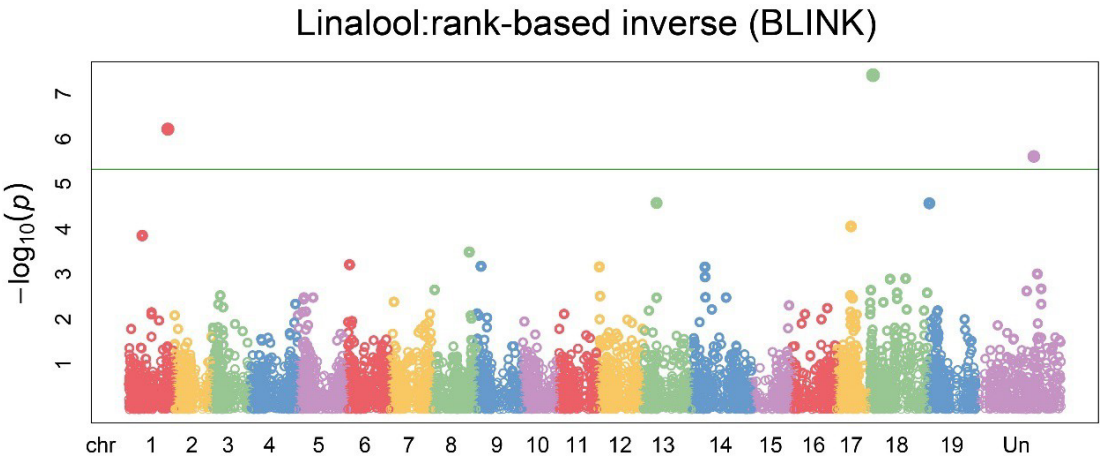
Australian Government

Wine Australia

CSIRO

Scion genetics and improvement: desirable consumer traits



FINAL REPORT TO WINE AUSTRALIA

Project Number: **CSA 1701-1.2**

Principal Investigator: **Paul Boss**

Research Organisation: **CSIRO**

Date: **3 February 2023**

Australia's National
Science Agency

CSA 1701-1.2 Scion genetics and improvement: desirable consumer traits

Paul Boss, Emily Nicholson, Ali Hassan and Christopher Davies

CSA 1701-1.2

February 2023

Wine Australia Final Report

Australian Government
Wine Australia

Citation

Boss PK, Hassan A, Nicholson EL and Davies C (2022) Scion genetics and improvement: desirable consumer traits. Final report to Wine Australia, Project CSA 1701-1.2. CSIRO, Australia.

Copyright

© Commonwealth Scientific and Industrial Research Organisation 2022. To the extent permitted by law, all rights are reserved and no part of this publication covered by copyright may be reproduced or copied in any form or by any means except with the written permission of CSIRO.

Important disclaimer

CSIRO advises that the information contained in this publication comprises general statements based on scientific research. The reader is advised and needs to be aware that such information may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific and technical advice. To the extent permitted by law, CSIRO (including its employees and consultants) excludes all liability to any person for any consequences, including but not limited to all losses, damages, costs, expenses and any other compensation, arising directly or indirectly from using this publication (in part or in whole) and any information or material contained in it.

Acknowledgements

We would like to thank Roger Maywald, SARDI Farm Manager at the Nuriootpa Research Centre, for providing access to the grape variety germplasm collection. We also thank Christine Böttcher, Crista Burbidge, Katie Harvey and Gerhard Rossouw for sampling fruit from the 1st generation mildew resistant varieties and producing the small-scale wines from these plants. Sue Maffei is also acknowledged for helping maintain the suite of instruments used for the analysis of grape and wine composition. This research was supported financially by CSIRO and the grape growers and winemakers of Australia through their investment body, Wine Australia, with matching funding from the Australian Government. CSIRO is a member of the Wine Innovation Cluster (www.wineinnovationcluster.com) Adelaide, SA.

Contents

Contents iv

1 Abstract 1

2 Executive summary 2

3 Background 4

4 Project Aims 7

5 Method 13

5.1 Overview 13

5.2 Grape Material and Sampling 13

5.2.1 Biparental crosses 13

5.2.2 Germplasm collections 14

5.2.3 Grape berry developmental series samples 17

5.3 Genotyping and Phenotyping 18

5.3.1 Biparental crosses 18

5.3.2 DArTseq for GWAS 18

5.3.3 SSR and SNP marker analysis 19

5.3.4 Micro-fermentation 19

5.3.5 Wine volatile compound analysis 19

5.3.6 Grape volatile compound analysis 24

5.3.7 Statistical analysis 25

5.4 Gene mapping and marker development 25

5.4.1 QTL analysis 25

5.4.2 Genome-wide association studies 26

5.4.3 Marker development 26

5.5 Candidate gene functional characterisation 27

5.5.1 Gene expression studies 27

5.5.2 Recombinant protein production and assays 27

6 Results/Discussion 31

6.1 Mapping traits using biparental crosses 31

6.1.1 Introduction 31

6.1.2 Genetic maps generated from established biparental crosses 31

6.1.3 QTL mapping of wine composition traits in the biparental crosses 39

6.2 Mapping traits using genome-wide association studies 68

6.2.1 Introduction 68

6.2.2 Data analysis 68

6.3 Candidate genes underlying QTLs of interest 99

6.3.1 Introduction 99

6.3.2 Candidate genes underlying the wine (*Z*)-3-hexenol concentration QTL . 99

6.3.3 Candidate genes linked to *VvGAI1* potentially involve in monoterpene biosynthesis 109

6.3.4 Candidate genes in the QTL associated with wine acetate ester concentrations 115

6.3.5 Candidate genes in the QTL associated with wine β-damascenone concentrations 123

6.4 Generation and/or examination of other populations for gene mapping 126

6.4.1 F2 population derived from self-pollinated Mv × R F1 plant #131 126

6.4.2 1st generation mildew-resistant vines 135

6.5 Development of markers to genes of known function 139

6.5.1 Introduction 139

6.5.2 Marker for acylated anthocyanins 140

6.5.3 Marker for methoxypyrazine production in berries 141

6.5.4 Marker for potential rotundone biosynthesis 141

6.5.5 Marker for red berry flesh 144

6.6 GCMS analysis of 1st generation mildew-resistant red wines 145

6.6.1 Introduction 145

6.6.2 Comparison of volatile compound profiles 145

6.6.3 Volatile compounds associated with the groups from the cluster analysis 147

6.6.4 Summary 151

7	**Outcome/Conclusion**	**152**
8	**Recommendations**	**155**
9	**Appendix 1: Communication**	**157**
10	**Appendix 2: Intellectual Property**	**159**
11	**Appendix 3: References**	**160**
12	**Appendix 4: Staff**	**167**

1 Abstract

The use of genetic markers for important consumer-preferred sensory traits will streamline the breeding process making it more rapid and less expensive. Initial studies have identified quantitative trait loci that explain differences in the volatile composition of wines from grapevine mapping populations that are related to 'fruity', 'green' and 'floral' sensory characters. Genome-wide association studies using a variety collection have identified quantitative trait loci and markers that predict the abundance of wine volatiles. Other markers have been designed to genes already shown to be important for wine phenotypes (methoxypyrazines, rotundone, anthocyanin stability, red flesh) to aid the selection of new scions for specific wine styles.

