adjacent transposable element annotations that were both heterozygous and contained within a single structural variant were merged into one record.

Transposable element annotations within introns and 1 kb of an orthologs TSS or stop codon were extracted and intersected to identify putative genic TE insertions. Exonic insertions points were identified by first extracting gene and CDS features from the Pinot Noir reference [9] genome and annotation sourced from Ensemble Plants (PN40024.v4).

Putative genic TEs were then leveraged against Nebbiolo, Chardonnay, Carménère, Zinfandel, Cabernet Sauvignon, Sauvignon Blanc, Riesling, and Merlot long-read data (see Table S1 accession number) structural variants using Sniffles v. 2.0.2 [48].

Publicly available short-read data from 23 common cultivars (see Table S1 for accession number) was used to determine if the TE insertion in *VviTPS24* is present in other cultivars. Data was first inspected for quality using ngsReports v. 2.0.1 [49], mapped to the Shiraz primary assembly using BWA-mem v. 0.7.17 [50], and filtered to remove any reads with MQ below 20 using SAMtools v. 1.16.1 [44]. Both the upstream and downstream breakpoints of the TE insertion were manually inspected for reads that contained both the *VviTPS24* coding region and TE sequence and for paired reads whose insert spanned the TE insertion point (i.e. reads that mapped to *VviTPS24* exon 5 whose mate mapped to the TE). For a sample to be considered to contain the TE, evidence must have been found at both the upstream and downstream breakpoints. The zygosity of the TE in each variety was then confirmed using sliding 31 mers of each breakpoint and the reconstructed functional allele with Jellyfish v. 2.3 [51].

3. Results and Discussion

3.1. Haplotype Phased Assembly of the Cultivar Shiraz. Recent advances in long-read DNA sequencing technologies have enabled the creation of high-quality, diploid genome assemblies for repeat-rich and highly heterozygous plant species such as *V. vinifera* [12]. Haplotype-phased genome assemblies have been produced for a handful of the most widely planted cultivars, including Cabernet Sauvignon [12], Merlot [18], and Chardonnay [10, 11]; however, there is no publicly available reference genome for Shiraz to date, despite it representing the fourth-most planted cultivar in the world. To address this knowledge gap, a reference genome for the cultivar Shiraz was produced using a hybrid sequencing approach that included 63-fold coverage of ONT long reads and 64-fold coverage of Illumina short-reads.

Clone 1654 was selected as the sourced plant material due to its widespread use in the Australian wine industry.

After assembly, polishing, and haplotype phasing, a 476 Mb primary assembly with an N_{50} of 1.9 Mb was obtained, falling well within the expected haploid size for *V. vinifera* and only 2% smaller than the inbred Pinot Noir reference genome (PN40024). While larger primary assemblies have been obtained for other wine cultivars (Cabernet Sauvignon: 590 Mb, Cabernet Franc: 570 Mb, Merlot: 606 Mb, Chardonnay: 490 Mb, Carménère: 623 Mb, Nebbiolo: 561 Mb, and Zinfandel: 591 Mb), not all of the reported grapevine assemblies have been processed with tools to optimize the reassignment of allelic contigs, and the larger sizes are likely due to the retention of both copies of highly heterozygous regions within the primary assembly contig pool.

Haplotype phasing generated a total of 356 Mb of associated haplotigs with an N_{50} of 199 kb (Table 1). The primary assembly included 95.1% of BUSCO orthologues and contained 56.1% repetitive content, which was represented primarily by gypsy and copia LTR retroelements (Table 1). After removal of repeat-associated gene models a total of 32,333 protein-coding genes were retained (Table 1).

To assess the degree and distribution of hemizygosity and homozygosity across the Shiraz genome, information from read-depth and heterozygous variant density were assessed across both the primary and scaffolded assemblies (Figure 1). Selection of regions characterized by half the median read-depth and low heterozygous variant density highlighted 54 Mb (10.5%) of the Shiraz assembly as hemizygous, with these regions predicted to encode 3017 genes (Figure 1). Similar genome wide hemizygosity levels have been reported in distantly-related cultivars such as Cabernet Sauvignon (15.5% of genes in primary assembly) and Chardonnay (14.6% of genes in primary assembly) [11], suggesting a basal level of hemizygosity that separates pairs of parental alleles in *V. vinifera*.

The functional consequences of hemizygous regions were assessed through gene ontology functional enrichment. This revealed an overrepresentation of several functional classes, including chitinases that form part of the systemic acquired resistance mechanism of *V. vinifera* [52] and terpene synthases (Table S2). This is not surprising as previous comparative genomic studies have suggested that hemizygosity and variation in gene content have a potential contribution towards the phenotypic differences between cultivars [11, 15].

Homozygous regions, which were categorized as areas of median read-depth and low heterozygous variant density, comprised 11.7% of the primary assembly. The longest run of homozygosity was a stretch of 4.8 Mb located at one end of chromosome 9 (Figure 1). In comparison, Chardonnay, which has been suggested to be a naturally inbred cultivar [10], was shown to contain twice the levels of homozygosity (22.4%) as Shiraz, providing support that the parental cultivars of Shiraz do not share a recent common ancestor.

Table 1: Shiraz assembly statistics.

	Primary assembly	Haplotigs
Assembly size (bp)	476,422,955	356,328,851
Contigs	435	2,408
N_{50}	1,969,387	199,587
Largest contig	7,765,998	2,855,789
Predicted proteins	32,333	22,725
Repetitive content (%)	56.1	62.3
Predicted hemizygous (bp)	50,426,450	
Predicted hemizygous genes	3,017	
Predicted homozygous (bp)	55,847,349	
Complete BUSCOs	1536 (95.1%)	
Complete and single-copy BUSCOs	1466 (90.8%)	
Complete and duplicated BUSCOs	70 (4.3%)	

The density of genes (Figure 1(c)) and LTR retrotransposons (Figure 1(d)) displayed a distinctive pattern whereby gene density decreased, with a concomitant increase in LTR density surrounding centromeric regions. However, a clear spike in LTR density was also observed outside of the centromeric region on Chr 10 (Figure 1(d)), suggesting a nested transposable element insertion region. Detailed annotation of this LTR repeat domain rich region on Chr 10 identified 29 separate LTR repeat domains and 127 internal domains, which were distributed across an 88 kb region (Figure S1). Long-terminal repeat insertions appeared to be nested within the internal domains, with multiple LTR domains occurring within small (<2 kb) windows. Analysis of the mapped reads (Figure S2) showed no indication of a mis-assembly across the region, with long reads spanning multiple nested LTR domains.