2 Executive summary

The winegrape varieties most extensively used throughout Australia are hundreds of years old and lack resistance to the common fungal pathogens powdery and downy mildew. Breeding programs around the world are using resistance genes from wild *Vitis* species introgressed into *vinifera* to produce new fungal disease resistant varieties, and this is true for Australia in the form of Wine Australia Project CSA 1701-1.1. The uptake of these new varieties will be improved if the advances in agronomic performance can be matched by these cultivars producing wines with high consumer acceptance. This project was designed to streamline the selection of such varieties by developing genetic markers for consumer-preferred sensory traits so that these plants can be identified without the need for small scale winemaking and sensory analysis. Added benefits from the development of such markers include a greater understanding of how grape composition influences wine style and the development of tools to predict how environment and management will affect grape and wine composition.

The project methodology consisted of four main parts. 1) Use of existing biparental crosses between the microvine and either Riesling or Cabernet Sauvignon to generate micro-scale wines and genetic linkage maps from the progeny. Volatile compound profiling from the wines would then allow the mapping of quantitative trait loci (QTLs) associated with the concentration of the wine compounds. Candidate genes in the genome regions underlying the QTLs would be identified, characterised and markers designed if found to be causing the phenotypic changes. 2) Sample leaves and berries from the varietal collection held by SARDI at the Nuriootpa Research Centre and use these to generate genome single nucleotide polymorphism (SNP) data using DArTseq and volatile compound profiles from micro-scale wines. Genome-wide association studies (GWAS) would then be used to identify SNPs associated with the concentration of wine volatile compounds and candidate genes characterised and/or markers then developed if appropriate. 3) Generate new germplasm or identify existing populations that are useful for mapping consumer-preferred traits. Sample leaves and berries and produce matching SNP and wine volatile compound datasets for mapping using linkage maps or GWAS. 4) Develop easy-to-use markers for laboratory detection of SNPs found to be responsible for consumer-preferred traits according to research published in peer-reviewed journals by other research teams.

The biparental crosses were successfully used to map genome regions associated with the following wine compounds: (*Z*)-3-hexenol, associated with 'leafy green' sensory characters; acetate esters, which cause 'fruity' and 'floral' type aromas; linalool, which contributes to 'citrus' and 'floral' sensory attributes and; β-damascenone, an enhancer of 'fruity' flavours and aromas. In the case of (*Z*)-3-hexenol, causal genes were identified in the genome region that catalyse the isomerisation of (*Z*)-3-hexenol to (*E*)-3-hexenal and markers were generated to predict this phenotype in breeding programs. For the acetate esters, SNPs in several candidate genes in the single QTL region identified, were tested against the data from the SARDI variety collection, but none were significantly associated with the acetate ester concentrations. Nevertheless, the mapping of this phenotype to the grape genome continues to provide evidence that grape composition can change acetate ester production, by yeast, during fermentation. Moderate linalool production, which imparts 'citrus' and 'floral' characters in varieties like Riesling and Viognier, appeared to be inhibited by a gene linked to

the dwarf locus in the microvine. Potential causal genes were characterised, and these NUDX hydrolases were shown to be able to dephosphorylate the precursors of monoterpenes and sesquiterpenes. However, genetic differences in these genes between the parents of the progeny in the biparental cross were not found and so the genotype:phenotype link was not proven. Finally, β-damascenone concentration in wines made from the microvine × Cabernet Sauvignon progeny mapped to a region on chromosome 2 that co-localised with the flower sex locus. It was found that the plants with female flowers had higher concentrations of β-damascenone in the wines, probably driven by a higher skin:flesh ratio in the very small berries produced from such flowers.

Valuable datasets were developed during the preparation for the GWAS, which can be used in future projects. Small-scale wines were produced from 177 grape varieties and these were analysed using gas chromatography-mass spectrometry to obtain volatile compound profiles. To compliment the phenotypic analysis, DNA was prepared from each variety and thousands of markers generated using the DArTseq platform of an external provider. The GWAS using these datasets identified QTLs for 17 wine volatile compounds and, in many cases, candidate genes were identified in the vicinity of the significant markers. Importantly, the GWAS confirmed the location of the QTL for (*Z*)-3-hexenol concentration in wine previously identified in the biparental crosses. The results of this work will inform future studies into the genetics of consumer-preferred traits that could feed into new germplasm improvement initiatives.

One new population was generated for genetic studies comprising F2 progeny derived from a self-pollinated F1 plant originally generated from the microvine × Riesling cross which produced Riesling-like concentrations of linalool. A dense genetic map was generated from this population and the fruit from the progeny was analysed for linalool concentrations using two extraction techniques. Subsequent QTL mapping found that the production of linalool in this population was multi-genic with seven QTLs identified. However, the future utility of this cross was demonstrated through the identification of a single major QTL responsible for the production of the monoterpene α-phellandrene with credible candidate genes located in the associated genome region. An existing population of 1st generation mildew resistant vines, located at the SARDI Nuriootpa Research Centre, was also used for GWAS to identify genome regions associated with wine volatile compounds. While some significant relationships were observed, a deeper examination of the data is required to determine if another mapping strategy would be more successful.

Markers for five traits were developed from previously published research and the work conducted in this project. The markers were designed as cleaved amplified polymorphic sequences (CAPS) markers for ease-of-use in molecular biology laboratories. The traits targeted were: acylated (stable) anthocyanins; methoxypyrazine production in berries; α-guaiene production (rotundone precursor); red-fleshed berries; (Z)-3-hexenol concentration in wine.

In addition to the identification of QTLs for many wine volatile compounds and the development of markers for mapping consumer-preferred traits, this project has generated materials, data and analytical pipelines that can be used to deliver superior germplasm to the Australian wine sector by identifying targets for DNA-free gene editing technology or conventional breeding.

3 Background

Three-fifths of the 2021 crush in Australia was made up of three varieties, Shiraz, Chardonnay and Cabernet Sauvignon, which are first mentioned in historical records in 1781, 1583 and 1777 respectively (Robinson, et al. 2013). This is a clear demonstration that the Australian wine sector relies on winegrape varieties that were selected many hundreds of years ago. Analysis of the genomes of hundreds of domesticated varieties of *Vitis vinifera* showed that there has been a 'weak domestication bottleneck' and since the selection of the current commonly used varieties centuries ago there has been widespread distribution via vegetative propagation (Myles, et al. 2011). As a result, the currently used varieties (and those present in germplasm collections) have high amounts of genetic diversity (Myles, et al. 2011). This serves as an advantage for grapevine breeders as it allows for the potential to produce unique characters and make large (positive) changes in phenotypes with very few, if not single, crosses between existing varieties. This is the envy of many breeders of other crops, like cereals, which are inbred and homozygous and where genetic gain is often minimal. It is also an advantage for geneticists as it represents a rich source of variation that can be used to produce populations that segregate for interesting phenotypes, so the genetic framework underpinning an important characteristic of a plant can be determined. This knowledge can then be converted into markers for use in breeding projects and targets to measure to improve our understanding of the phenotype and how we might manipulate it in existing vineyards.

The ancient origin of the current commercial varieties raises the question as to why these chance seedlings, selected hundreds of years ago and long before the publication of Mendel's discovery of 'genetics' (Mendel 1866) and Watson and Crick's structure of DNA (Watson and Crick 1953), are still preferred to those that have been bred in more recent breeding programs. The French- American hybrids had obvious issues with wine quality, as they were F1 hybrids between classic *vinifera* winegrape varieties and wild North American *Vitis* species. So, while they provided a solution to disease problems, the quality issues (due to genes from the wild species parents) were never going to allow them to supplant the *vinifera* varieties, especially when the development of rootstocks alleviated issues with phylloxera.

The solution to the 'quality' problem is to introgress the desired genes from the wild species into vinifera through a process of selecting the progeny with the targeted phenotype from an initial cross and then backcrossing to *vinifera*, still selecting the desired phenotype, but removing more of the wild species DNA with each cross. In the past, this solution was a lengthy process in grapevine, and it was the foresight of Dr Alain Bouquet from INRA in France that has allowed the production of many of the recently released mildew-resistant varieties around the world. Dr Bouquet embarked on a backcrossing strategy in the mid 1970s to introgress mildew resistance genes from *Muscadinia rotundifolia* into *vinifera* and 23 years and five backcrosses later, he had plants that were mildew-resistant, but ~98% of the genome was derived from *vinifera* (Pauquet, et al. 2001). One of these plants was the female parent for the 1st generation mildew-resistant varieties produced by CSIRO as part of Wine Australia project CSP 0904. However, durable resistance requires the use of multiple resistance genes, ideally from different wild species, so that the pathogen has less chance of breaking the immunity. Fortunately, our ability to introgress genes from wild grapevine species has been significantly enhanced with the discovery of the microvine (Boss and Thomas 2002), which

greatly shortens the generation time due to its rapid flowering phenotype and the fact that it can be grown in controlled environments year-round. This has formed the basis of the work in Project CSA 1701-1.1 'Scion genetics and improvement: development of new disease-resistant varieties', where grapevines with multiple resistance genes to powdery and downy mildew have been generated from microvine breeding lines with stacked resistance loci introgressed from wild species.

The adoption of these new scion genotypes will be enhanced by the presence of additional traits that influence consumer acceptance of the wines made from these varieties. This works in parallel with the development of scions that possess improved agronomic traits (e.g. disease resistance) that are important for sustainable wine production. The use of genetic markers for important consumer-preferred traits would streamline the selection process making it more rapid and less expensive. However, to select for such traits during the breeding process, we require markers that indicate to what extent the fruit from a scion will be able to produce certain sensorially-active compounds. Ultimately, it would be possible to remove genotypes that produce compounds with negative sensory characteristics and to select for those with enhanced production of compounds with positive attributes before the expensive processes of field trials and sensory evaluation are undertaken.

The current issue with such an approach is that there is a gap in our knowledge about the genetics of wine consumer-preferred traits in grapevine. While there are some examples of genes being identified that control important consumer-preferred traits, for example, colour (Kobayashi, et al. 2004) and 'Muscat' flavour (Battilana, et al. 2011), there are many other wine sensory attributes for which the genetic basis is unclear or unknown. The CSIRO research group in Adelaide has demonstrated that the microvine is a valuable tool in dissecting the genetics of wine consumer-preferred traits by using the system to identify the gene responsible for the production of methoxypyrazines in grape berries (Dunlevy, et al. 2013). The group has also demonstrated the use of transgenic plants for gene discovery, with the identification of the gene responsible for the production of stable, acylated anthocyanins (Rinaldo, et al. 2015). The main value of these techniques is that they greatly increase the speed at which these studies can be conducted.

There are also new genetic techniques that can be used to link genes with traits. Genome wide association studies (GWAS) combine the power of next generation sequencing techniques with advanced statistical algorithms to map difference in genomes to differences in traits (Marigorta, et al. 2018). This methodology was first developed for mapping human diseases where access to other genetic tools such as biparental crosses is not possible, but it has been successfully adapted in plants (Korte and Farlow 2013) and represents a good technique to compliment other genetic approaches. Access to extensive grapevine variety collections gave us the opportunity to explore this approach to generate lists of markers associated with traits of interest that may be used for gene discovery or selection of progeny with desired traits.

The study of consumer-preferred traits is hampered by the complexity of wine chemistry and how this relates to the sensory properties of wine. The determination of a compound or group of compounds that contribute to a certain flavour or aroma character is a goal for scientists trying to develop a better understanding of how wine style is constructed chemically. Some compounds that contribute to varietal sensory characteristics have been identified, such as the methoxypyrazines, which contribute vegetal or earthy aromas to certain varieties (e.g. Cabernet Sauvignon and Sauvignon Blanc) and varietal thiols, which are important contributors to Sauvignon Blanc flavour and aroma

(for review see Robinson, et al. 2014). However, such compounds are not solely responsible for the sensory attributes of the respective varietal wines; changes in the total volatile compound profiles will alter the sensory perception of the wines and ultimately determine consumer preference. It has been shown that the volatile compound profiles of wine depend greatly on the composition of the grapes used to produce that wine, even those compounds that are synthesised by the yeast during fermentation (e.g. Blackford, et al. 2022, Boss, et al. 2015, Dennis, et al. 2012). The profound influence that grape composition has on wine style provides the potential to manage wine flavour and aroma more effectively in the vineyard. Therefore, the identification of grape-derived precursors contributing to wine flavour and aroma is a key research outcome that is essential for growing grapes fit for purpose. Gene mapping is a powerful tool to discover how fruit composition influences wine composition, especially if the phenotyping uses wine made from individual progeny and not the fruit. The identification of genes controlling the concentrations of wine volatile compounds would give us insights into the metabolic pathways and precursors that modify wine style. So, additional benefits accrue from determining the genetics of consumer-preferred traits, as knowledge of the genes that regulate traits will allow us to understand the biochemistry behind the production of important molecules in grapes and wine. This, in turn, facilitates the design of strategies to manipulate the production of these compounds, in the vineyard and winery, to control wine style in existing cultivars.