3.2. Phylogenetic Reconstruction and Parentage of Shiraz. The availability of several long-read grapevine genomes offered the ability to assess phylogenetic relatedness that encompassed information inherent within both phased alleles across the diploid genome. SCOs were identified across the ten long-read *V. vinifera* genomes, in addition to *V. vinifera sylvestris* and *V. rotundifolia* as outgroups. Phylogenetic reconstruction using these 437 SCOs revealed Pinot Noir as the closest relative of Shiraz (Figure 2), providing further support for their proposed kinship [53]. Cabernet Franc and its offspring Cabernet Sauvignon, Carménère, and Merlot (Figure 2) were closely related, yet clade branching placed Cabernet Franc as the most derived and Merlot as more closely related to Riesling than the Cabernet clade. This suggests Merlot and Riesling may share a close relative as alleles from the non-Cabernet Franc parental haplotype may be influencing the topology. The observed topology is likely due to the "primary" assembly of each of the cultivars containing the most contiguous reference allele, producing pseudohaplotypes that represent a random blending of the two true parental haplotypes. Therefore, care should be taken in future studies when interpreting the trees containing closely related crop plants

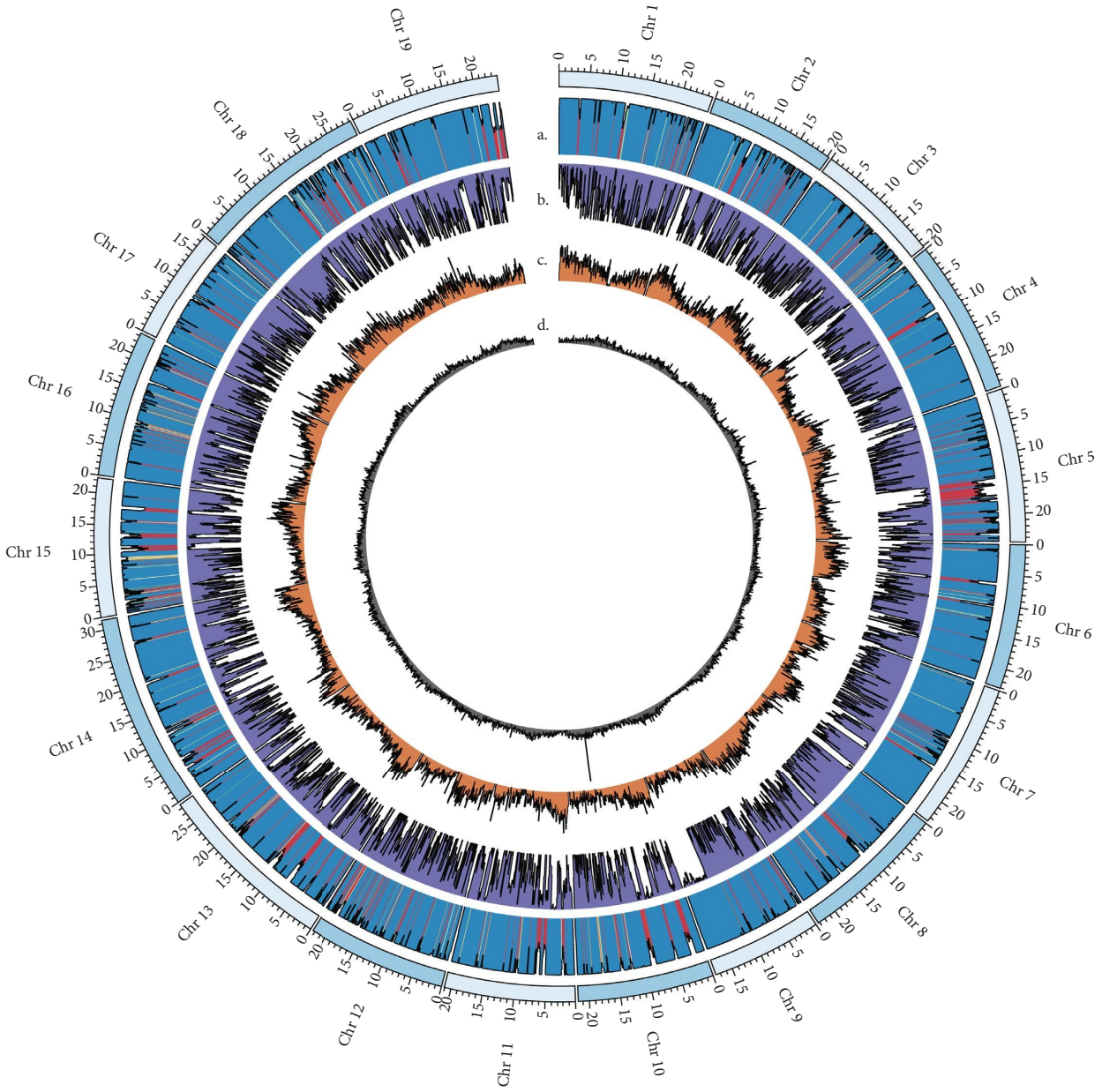


FIGURE 1: Genome assembly of *V. vinifera* cv. Shiraz. Circos plot depicts chromosome-scaffolded primary assembly using the Pinot Noir 12X reference genome (accession GCF_000003745.3). (a) Read-depth of Illumina reads mapped to the primary assembly with color scale ranging from ≥ median read-depth (blue) to half median read-depth (red), (b) heterozygous variants, (c) gene, and (d) LTR retrotransposon density.

with mixed ancestry unless true haplotypes can be resolved and not relying in algorithmic phasing.

As SSR data was previously used to suggest that Mondeuse Blanche and Dureza comprise the parents of Shiraz [1], existing RAD-Seq data from these two cultivars [42] was utilized to provide further support to their relationship to Shiraz. Whole genome variant calling was performed using the RAD-Seq data of Mondeuse Blanche and Dureza against the Shiraz genome sequence, producing a space of 81,551 variants for analysis. Quality filtering, removing calls within annotated repeats, and RAD allele dropout decreased the total number of useable genotypes to 22,358 for kinship estimation. During filtering, allelic ratio was used to calculate a binomial probability (np = 0.5) at each heterozygous Shiraz variant to remove variants that may be somatic from the Shiraz genotypes. Kinship estimation and identity by descent (IBD) calculation was then carried out using the MOM method. A kinship matrix consistent with a parent-offspring relationship was calculated for Dureza and Shiraz (IBD = 0.25, (k_0 0, k_1 1)). However, values that would be considered consistent with a true parent-offspring relationship were not recovered for Mondeuse Blanche (IBD = 0.192, (k_0 0.23, k_1 0.77)). This suggests that

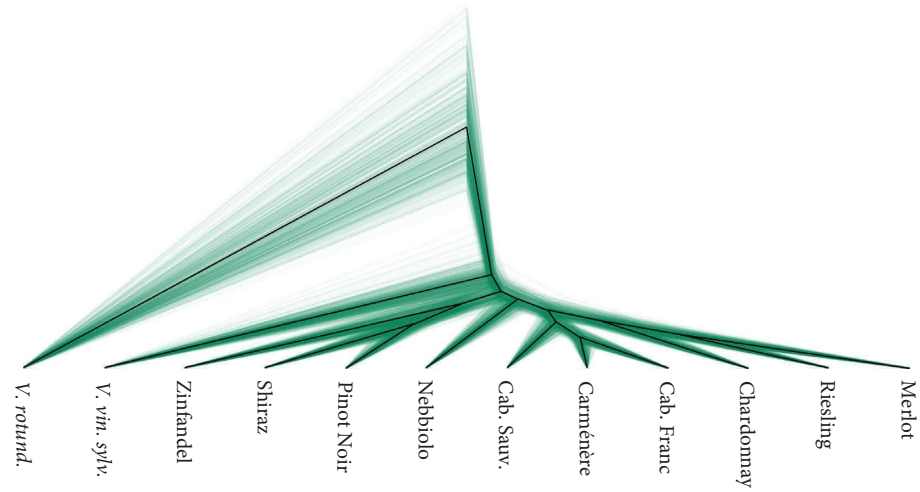


Figure 2: Bayesian phylogeny of Shiraz and other common cultivars with *Vitis vinifera sylvestris* (*V. vin. sylv.*) and *Vitis rotundifolia* (*V. rotund.*) used as outgroups. Cab. Sauv.: Cabernet Sauvignon, Cab. Franc: Cabernet Franc.

the relationship between Mondeuse Blanche and Shiraz may be more complex than previously estimated using SSR markers.