4 Project Aims

The main objective of this project was to develop tools to enable a more rapid screening of scion selections for desirable, consumer-preferred traits. Preliminary work had shown that the concentrations of positive wine volatile compounds, such as acetate esters and linalool, and those with negative attributes, such as the C_6 alcohols, segregate in our existing grapevine populations resulting from biparental crosses with the microvine. Understanding how positive and negative sensory attributes are controlled genetically in grapes would enable the development of markers to select for specific traits or to eliminate those with potentially negative attributes at the seedling stage. This would save both time and vineyard space and improve the value of wines produced from new disease resistant cultivars and offer alternative wines styles. Another aim was to use the genetic knowledge to improve our understanding of the biochemistry behind the expression of these traits in grapes and wine. This would ultimately be linked to viticulture interventions so that wine style can be effectively tailored in the vineyard. Also, wines from current disease resistant germplasm would be analysed to understand how their volatile profiles relate to those of currently used cultivars.

Year 3 outputs:

- Mapping of at least two consumer-preferred traits related to important wine sensory properties in existing populations to develop markers for rapid breeding.
- Identification of suitable germplasm for mapping of other important consumer-preferred traits such as sugar accumulation or wine volatile production.

Year 5 outputs:

- Establishment of suitable populations for the mapping of sugar accumulation and wine volatile phenotypes.
- Development of markers for at least four consumer-preferred traits using existing populations and data from the literature.
- Benchmarking of the volatile composition of wines made from elite disease resistant varieties against commonly grown cultivars.

Year 10 aspirational outcomes:

- A demonstrated pipeline for the rapid development of new scions with robust disease resistance and superior consumer-preferred traits.
- Greater understanding of the biochemistry underpinning important consumer-preferred traits that informs vineyard management in the production of specific wine styles.

The planned outputs as described in the original funding application are listed below.

Year 1:

OUTPUT	TARGET	ACTIVITIES
Project staff appointed	1/10/17	PDF position on identifying and mapping desirable consumer traits advertised and appointed.
Data set listing differences in concentration of wine volatile compounds amongst the progeny of one of microvine x Riesling F1 or microvine x Cabernet Sauvignon F1 crosses (~90 individuals in each).	30/6/18	Veraison noted on individual bunches on flowering dwarf vines. Bunches harvested at 7 weeks post veraison and berries ground into powder. Brix noted for each sample and fermentations conducted after normalising for sugar content. Fermentations using ~25g lots followed by weight lost, and wines clarified by centrifugation before storage at 4 °C. Wines produced from progeny analysed by head space solid phase microextraction gas chromatography (HS-SPME-GC) using multiple internal standards. Wine volatile compounds quantified using extracted ions and normalised using appropriate standards. Data analysed for normality and repeatability in two replicates.
Genetic maps of both F1 crosses generated using at least 200 markers.	30/6/18	DNA purified from parents and progeny and analysed using suitable markers spread across the genome. Data collated and used to draw genetic maps for the microvine x Riesling and microvine x Cabernet Sauvignon populations.
Markers designed to consumer traits from existing knowledge.	30/6/18	Literature searched for important genes shown to be potentially involved in consumer-driven traits (e.g., acylated anthocyanins, rotundone). Markers designed and tested in appropriate material.
New mapping populations.	30/6/18	Exploration of existing germplasm to identify material for future consumer trait mapping. Carry out crosses if and when appropriate.

Year 2:

OUTPUT	TARGET	ACTIVITIES
Data set listing differences in concentration of wine volatile compounds amongst the progeny of a microvine x Cabernet Sauvignon F1 cross (~90 individuals in each).	30/6/19	Veraison noted on individual bunches on flowering dwarf vines. Bunches harvested at 7 weeks post veraison and berries ground into powder. Brix noted for each sample and fermentations conducted after normalising for sugar content. Fermentations using ~25g lots followed by weight lost, and wines clarified by centrifugation before storage at 4 °C. Wines produced from progeny analysed by head space solid phase microextraction gas chromatography (HS-SPME-GC) using multiple internal standards. Wine volatile compounds quantified using extracted ions and normalised using appropriate standards.

QTLs of at least 2 important wine volatiles in the microvine x Riesling progeny identified and fine mapping initiated.	30/12/18	Data from the volatile analysis of the microvine x Riesling wines analysed for normality and repeatability in two replicates. Data scaled and standardised when appropriate. Data analysed for significant QTLs linked to the production of wine volatiles using the microvine x Riesling map produced in previous year. QTLs selected for further study and new markers around the loci designed for fine mapping.
Candidate genes underlying QTL identified and characterization initiated.	30/6/19	Fine mapping of QTL loci used to develop list of genes within the region. Likely candidates selected based on proposed function and phenotypes of the progeny. Polymorphisms analysed to see if any relationship to the phenotype can be identified. Candidate genes cloned into relevant vectors for functional analysis. Functional analysis initiated in heterologous systems (*E. coli*, tobacco, yeast).
Association mapping data set to confirm QTL mapping of traits in the biparental crosses.	30/6/19	Harvest ripe fruit from a selection of grapevine cultivars. Leaf material sampled from the same vines and DNA purified. Micro-scale fermentations conducted on grape material.
Markers designed to relevant traits linked to specific genes by other research groups if information becomes available.	30/6/19	Literature continually searched for important genes shown to be potentially involved in consumer-driven traits. Markers designed and tested in appropriate material.

Year 3:

OUTPUT	TARGET	ACTIVITIES
QTLs of at least 2 important wine volatiles in the microvine × Cabernet Sauvignon progeny identified, and fine mapping initiated.	30/6/20	Data from the volatile analysis of the microvine × Cabernet Sauvignon wines analysed for normality and repeatability in two replicates. Data scaled and standardised when appropriate. Data analysed for significant QTLs linked to the production of wine volatiles using the microvine × Cabernet Sauvignon map produced in year 1. Traits being analysed from the microvine × Riesling population tested in the microvine × Cabernet Sauvignon cross to confirm and improve QTLs and modify gene lists underlying the QTLs when appropriate. Novel QTLs selected for further study and new markers around the loci designed for fine mapping.

Candidate genes underlying QTLs identified and characterization initiated.	30/6/20	Fine mapping of QTL loci underlying: a) unsaturated C_6 aldehyde/alcohol production; b) acetate ester concentration; and c) monoterpene synthesis conducted to refine candidate gene lists. Likely candidate genes in the QTLs selected from the annotated genome based on proposed function. Polymorphisms analysed to see if any relationship to the phenotype can be identified. Candidate genes cloned into relevant vectors for functional analysis. Functional analysis initiated in heterologous systems (e.g., *E. coli*, tobacco, yeast). Functional analysis in grapevine initiated in collaboration with subproject 1.7.
Association mapping data set to confirm QTL mapping of traits in the biparental crosses.	30/6/20	Harvest ripe fruit from a selection of grapevine cultivars not examined in the previous year of the project. Leaf material sampled from the same vines and DNA purified. Micro-scale fermentations conducted on grape material.
Generate new germplasm or identify existing germplasm with important genotypes of phenotypes.	30/6/20	Begin to generate a F2 population of ~100 plants from the microvine × Riesling #131 F1 plant that we predict will segregate for moderate monoterpene production (i.e., aromatic white wine concentrations, but not high Muscat-like monoterpene production) Use primers developed from last year to screen existing crosses and germplasm for genes important for increasing anthocyanin production. Develop a strategy to pyramid genes for high anthocyanin production into a microvine background.
Markers designed to relevant traits linked to specific genes by other research groups if information becomes available.	30/6/20	Literature continually searched for important genes shown to be potentially involved in consumer-driven traits. Markers designed and tested in appropriate material.

Year 4:

AIM (YEAR 3 / YEAR 5 DELIVERABLES)	OUTPUTS	ACTIVITIES	ACTIVITY TARGET DATE
Yr 5 output: Development of markers for at least four consumer traits using existing populations and data from the literature.	Association mapping data set to confirm QTL mapping of traits in the biparental crosses and the identification of new QTL regions for important wine volatile compounds by genome-wide association studies.	Analyse the chromatograms from the GCMS analysis of the micro-scale wines made from the various varieties sampled from Nuriootpa. Develop markers from candidate genes identified from biparental crosses and test for significant association with relevant phenotypes across the variety collection. Conduct relevant association mapping analyses with the SNP data to see if QTLs co-locate with those from microvine biparental crosses. Conduct GWAS studies to identify potential QTLs associated with important wine volatiles not mapped using the biparental crosses.	30/6/2021

Yr 5 output: Development of markers for at least four consumer traits using existing populations and data from the literature.	Candidate genes underlying QTLs identified, characterised and markers developed when appropriate.	Continue to characterise putative (Z)-3:(E)-2-hexenal isomerase genes underlying the QTL for (Z)-3-hexenol concentration by: a) analysing polymorphisms to see if they explain the phenotypes observed in crosses; b) correlating expression during berry development with C_6 compound composition; c) designing markers and testing for association with the C_6 compound phenotype using data from the variety collection; d) characterising gene function using gene-editing in collaboration with subproject 1.7. Fine mapping of QTL loci underlying acetate ester concentration to refine candidate gene lists. Likely candidate genes in this QTL selected from the annotated genome based on proposed function. Polymorphisms analysed to see if any relationship to the phenotype can be identified. Any candidate genes identified cloned into relevant vectors for functional analysis. Functional analysis initiated in heterologous systems (e.g., *E. coli*, tobacco, yeast).	30/6/2021
Yr 5 output: Establishment of suitable populations for the mapping of sugar accumulation and wine volatile phenotypes.	Mapping population tested for segregation of the moderate production of monoterpenes seen in non-Muscat aromatic varieties.	Finish the generation of a F2 population of ~100 plants from the microvine × Riesling #131 F1 plant that we predict will segregate for moderate monoterpene production. Begin phenotyping of progeny for monoterpene concentration as they begin producing fruit. Genotype progeny using SNP analysis if segregation observed in this population. Continue to characterise NUDX hydrolase genes, linked to the dwarf locus, that may reduce the concentration of geranyl diphosphate, the substrate of terpene synthase	30/6/2021
Yr 5 output: Establishment of suitable populations for the mapping of sugar accumulation and wine volatile phenotypes.	Generate new germplasm or identify existing germplasm with important genotypes of phenotypes.	Continue to screen existing crosses and germplasm for genes important for increasing anthocyanin production. Begin to implement the strategy to pyramid genes for high anthocyanin production into a microvine background.	30/6/2021
Yr 5 output: Development of markers for at least four consumer traits using existing populations and data from the literature.	Markers designed to relevant traits linked to specific genes by other research groups if information becomes available.	Literature continually searched for important genes shown to be potentially involved in consumer-driven traits. Markers designed and tested in appropriate material.	30/6/2021

Year 5:

AIM (YEAR 5 DELIVERABLES)	OUTPUTS	ACTIVITIES	ACTIVITY TARGET DATE
Yr 5 output: Development of markers for at least four consumer traits using existing	GWAS results for ~30 target volatile compounds obtained, and markers developed when appropriate.	Integrate and collate the GCMS data from targeted compounds identified in the small-scale wines produced from the red and white varieties in the SARDI collection at Nuriootpa.	30/6/2022

populations and data from the literature.		Conduct appropriate quality control on genotyping results and prepare datasets for GWAS. Identify major QTLs underlying targeted wine volatile compounds. Design markers to these QTLs for use in future breeding programs using appropriate methodologies.	
Yr 5 output: Development of markers for at least four consumer traits using existing populations and data from the literature.	Confirmation of QTLs and markers developed for traits mapped in the biparental crosses and associated with important wine volatile compounds.	Finalise characterisation of hexenal isomerase genes and design markers to assist in the selection of low (Z)-3-hexenol producing plants from breeding programs. Use new genotyping data from the biparental crosses to fine map and better define the borders of the locus for enhanced acetate ester production. Refine the list of candidate genes influencing acetate ester concentration in wine. Develop markers from candidate genes identified and test for significant association with relevant phenotypes across the variety collection. Identify potential targets for gene editing technologies to enhance consumer traits in existing varieties.	30/6/2022
Yr 5 output: Development of markers for at least four consumer traits using existing populations and data from the literature.	Markers designed to relevant traits linked to specific genes by other research groups if information becomes available.	Literature continually searched for important genes shown to be potentially involved in consumer-driven traits. Markers designed and tested in appropriate material.	30/6/2022
Yr 5 output: Establishment of suitable populations for the mapping of sugar accumulation and wine volatile phenotypes.	Identify existing germplasm with important genotypes or phenotypes.	Use the data generated from the variety collection to identify any novel wine volatile phenotypes that can be integrated into the current breeding program. Analyse grape chemistry, GCMS and genotyping data from 1st generation mildew resistant varieties to identify consumer traits that can be exploited in future breeding, gene discovery and/or gene editing projects. Identify useful plants from the initial steps of the high anthocyanin microvine breeding that would be useful to combine acylation, red flesh and diglucoside phenotypes in any future project.	30/6/2022
Yr5 output: Generation of background knowledge to inform future research into consumer traits	Communication of project results and recommendations to interested parties	Prepare final report, prepare papers for publication, disseminate final project outcomes to interested parties.	30/6/2022