3.3. Repeat Characterization in Shiraz Reveals TEs near Genes Absent from Other Cultivars. Somatic variation and specifically transposable element (TE) insertions have been shown to drive adaptation [54, 55] and to provide an important source of genetic diversity for trait selection and breeding in clonally propagated crop plants [56, 57]. Disruption of proper gene function via the insertion of TEs can either occur directly through disruption of the open reading frame or indirectly through influencing transcription by insertion into regulatory regions or through chromatin availability and epigenetic silencing [58–60]. In *V. vinifera*, TE insertions have been identified as the causative mutation that underpins important phenotypes such as differences in berry color [61], which have convergently occurred across multiple lineages [61, 62] through insertion of the Gret1 LTR into the promoter region of *VviMYBA1* [61, 63].

Genome-wide characterization of TE content identified 125,736 TE annotations in the Shiraz primary assembly (Figure 3(a)). Long terminal repeats (LTR) comprised the most frequent TE class, with Copia and Gypsy LTRs representing 50.6% of all annotated TEs. LINE elements (17.8%) were the third most common class, followed by MITEs (14%) (Figure 3(a)).

As TE insertions within genic regions can impact gene function, these were mapped across the Shiraz genome and compared to those in other cultivars. First, the Shiraz gene and repeat annotations were utilized to identify TEs within 1 kb of either the 5′ or 3′ termini of each gene model, in addition to those within intronic regions. This identified 10,570 TE annotations upstream (Table S3) and 10,223 annotations downstream (Table S3), which may affect gene regulation (Figure 3(b)). In contrast to the genome-wide data, in which LTR elements are the most frequent TE across the genome, MITE elements were most commonly observed upstream (24.2%) and downstream (21.6%) of genes, agreeing with past studies [58, 64]. LINE insertions were the most frequently observed TE within intronic regions, comprising 52.9% of the 24,721 intronic TEs (Table S3).

Insertions of TEs within exons would evade detection by the previous methodology, as it is likely that exonic insertion would interfere with correct gene annotation within Shiraz. To overcome this limitation, a secondary methodology was applied in which gene annotations were extracted from the Ensembl Pinot Noir reference genome entry and mapped to the Shiraz primary assembly. This identified a total of 31,839 putative gene annotations, which were then overlapped with the Shiraz TE annotations, revealing 83 potential exonic insertions in Shiraz (Table S3) (Figure 3(b)).

To identify genic TE insertions that are variable between cultivars or specific to Shiraz, structural variants called from long-read data of Cabernet Sauvignon, Chardonnay, Carménère, Merlot, Nebbiolo, Riesling, Shiraz, Sauvignon Blanc, and Zinfandel were cross-referenced against the set of putative genic TEs. This identified 4,554 genic TE insertions variable between these cultivars (Table S4), which may contribute to their phenotypic diversity. Shiraz-specific TE insertions were also identified (Table S5), 34 upstream (LTR: 82.3%; MITE: 11.8%; and LINE: 5.9%) (Figure 3(c) (i)), 28 downstream (LTR: 92.9% and MITE: 7.1%) (Figure 3(c) (ii)), 45 intronic (LTR: 55.6%; MITE: 2.2%; and LINE: 42.2%) (Figure 3(c) (iii)), and 6 exonic (LTR: 100%) (Figure 3(c) (iv)). Furthermore, the majority of Shiraz-specific TEs were in the heterozygous state (88.6%).

3.4. The VviTPS24 Locus Comprises a Distinct Genotype in the Cultivar Shiraz. From the few characterized enzymes involved in the production of aroma compounds or their precursors, terpene synthases have received particular attention due to their role in the biosynthesis of volatile terpenoids that define the varietal characters of several grape cultivars [65, 66]. The potent bicyclic sesquiterpene

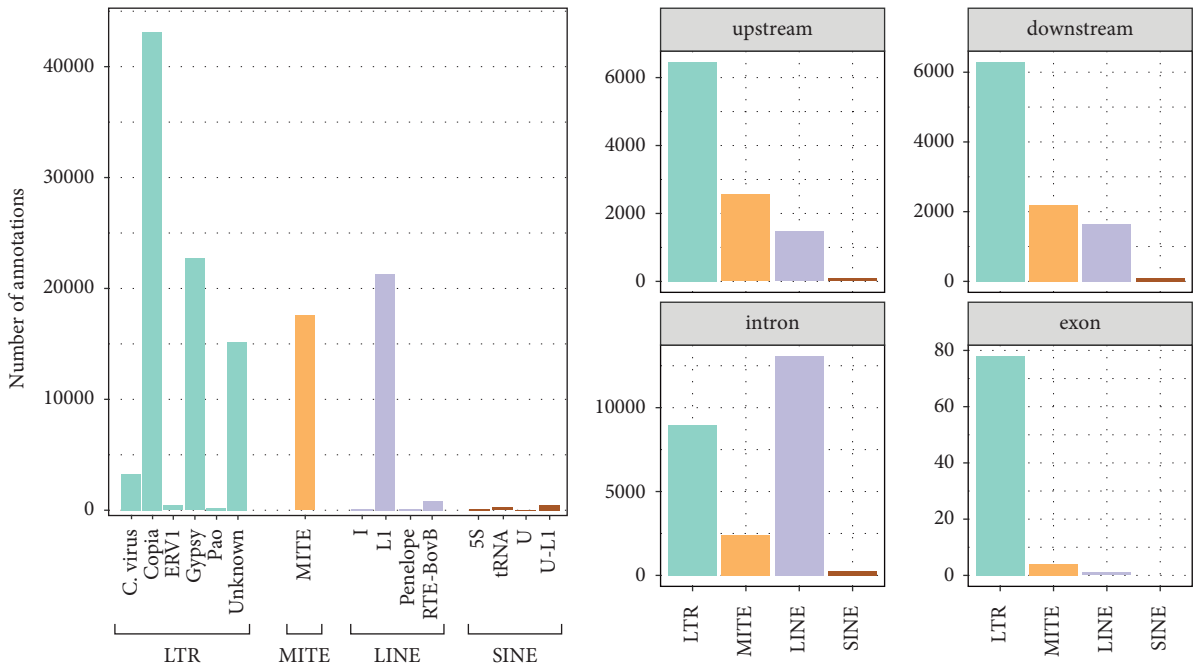


Figure 3: Continued.

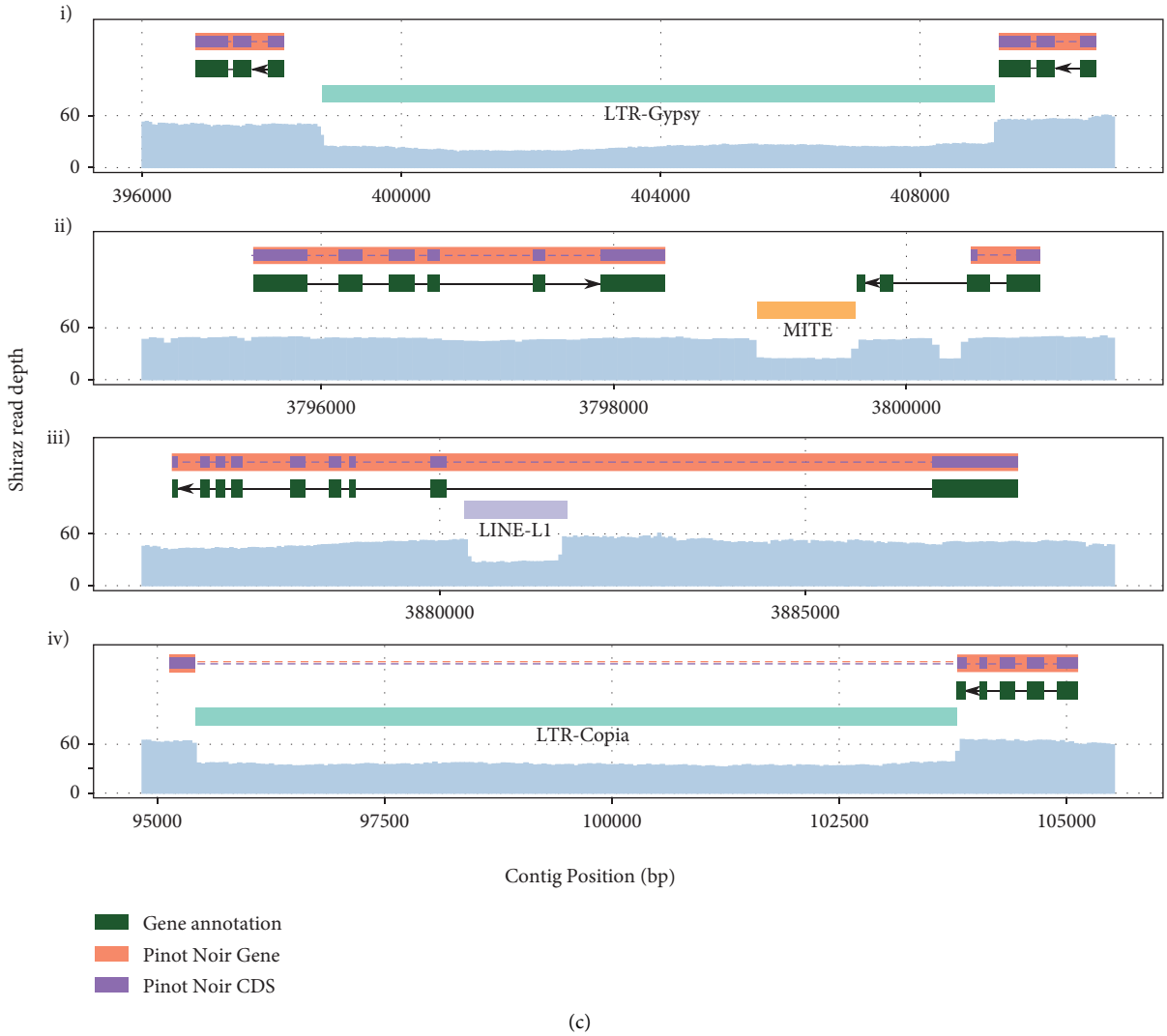
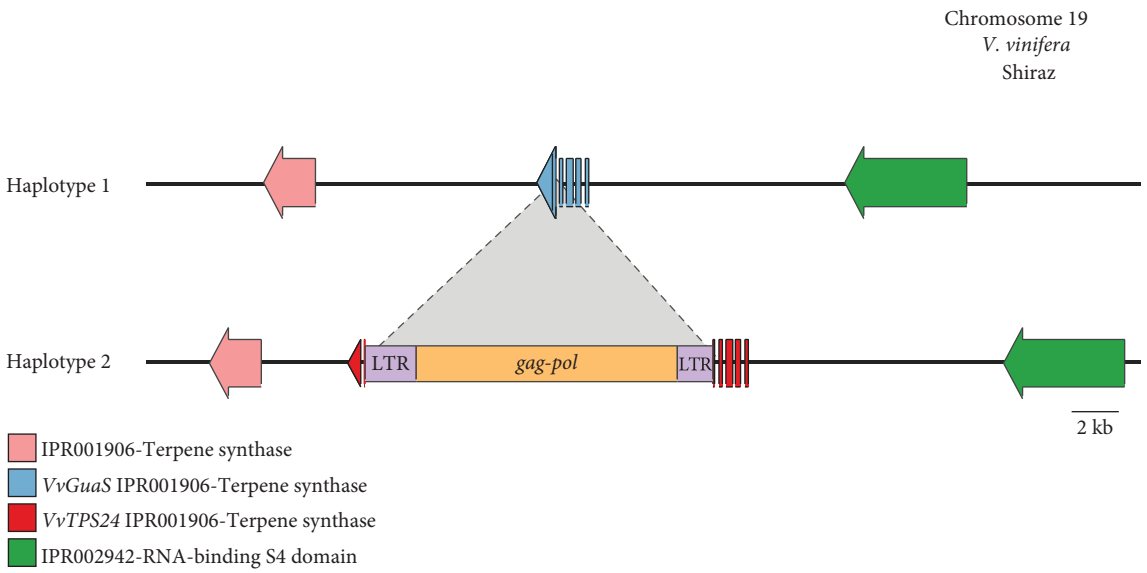


FIGURE 3: Repeat structure of *V. vinifera* Shiraz clone 1654. (a) Genome wide classification of transposable elements (TEs) annotated in the Shiraz primary assembly. LTR, MITE, LINE, RC, and SINE TEs are broken down into subclasses and colored according to their class. (b) Classification of TEs that are annotated 1 kb upstream, downstream, and intronic of an annotated gene or exonic of a Pinot Noir gene/CDS combo. Gaps in Pinot Noir gene/CDS alignments to the Shiraz reference are depicted as dashed lines. (c) TEs that are specific to Shiraz after overlapping genic TEs with structural variants from Nebbiolo, Chardonnay, Carménère, Zinfandel, Cabernet Sauvignon, Sauvignon Blanc, Riesling, and merlot long-read data: (i) upstream, (ii) downstream, (iii) intronic, and (iv) exonic.

rotundone is characterized by a peppery aroma and is responsible for one of the key varietal characteristics of Shiraz [2]. The biosynthesis of rotundone involves a two-step process, involving the production of the precursor α-guaiene from farnesyl pyrophosphate by an allele of the sesquiterpene synthase *VviTPS24* [5] and the subsequent oxidation of α-guaiene into rotundone [6, 7]. Wildtype *VviTPS24* produces an array of sesquiterpenes, of which only a minor fraction is α-guaiene [66]. However, a high α-guaiene producing variant of *VviTPS24* has been recently reported in Shiraz, which contains two polymorphisms in the active site of the protein [5]. This is likely to be linked to the ability to synthesize high levels of rotundone, although it remains to be determined if this variant is present in other cultivars.