5 Method

5.1 Overview

In developing this project, our hypothesis was that certain wine volatile compounds are dependent on the composition of grapes and as such may be under the control of grapevine genes. Accordingly, it should be possible to map the genes responsible for controlling the concentration of these compounds in wine if relevant genetic differences exist in a population of plants resulting from a biparental cross. In order to achieve this, existing crosses were genotyped in order to construct genetic maps for use in quantitative trait loci (QTL) analysis. Corresponding micro-scale fermentations were conducted under controlled conditions to eliminate winemaking variables and the volatile compounds quantified using solid phase microextraction - gas chromatography mass spectrometry. These data sets were then combined in a QTL analysis to identify genomic regions associated with the concentration of volatile compounds in wine. In some cases, the regions delineated by the QTL analysis as being associated with wine volatile composition were made smaller through the design of cleaved amplified polymorphic sequences (CAPS) or derived-CAPS (dCAPS) markers. In other cases, candidate genes in the QTLs were studied for their role in the phenotype observed with the idea of producing perfect markers.

In parallel to the QTL analyses, we also conducted genome-wide association studies (GWAS) to confirm the biparental mapping or to reveal new QTLs for traits not mapped in the biparental crosses. GWAS looks for statistically significantly associated genome variants that are associated with a certain phenotype. In our case we used genotyping by sequencing technologies provided by Diversity Arrays Technology (DArT) and used the markers generated to look for associations with the concentration of volatile compounds from micro-scale fermentations. In this case the plant material used was sourced from the variety collection located at the SARDI vineyards in Nuriootpa. Significant associations were then identified on the grape genome and candidate genes in the vicinity of the markers identified from curated gene lists.

Molecular and biochemical methods (e.g. enzyme and gene expression assays) were also used to study candidate genes identified in the mapping studies. The ultimate aim is to then produce perfect markers based on the causal gene and to improve our knowledge of how the biochemical pathways of secondary metabolite accumulation are controlled during berry development.

5.2 Grape Material and Sampling

5.2.1 Biparental crosses

Biparental crosses between the *VvGAI1*/*Vvgai1* heterozygous mutant dwarf grapevine (Boss and Thomas 2002) originating from the outer cell layer of Pinot Meunier (henceforth referred to as either ‘microvine’ and abbreviated as ‘Mv’) and either Riesling (R) or Cabernet Sauvignon (CS) had been conducted previously in internal CSIRO projects for gene mapping of grapevine traits. To produce these F1 populations, reciprocal crosses were conducted, and seed randomly chosen from

bunches for germination via methods already described (Chaib, et al. 2010). Plants were grown under 16 h days at 25-30°C and 20-25°C nights in a glasshouse. Bunches were tagged when berries began to soften according to touch and all of the berries on each bunch were sampled seven weeks later and frozen at -80°C pending further analysis.

An F2 population from a single genotype from the Mv × R F1 population (plant #131) was also generated during the project. Seed was collected from this plant and vegetatively propagated clones and again germinated using methods described previously (Chaib, et al. 2010). Plants were grown under the same conditions as above and berries sampled seven weeks after veraison.

5.2.2 Germplasm collections

We are grateful to SARDI for providing access to the variety collection based at their research vineyard in Nuriootpa. Despite the differences in the date of grape maturity among the accessions, we chose to sample berries from the cultivars on the same day, aiming to be somewhat later in the season so the majority of genotypes would be above 21 °Brix. In 2019, mostly red varieties were sampled (Table 1) and the mean °Brix was 22.8, with a range from 11.6 to 29.6, although only 13 of the 96 sampled were below 20 °Brix. In 2020, white varieties were mainly sampled (Table 1) and the mean °Brix was 23.0, ranging from 13.4 to 28.1. Since we normalise sugar concentration during the winemaking, this variable is removed from the winemaking process, but obviously the stage of ripening is quite different. Nevertheless, we thought that sampling on one day would remove environmental effects that could differ when fruit is exposed to different conditions throughout the differing ripening periods the range of varieties would experience if all were taken at the same °Brix. Also, we considered that any significant associations observed would be very robust given they would be independent of grape maturity. We also conducted GWAS on the °Brix values to see if significant QTLs were detected. If so, any volatile compound QTLs identified in the same genome regions as the °Brix QTLs may have mapped there due to the effect of grape maturity on the concentration of these compounds in the wines.

Table 1. Varieties from the SARDI collection at Nuriootpa sampled in 2019 and 2020.

VARIETY NAME	VARIETY NAME	VARIETY NAME
	Sampled in 2019	
Abouriou	Aleatico	Alicante Bouschet
Americano	Aramon	Arinarnoa
Auldana No. 1	Auldana No. 3	Baco noir
Barbera	Bastardo	Cabernet Franc
Cabernet Sauvignon	Calitor noir	Carignan
Cascade	CG 1730	Chambourcin
Chardonnay	Cinsaut	Clairette
Colombard	Constantia	Couderc noir
Criolla negra	De Chaunac	Egiodola
Emerald Riesling	Fer	Flora

Foch	Freisa	Gamay (D'Arcenant)
Ganson	Gascon	Gouais
Graciano	Gramon	Grec Rose
Grenache	Jacquez	Kadarka
Lagrain	Limberger	Malbec
Malvasia blanc	Mammolo	Mantley
Mataro	Merlot	Mondeuse
Monerac	Montils	Monvedro
Mueller Thurgau	Muscat a petite grains rouge	Muscat Hamburg
Nebbiolo Bourgu	Nebbiolo Fino	Palomino
Pedro Ximenes	Petit Verdot	Pinot Meunier
Piquepoul noir	Portan	Procupak
Quick's Seedling	Raboso Piave	Royalty
Rubired	Ruby Cabernet	Sangiovese
Saperavi	Shiraz	Shiraz (HBP)
Souzao	Sumoll	Tannat
Tarrango	Teroldego	Terret noir
Tinta Ameralla	Tinta Cao	Tinta Carvalha
Tinta Molle	Touriga	Traminer
Trollinger	Valdepenas Tempranillo	Valdiguie
Villard blanc	Zinfandel	
	Sampled in 2020	
Agestsage blanc	Ansonica	Arrilobe
Aurore	Auxerrois	Baroque
Baxter's Sherry	Biancone	Buckland's Sweetwater
Chardonnay	Chasan	Chasselas dore
Chenin Blanc	Clairette Blanche	Crouchen
Doradillo	Dourado	Elvira
Enhresfelser	Fernao Pires	Fetyeska
Furmint	Gouais	Goyura
Green Veltliner	Gros Meslier	Harslevelu
Helena	Himrod	Kekneylu
Les de L'el	Lignan	Liliorila
Maccabeu	Malvasia Bianca	Marsanne
Melon	Melvasia Rei	Montils
Mtsvase	Mueller Thurgau	Ondenc
Ortruge	Palomino	Pannaonia Gold
Parellada	Pedro Ximenez	Perdea
Perle	Petit Meslier	Pinot Blanc