Investigation of the *VviTPS24* locus in the diploid genome assembly of Shiraz revealed a single predicted gene model (Gene ID 002051) with 98.9% protein similarity to the *VviTPS24* ortholog from Pinot Noir (NCBI accession XP_002282488) that was present in the haplotig pool. The protein predicted by this gene annotation in Shiraz contained both the T414S and V530M substitutions that have been previously associated with higher production of α-guaiene [5] (Figure S3).

Given the diploid nature of the Shiraz assembly, a second allele of *VviTPS24* would also be expected to be present in the primary contigs of the genome assembly. To determine if this second allele was present, but missing from the initial annotation, splice-aware mapping of the CDS of *VviTPS24* was performed against the primary assembly. Results



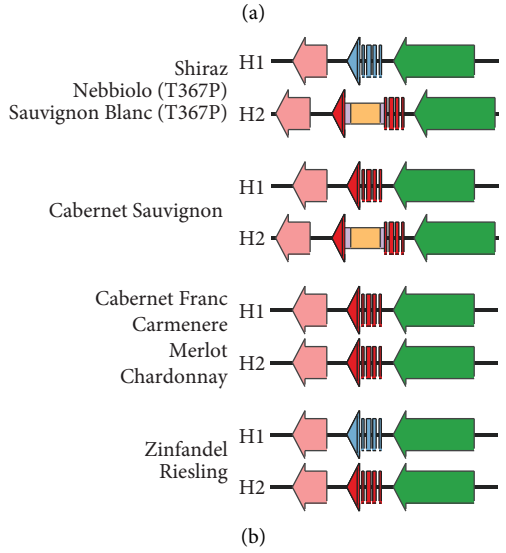


Figure 4: Continued.

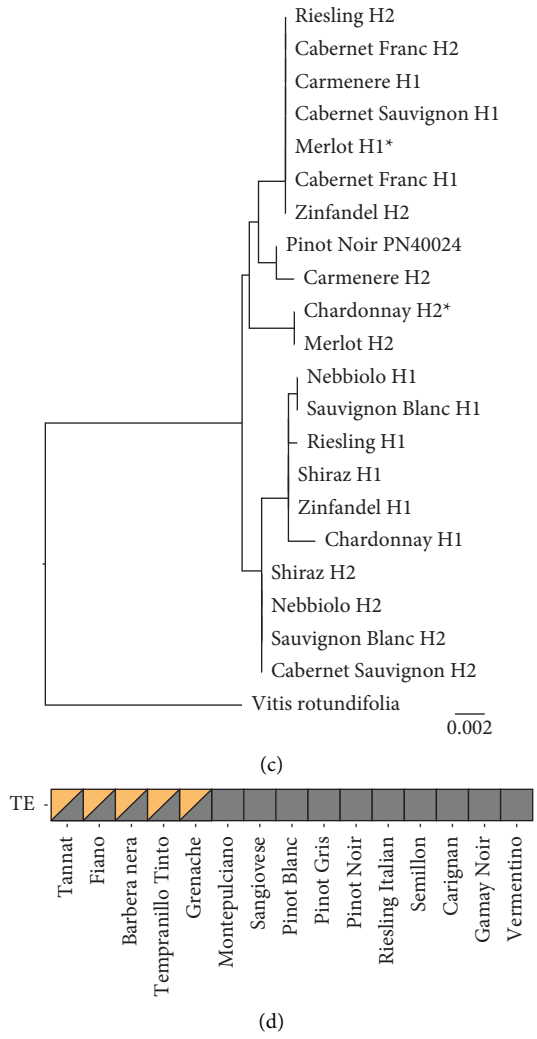


Figure 4: The *VviTPS24* locus in *V. vinifera*. (a) Functional annotation of the *VviTPS24* locus in the haplotypes of *V. vinifera* Shiraz. *VviGuaS* (blue) represents the polymorphic variant of *VviTPS24* containing the T414S and V530M amino acid substitutions previously reported [5]. *VviTPS24* (red) represents an allele without the T414S and V530M amino acid substitutions. Dotted lines depict the insertion point of a transposable element within exon 5 of *VviTPS24*. (b) Schematic representation of the *VviTPS24* locus in the haplotype phased assemblies of nine additional *V. vinifera* cultivars. (c) Maximum likelihood phylogenetic tree predicted from the CDS of *VviTPS24* for nine *V. vinifera* cultivars with *V. rotundifolia* as an outgroup. The CDS for all alleles including retroelement disrupted variants were manually predicted using the *VviTPS24* CDS of the pinot noir PN40024 reference genome. Cultivars denoted with an asterisk (*) contain an insertion/deletion within the CDS of *VviTPS24*. (d) Presence (orange) and absence (grey) of a retroelement within *VviTPS24* for 15 *V. vinifera* cultivars predicted using publicly available short-read data (Table S1).

showed the presence of a second putative allele of *VviTPS24* that contained a large, 15 kb insertion within exon 5 (Figure 4(a)). Annotation of this large insertion using the RepBase database revealed the insertion to be a Ty3-gypsy-type retrotransposon with high similarity to the Gypsy18 LTR of grapevine (NCBI accession AM476928) [67]. Manual annotation of this presumably inactivated allele of *VviTPS24* indicated that, in the absence of the LTR insertion, the protein sequence of the second allele would have 99.5% similarity to the Pinot Noir *VviTPS24* and only contain the T414S substitution (Figure S3). Overall, these results indicate that due to this unique combination of SNP and structural variation in Shiraz, the main product of the *VviTPS24* locus would be α-guaiene.

To obtain a broader understanding of the potential metabolic pathways underpinning rotundone biosynthesis in other cultivars, the *VviTPS24*, *VviFPPS* (farnesyl diphosphate synthase) [68], and *VviSTO2* (α-guaiene 2-oxidase) [7] loci, which have all been linked to the biosynthesis of rotundone, were genotyped across the eight *V. vinifera* cultivars with available haplotype phased assemblies (Figure 4(b)). No structural differences were observed for either the *VviFPPS* or *VviSTO2* genes (data not shown); however, an allele of *VviTPS24* with a LTR retrotransposon inserted within exon 5 was observed in the cultivars Nebbiolo and Cabernet Sauvignon (Figure 4(b)). Alignment of the theoretical VviTPS24 proteins indicated that, such as Shiraz, the *V. vinifera* cultivars Zinfandel, Riesling,

Nebbiolo, and Sauvignon Blanc also carry a copy of *VviTPS24* containing the T414S and V530M substitutions (Figure 4(b), Figure S3). An additional amino acid substitution (T367P) relative to the *VviTPS24* of Shiraz was also observed in the H1 haplotype of Nebbiolo and Sauvignon Blanc (Figure S3). Highly variable rotundone concentrations in grapes have been linked with environmental conditions and phenological stages of grape ripening [69–71]; however, these results also suggest a molecular basis may prime an inherent variability of rotundone concentrations between cultivars of *V. vinifera* [4, 72].

Phylogenetic reconstruction of *VviTPS24* annotations revealed that the alleles specifying the T414S and V530M amino acid substitutions share an evolutionary origin (Figure 4(c)). The topology of the gene tree was largely congruent with the species tree (Figure 2), with the lineages derived from Cabernet Franc being monophyletic in at least one allele. The H2 allele of Chardonnay and Merlot were also shown to be closely related (Figure 4(c)), providing additional support for the shared ancestry suggested by the species tree (Figure 2). After excising the TE insertion from exon 5, all alleles containing the TE were found to be monophyletic (Figure 4(c)), indicating that Shiraz inherited both the inactivated and high α-guaiene producing alleles of *VviTPS24* through an ancestral outcrossing event. Short-read data from a further 15 cultivars was mapped to the Shiraz primary, assembly identifying the TE insertion in exon 5 of *VviTPS24* in a further five cultivars (Figure 4(d)). While these results might suggest these cultivars share an ancestry with Shiraz, long-read sequencing data will be required to confirm the genotype of the putative functional copy of *VviTPS24* in these cultivars. The correlation of these results with measured rotundone levels in other cultivars might provide further insights into the relevance of the detected allelic differences in the *VviTPS24* locus once the environmental triggers of sesquiterpene biosynthesis are more thoroughly understood.

4. Conclusions

The availability of a reference genome for Shiraz expands the pool of genomes available for wine grapes while providing a foundation resource for whole-genome studies involving this iconic cultivar, including intracultivar variant identification and transcriptomic studies using a matching reference genome, rather than a disparate proxy. The identification of a pair of specific genomic variants involving the *VviTPS24* gene, outline a potential genetic basis for the propensity of Shiraz (and other cultivars) to be primed for the formation of α-guaiene-type sesquiterpenes, such as rotundone, when exposed to appropriate environmental triggers. Following appropriate confirmation, this study could provide a genetic marker for the production of cool climate-associated peppery characters in future grape breeding strategies.

Data Availability

The sequencing data and genome assembly of the cultivar Shiraz and Sauvignon Blanc are publicly available at NCBI under BioProjects PRJNA907382 and PRJNA939136. The GFF files with the gene annotations are included in the supplementary data.

Disclosure

This article has been previously published as a preprint at bioRxiv [73].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

C.A.O. and C. M.W. performed the sequencing, data analysis, and drafting of the manuscript. A.R.B supervised and provided guidance through the project and drafted the manuscript. Cristobal A. Onetto and Christopher M. Ward contributed equally to this work.

Acknowledgments

This work was supported by the Wine Australia, with levies from Australia's grapegrowers and winemakers and matching funds from the Australian Government. Support for DNA sequencing was provided by Bioplatforms Australia as part of the National Collaborative Research Infrastructure Strategy, an initiative of the Australian Government. The AWRI is a member of the Wine Innovation Cluster (WIC) in Adelaide. The authors would like to thank AWITC for the invitation to publish our research in this special issue. The authors would also like to thank Benjamin Pike for providing access to the Coombe Vineyard and Darren Graetz for collecting the samples of the cultivar Sauvignon Blanc.

Supplementary Materials

Supplementary_figures.pdf: Figures S1–S3. Tables_S1.txt: SRA accessions included in this study. Tables_S2.txt: GO molecular function enrichment analysis of hemizygous genes predicted in the assembly of Shiraz. Tables_S3.txt: Genic transposable element insertions in the Shiraz genome. Tables_S4.txt: Variable transposable element (TE) insertions between cultivars of *V. vinifera*.Tables_S5.txt: Shiraz specific transposable element (TE) insertions. SHZ_1654_primary_annotation.txt: Gene annotations in GFF3 format of the primary assembly of Shiraz. SHZ_1654_haplotigs_annotation.txt: Gene annotations in GFF3 format of the haplotigs of the genome assembly of Shiraz. (*Supplementary Materials*)

References

[1] J. E. Bowers, R. Siret, C. P. Meredith, P. This, and J.-M. Boursiquot, "A single pair of parents proposed for a group of grapevine varieties in Northeastern France," *Acta Horticulturae*, no. 528, pp. 129–132, 2000.

[2] C. A. Black, M. Parker, T. E. Siebert, D. L. Capone, and I. L. Francis, "Terpenoids and their role in wine flavour: recent

advances: terpenoids: role in wine flavour," *Australian Journal of Grape and Wine Research*, vol. 21, pp. 582–600, 2015.

[3] C. Wood, T. E. Siebert, M. Parker et al., "From Wine to pepper: rotundone, an obscure sesquiterpene, is a potent spicy aroma compound," *Journal of Agricultural and Food Chemistry*, vol. 56, no. 10, pp. 3738–3744, 2008.

[4] O. Geffroy, D. Kleiber, and A. Jacques, "May peppery wines be the spice of life? A review of research on the 'pepper' aroma and the sesquiterpenoid rotundone," *OENO One*, vol. 54, no. 2, 2020.

[5] D. P. Drew, T. B. Andersen, C. Sweetman, B. L. Møller, C. Ford, and H. T. Simonsen, "Two key polymorphisms in a newly discovered allele of the *Vitis vinifera TPS24* gene are responsible for the production of the rotundone precursor *α*-guaiene," *EXBOTJ*, vol. 67, no. 3, pp. 799–808, 2016.

[6] A.-C. Huang, S. Burrett, M. A. Sefton, and D. K. Taylor, "Production of the pepper aroma compound (−)-Rotundone, by aerial oxidation of *α*-guaiene," *Journal of Agricultural and Food Chemistry*, vol. 62, no. 44, Article ID 10809, 2014.

[7] H. Takase, K. Sasaki, H. Shinmori et al., "Cytochrome P450 CYP71BE5 in grapevine (*Vitis vinifera*) catalyzes the formation of the spicy aroma compound (−)-rotundone," *EXBOTJ*, vol. 67, no. 3, pp. 787–798, 2016.

[8] R. Velasco, A. Zharkikh, M. Troggio et al., "A high quality draft consensus sequence of the genome of a heterozygous grapevine variety," *PLoS One*, vol. 2, no. 12, p. 1326, 2007.