Pinot gris	Putzscheere	Rabener
Raffiet de Moncade	Red Palamino	Reichensteiner
Rhine Riesling	Rkaziteli	Rolle
Rosulus	Rousanne	Sauvignon blanc
Sauvignonasse	Semillon	Senecca
Seyval	Siegarrebe	Smederevka
Sylvaner	Trebbiano	Tulillah
Valensi blanc	Verdelho	Verdelot
Viognier	Xarelle	Muscat Gordo Blanco
Muscadelle	Canada Muscat	White Muscat
Muscadelle du Borde laise	Muscat Ottenel	Orange Muscat
Morio Muscat		

The second collection used was the 131 remaining 1st generation mildew resistant lines bred as part of previous Wine Australia projects. These plants were generated through crossing the female parent 'R23' which contains the mildew resistance loci with 11 different male parents (Table 2). R23 is a complex cross with its DNA content approximately made up of 50% Cabernet Sauvignon, 25% Aubun, 12.5% Merlot, 6.25% Grenache, 3.12% Cabernet Sauvignon, 1.56% Malaga with 1.56 % *M. rotundifolia* (the total Cabernet Sauvignon content is therefore ~53.12%). The vines were also grown at the SARDI Research Station at Nuriootpa, SA. Each line was represented by six vines and vines were pruned to 20 shoots, 40 buds. To try to account for seasonal effects, data were collected over two successive seasons and grapes for the micro-scale winemaking were harvested at 20 °Brix + 2 weeks in 2020 and 20 °Brix + 3 weeks in 2021.

Table 2 Male parents and number of progeny for the 1st generation mildew resistant lines.

MALE PARENT	NUMBER OF LINES
Chardonnay	5
Dunkelfelder	36
Frontignac	25
Muscadelle	2
Muscat Gordo Blanco	17
Riesling	23
Sangiovese	1
Shiraz	2
Tannat	14
Tempranillo	1
Verdelho	5

5.2.3 Grape berry developmental series samples

Cabernet Sauvignon, Riesling and Pinot gris berry samples were harvested during the 2018/19 season from a commercial vineyard located in Eden Valley, South Australia. Samples were taken in triplicate from three areas of the vineyard at fortnightly intervals beginning 20 November 2018 with flower samples taken from all three varieties. On this date, cap-fall had occurred in Pinot gris, and was close to occurring in the Riesling inflorescences. As such flowering (defined as 50% cap fall) was predicted to have occurred on 20 November for Pinot gris, 27 November for Riesling and 4 December for Cabernet Sauvignon. Samples were taken as whole bunches, placed in plastic bags and transported on ice to the laboratory at the Waite Campus where individual berries were removed from bunches, snap frozen in liquid nitrogen and stored at -80 °C until further use.

A subsample of fruit was taken from each variety at each time point after two weeks post-flowering (wpf) and 30 berries measured individually for berry weight and °Brix using an RFM710 digital refractometer (Bellingham Stanley, Tunbridge Wells, UK). The changes in °Brix followed the normal pattern (Figure 1), with veraison estimated at between 8-9 wpf for Cabernet Sauvignon and Pinot gris, and between 9-10 wpf for Riesling. Berry weight matched the °Brix results for Pinot gris and Cabernet Sauvignon, with the end of the lag phase in berry weight increase coinciding with the increase in sugar accumulation at some time after 8 wpf (Figure 2). However, the Riesling fruit began to expand between 7 and 9 wpf before the onset of sugar accumulation. Commercial harvest was just after 16 wpf for Cabernet Sauvignon, 15 wpf for Riesling and 14 wpf for Pinot gris.

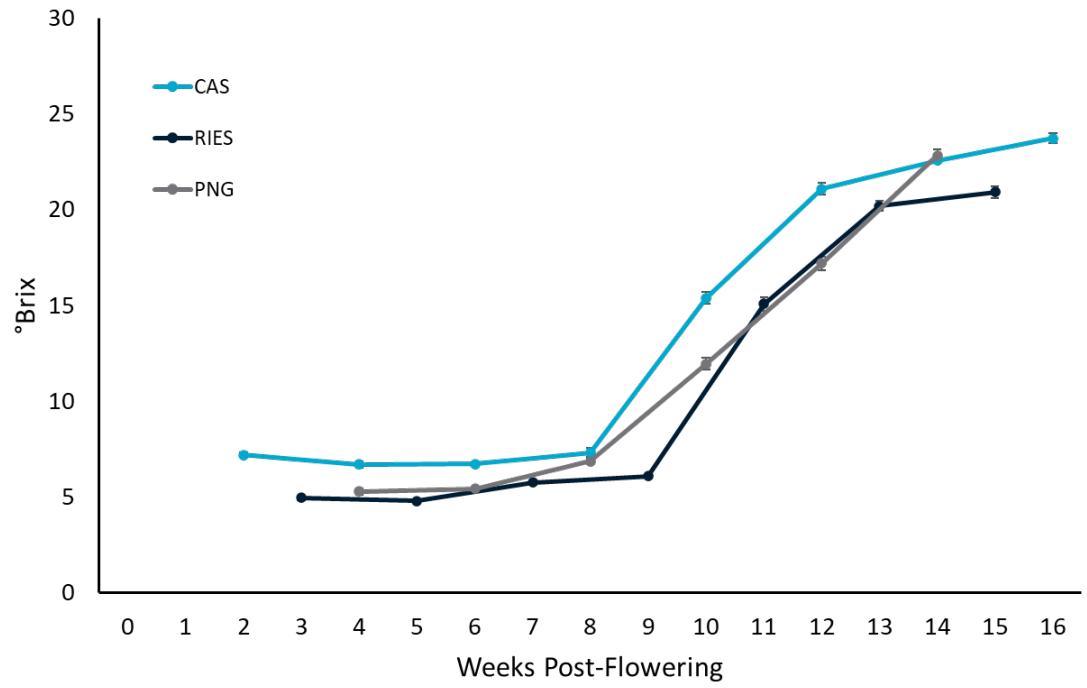


Figure 1 Changes in °Brix during the 2018/19 growing season in Cabernet Sauvignon (CAS), Riesling (RIES) and Pinot gris (PNG) fruit sampled at regular intervals from the same vineyard in Eden Valley.