[9] The French–Italian Public Consortium for Grapevine Genome Characterization, J.-M. Aury, B. Noel, A. Policriti, C. Clepet, and A. Casagrande, "The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla," *Nature*, vol. 449, no. 7161, pp. 463–467, 2007.

[10] M. J. Roach, D. L. Johnson, J. Bohlmann et al., "Population sequencing reveals clonal diversity and ancestral inbreeding in the grapevine cultivar Chardonnay," *PLoS Genetics*, vol. 14, no. 11, Article ID e1007807, 2018.

[11] Y. Zhou, A. Minio, M. Massonnet et al., "The population genetics of structural variants in grapevine domestication," *Native Plants*, vol. 5, no. 9, pp. 965–979, 2019.

[12] C.-S. Chin, P. Peluso, F. J. Sedlazeck et al., "Phased diploid genome assembly with single-molecule real-time sequencing," *Nature Methods*, vol. 13, no. 12, pp. 1050–1054, 2016.

[13] A. Minio, M. Massonnet, R. Figueroa-Balderas, A. Castro, and D. Cantu, "Diploid genome assembly of the wine grape Carménère," *G3 Genes Genomes Genetics*, vol. 9, no. 5, pp. 1331–1337, 2019.

[14] A. M. Vondras, A. Minio, B. Blanco-Ulate et al., "The genomic diversification of grapevine clones," *BMC Genomics*, vol. 20, no. 1, p. 972, 2019.

[15] S. Maestri, G. Gambino, G. Lopatriello et al., "Nebbiolo" genome assembly allows surveying the occurrence and functional implications of genomic structural variations in grapevines (*Vitis vinifera* L.)," *BMC Genomics*, vol. 23, no. 1, p. 159, 2022.

[16] A. M. Vondras, L. Lerno, M. Massonnet et al., "Rootstock influences the effect of grapevine leafroll-associated viruses on berry development and metabolism via abscisic acid signalling," *Molecular Plant Pathology*, vol. 22, no. 8, pp. 984–1005, 2021.

[17] C. Zou, M. Massonnet, A. Minio et al., "Multiple independent recombinations led to hermaphroditism in grapevine," *Proceedings of the National Academy of Sciences of the U S A*, vol. 118, no. 15, Article ID e2023548118, 2021.

[18] M. Massonnet, N. Cochetel, A. Minio et al., "The genetic basis of sex determination in grapes," *Nature Communications*, vol. 11, no. 1, p. 2902, 2020.

[19] S. Koren, B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, and A. M. Phillippy, "Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation," *Genome Research*, vol. 27, no. 5, pp. 722–736, 2017.

[20] M. Kolmogorov, J. Yuan, Y. Lin, and P. A. Pevzner, "Assembly of long, error-prone reads using repeat graphs," *Nature Biotechnology*, vol. 37, no. 5, pp. 540–546, 2019.

[21] M. Chakraborty, J. G. Baldwin-Brown, A. D. Long, and J. J. Emerson, "Contiguous and accurate *de novo* assembly of metazoan genomes with modest long read coverage," *Nucleic Acids Research*, vol. 44, no. 19, p. e147, 2016.

[22] B. J. Walker, T. Abeel, T. Shea et al., "Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement," *PLoS One*, vol. 9, no. 11, Article ID e112963, 2014.

[23] M. J. Roach, S. A. Schmidt, and A. R. Borneman, "Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies," *BMC Bioinformatics*, vol. 19, no. 1, p. 460, 2018.

[24] M. Manni, M. R. Berkeley, M. Seppey, F. A. Simão, and E. M. Zdobnov, "BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes," *Molecular Biology and Evolution*, vol. 38, no. 10, pp. 4647–4654, 2021.

[25] H. Liu, S. Wu, A. Li, and J. Ruan, "SMARTdenovo: a de novo assembler using long noisy reads," *Gigabyte*, vol. 2021, pp. 1–9, 2021.

[26] M. Alonge, S. Soyk, S. Ramakrishnan et al., "RaGOO: fast and accurate reference-guided scaffolding of draft genomes," *Genome Biology*, vol. 20, no. 1, p. 224, 2019.

[27] J. M. Flynn, R. Hubley, C. Goubert et al., "RepeatModeler2 for automated genomic discovery of transposable element families," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 17, pp. 9451–9457, 2020.

[28] S. Ou and N. Jiang, "LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons," *Plant Physiology*, vol. 176, no. 2, pp. 1410–1422, 2018.

[29] J. M. Crescente, D. Zavallo, M. Helguera, and L. S. Vanzetti, "MITE Tracker: an accurate approach to identify miniature inverted-repeat transposable elements in large genomes," *BMC Bioinformatics*, vol. 19, no. 1, p. 348, 2018.

[30] A. F. A. Smit, R. Hubley, and P. Green, "RepeatMasker open-4.0. 2013-2015," 2015, https://www.repeatmasker.org/.

[31] J. Palmer and J. Stajich, *Nextgenusfs/Funannotate: Funannotate v1.5.3*, Zenodo, Geneva, Switzerland, 2019.

[32] V. Ter-Hovhannisyan, A. Lomsadze, Y. O. Chernoff, and M. Borodovsky, "Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training," *Genome Research*, vol. 18, no. 12, pp. 1979–1990, 2008.

[33] I. Korf, "Gene finding in novel genomes," *BMC Bioinformatics*, vol. 5, no. 1, p. 59, 2004.

[34] M. Stanke and B. Morgenstern, "AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints," *Nucleic Acids Research*, vol. 33, pp. W465–W467, 2005.

[35] A. Delcher, "Improved microbial gene identification with GLIMMER," *Nucleic Acids Research*, vol. 27, no. 23, pp. 4636–4641, 1999.

[36] P. J. Fabres, L. Anand, N. Sai et al., "Tissue and regional expression patterns of dicistronic tRNA–mRNA transcripts in grapevine (*Vitis vinifera*) and their evolutionary co-appearance with vasculature in land plants," *Horticultural Research*, vol. 8, no. 1, p. 137, 2021.

[37] D. C. Koboldt, Q. Zhang, D. E. Larson et al., "VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing," *Genome Research*, vol. 22, no. 3, pp. 568–576, 2012.

[38] A. R. Quinlan and I. M. Hall, "BEDTools: a flexible suite of utilities for comparing genomic features," *Bioinformatics*, vol. 26, no. 6, pp. 841-842, 2010.

[39] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Research*, vol. 32, no. 5, pp. 1792–1797, 2004.

[40] R. Bouckaert, J. Heled, D. Kühnert et al., "Beast 2: a software platform for Bayesian evolutionary analysis," *PLoS Computational Biology*, vol. 10, no. 4, Article ID e1003537, 2014.

[41] A. J. Drummond, S. Y. W. Ho, M. J. Phillips, and A. Rambaut, "Relaxed Phylogenetics and dating with confidence," *PLoS Biology*, vol. 4, no. 5, p. 88, 2006.

[42] T. Flutre, L. Le Cunff, A. Fodor et al., "A genome-wide association and prediction study in grapevine deciphers the genetic architecture of multiple traits and identifies genes under many new QTLs," *G3 Bethesda Md*, vol. 12, no. 7, Article ID jkac103, 2022.

[43] H. Li, "Minimap2: pairwise alignment for nucleotide sequences," *Bioinformatics*, vol. 34, no. 18, pp. 3094–3100, 2018.

[44] H. Li, B. Handsaker, A. Wysoker et al., "The sequence alignment/map format and SAMtools," *Bioinformatics*, vol. 25, no. 16, pp. 2078-2079, 2009.

[45] C. M. Ward, A. J. Ludington, J. Breen, and S. W. Baxter, *Genomic Evolutionary Analysis in R with geaR*, Cold Spring Harbor Laboratory, Laurel Hollow, New York, 2020.

[46] X. Zheng, D. Levine, J. Shen, S. M. Gogarten, C. Laurie, and B. S. Weir, "A high-performance computing toolset for relatedness and principal component analysis of SNP data," *Bioinformatics*, vol. 28, no. 24, pp. 3326–3328, 2012.

[47] S. Lee, D. Cook, and M. Lawrence, "Plyranges: a grammar of genomic data transformation," *Genome Biology*, vol. 20, no. 1, p. 4, 2019.

[48] F. J. Sedlazeck, P. Rescheneder, M. Smolka et al., "Accurate detection of complex structural variations using single-molecule sequencing," *Nature Methods*, vol. 15, no. 6, pp. 461–468, 2018.

[49] C. M. Ward, T.-H. To, and S. M. Pederson, "ngsReports: a Bioconductor package for managing FastQC reports and other NGS related log files," *Bioinformatics*, vol. 36, no. 8, pp. 2587-2588, 2020.

[50] H. Li, "Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM," 2013, https://arxiv.org/abs/1303.3997.

[51] G. Marçais and C. Kingsford, "A fast, lock-free approach for efficient parallel counting of occurrences of k-mers," *Bioinformatics*, vol. 27, no. 6, pp. 764–770, 2011.

[52] G. Busam, H. Kassemeyer, and U. Matern, "Differential of in *Vitis vinifera* L. To or," *Plant Physiology*, vol. 115, no. 3, pp. 1029–1038, 1997.

[53] J. F. Vouillamoz and M. S. Grando, "Genealogy of wine grape cultivars: 'Pinot' is related to 'Syrah,'" *Heredity*, vol. 97, no. 2, pp. 102–110, 2006.

[54] J. González and D. A. Petrov, "The adaptive role of transposable elements in the *Drosophila* genome," *Gene*, vol. 448, no. 2, pp. 124–133, 2009.

[55] X.-M. Niu, Y.-C. Xu, Z.-W. Li et al., "Transposable elements drive rapid phenotypic variation in *Capsella rubella*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 14, pp. 6908–6913, 2019.

[56] D. McKey, M. Elias, B. Pujol, and A. Duputié, "The evolutionary ecology of clonally propagated domesticated plants," *New Phytologist*, vol. 186, no. 2, pp. 318–332, 2010.

[57] L. Wang, Y. Huang, Z. Liu et al., "Somatic variations led to the selection of acidic and acidless orange cultivars," *Native Plants*, vol. 7, pp. 954–965, 2021.

[58] Q. H. Le, S. Wright, Z. Yu, and T. Bureau, "Transposon diversity in *Arabidopsis thaliana*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 13, pp. 7376–7381, 2000.

[59] E. E. Grundy, N. Diab, and K. B. Chiappinelli, "Transposable element regulation and expression in cancer," *FEBS Journal*, vol. 289, no. 5, pp. 1160–1179, 2022.

[60] C. D. Todd, Ö. Deniz, D. Taylor, and M. R. Branco, "Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells," *Elife*, vol. 8, Article ID e44344, 2019.

[61] S. Kobayashi, N. Goto-Yamamoto, and H. Hirochika, "Retrotransposon-induced mutations in grape skin color," *Science*, vol. 304, no. 5673, p. 982, 2004.

[62] Y. Yang, J. Ke, X. Han, W. A. Wuddineh, G. Song, and G.-Y. Zhong, "Removal of a 10-kb *Gret1* transposon from *VvMybA1* of *Vitis vinifera* cv," *Horticulture Research*, vol. 9, Article ID uhac201, 2022.

[63] D. Lijavetzky, L. Ruiz-García, J. A. Cabezas et al., "Molecular genetics of berry colour variation in table grape," *Molecular Genetics and Genomics*, vol. 276, no. 5, pp. 427–435, 2006.

[64] Q. Zhang, J. Arbuckle, and S. R. Wessler, "Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family *Heartbreaker* into genic regions of maize," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 3, pp. 1160–1165, 2000.

[65] J. Lin, M. Massonnet, and D. Cantu, "The genetic basis of grape and wine aroma," *Horticultural Research*, vol. 6, no. 1, p. 81, 2019.

[66] D. M. Martin, S. Aubourg, M. B. Schouwey et al., "Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays," *BMC Plant Biology*, vol. 10, no. 1, p. 226, 2010.

[67] O. Kohany and J. Jurka, "LTR retrotransposons from grapevine," *RepBase Reports*, vol. 7, p. 684, 2007.

[68] H. Takase, K. Sasaki, G. Ikoma, H. Kobayashi, H. Matsuo, and S. Suzuki, "Farnesyl diphosphate synthase may determine the accumulation level of (–)-rotundone in "Syrah" grapes," *VITIS - Journal of Grapevine Research*, vol. 34, pp. 99–106, 2016.

[69] L. Caputi, S. Carlin, I. Ghiglieno et al., "Relationship of changes in rotundone content during grape ripening and winemaking to manipulation of the 'peppery' character of wine," *Journal of Agricultural and Food Chemistry*, vol. 59, no. 10, pp. 5565–5571, 2011.

[70] R. G. V. Bramley, T. E. Siebert, M. J. Herderich, and M. P. Krstic, "Patterns of within-vineyard spatial variation in the 'pepper' compound rotundone are temporally stable from year to year: spatial variation of rotundone is temporally stable," *Australian Journal of Grape and Wine Research*, vol. 23, no. 1, pp. 42–47, 2017.

[71] V. V. S. R. Gupta, R. G. V. Bramley, P. Greenfield, J. Yu, and M. J. Herderich, "Vineyard soil microbiome composition related to rotundone concentration in Australian cool climate 'peppery' Shiraz grapes," *Frontiers in Microbiology*, vol. 10, p. 1607, 2019.

[72] M. J. Herderich, T. E. Siebert, M. Parker, D. L. Capone, D. W. Jeffery, and P. Osidacz, "Spice up Your life: analysis of key aroma compounds in Shiraz," in *ACS Symposium Series*, M. C. Qian and T. H. Shellhammer, Eds., American Chemical Society, Washington, DC, USA, 2012.

[73] C. A. Onetto, C. M. Ward, and A. R. Borneman, "The phased diploid genome assembly of Vitis vinifera cv," Cold Spring Harbor Laboratory, Laurel Hollow, New York, 2022.