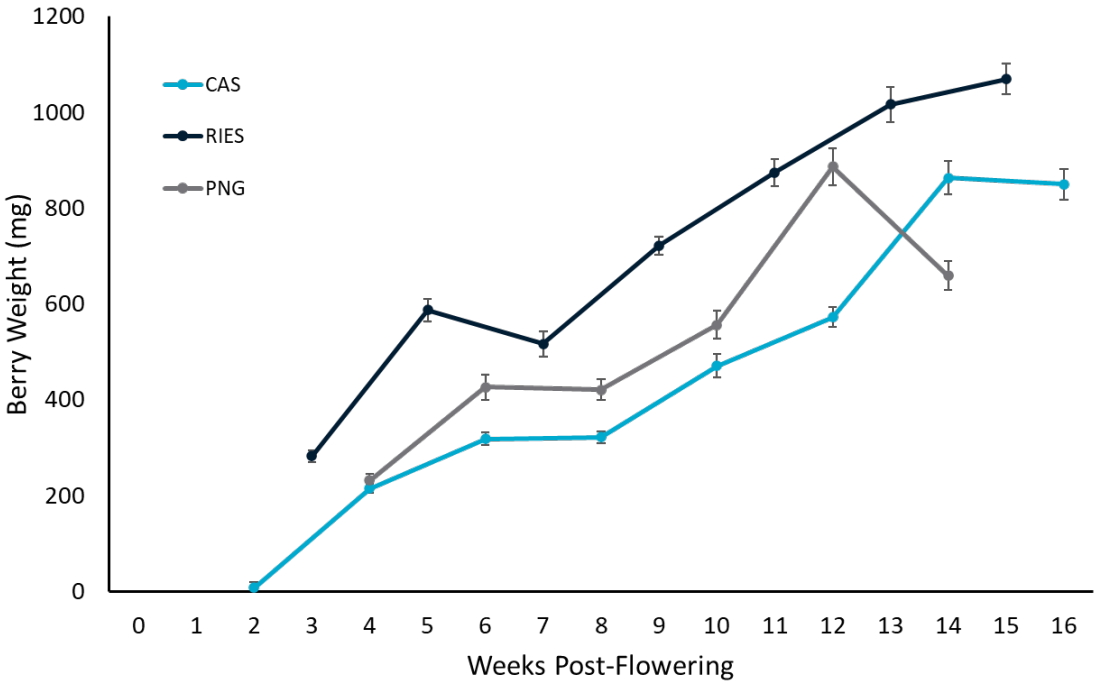


Figure 2 Changes in berry weight during the 2018/19 growing season in Cabernet Sauvignon (CAS), Riesling (RIES) and Pinot gris (PNG) fruit sampled at regular intervals from the same vineyard in Eden Valley.

5.3 Genotyping and Phenotyping

5.3.1 Biparental crosses

Cabernet Sauvignon, Riesling, Mv and F1 and F2 progeny from reciprocal crosses between the Mv and the wild-type parents, were grown under 16 h days at 25-30°C and 20-25°C nights in a glasshouse. Bunches were tagged when berries began to soften according to touch and all of the berries on each bunch were sampled seven weeks later and frozen at -80°C pending further analysis.

5.3.2 DArTseq for GWAS

DNA extraction was carried out by AGRF (Waite Campus, Adelaide, Australia). SNP genotyping was carried out by Diversity Array Technologies (University of Canberra, Australia) using the DArTseq technology. For the GWAS pane 112026 markers were generated, whereas 30899 SNPs were obtained for the 1st generation mildew resistant plants and the F2 population of plant #131 from the Mv × R cross (Mv×R131). SNP data were analysed and filtered using the DartR package (Gruber, et al. 2018) in RStudio version 1.4.1717 (R Core Team 2019). SNPs were filtered based on the following parameters: call rates (callrate) > 0.95, average repeatability (Repavg) > 0.995 and read depth (rdepth): lower= 5, upper= 60. SNPs with a minor allele frequency (maf) < 0.05 were excluded from further analysis. SNPS were mapped to the *Vitis vinifera* PN40024_V1 genome 12x.

5.3.3 SSR and SNP marker analysis

Progeny from the Mv × CS cross were genotyped using 251 SNP and 33 SSR markers and the Mv × R progeny using 229 SNP, 37 SSR and 23 dCAPS markers. DNA extraction was conducted by the Australian Genome Research Facility (AGRF) using a NucleoSpin 96 Plant II kit (Macherey-Nagel, Düren, Germany). The SSR markers were analysed by AGRF according to the protocol described in Dunlevy et al. (2013) and the SNP genotyping was performed using he MassARRAY® iPLEX Gold assay with matrix assisted laser desorption ionisation time-of-flight mass spectrometry (MALDI-TOF MS) detection (Sequenom, San Diego, CA). dCAPS markers were analysed as outlined in Dunlevy et al. (2013).

5.3.4 Micro-fermentation

Micro-fermentation was carried out essentially as described in Blackford et al. (2022) with some modifications. First, the volume of grape material was reduced to 7.5 g and then 7.5 mL total of a mixture of a sterile 550 g/L fructose:glucose (1:1) solution plus sterile water to achieve the same final sugar content in each must was added to a sterile 50 mL bioreactor tube (Techno Plastic Products, Trasadingen, Switzerland). In a laminar flow cabinet, 75 µL of 60 g/L ammonium chloride and 150 µL of an overnight culture of EC1118 yeast that had been washed and diluted to an OD_{600} = 1 were added to each must. The tubes were closed and allowed to ferment in the dark with mixing twice daily. When the weight loss had stabilised, the tubes were centrifuged at 2000 × *g*, the supernatant decanted into 15 mL falcon tubes and stored at 4°C until analysis by GC-MS. Fermentations were done in biological duplicate at least two weeks apart to test for yeast and fermentation effects on volatile compound production.

5.3.5 Wine volatile compound analysis

Wine volatile compounds were analysed using the method described in Boss et al. (2014). One hundred and seventy-seven compounds were quantified in the various wines and their retention indices and method of identification are listed in Table 3.

Table 3. Compounds for which relative quantities were obtained in the volatile fractions of the wine samples.

COMPOUND	RETENTION INDICES[1]	IDENTIFICATION[2]
Ethyl acetate	888	A
Ethanol	936	A
Ethyl propanoate	953	A
Ethyl isobutyrate	960	A
Propyl acetate	968	A
Diacetyl	972	B
Decane	992	A
α-Pinene	1009	A
Isobutyl acetate	1009	A
Ethyl butanoate	1033	A
1-Propanol	1040	A

COMPOUND	RETENTION INDICES[1]	IDENTIFICATION[2]
Ethyl 2-methylbutanoate	1050	A
2,3-Pentadione	1059	A
Ethyl isovalerate	1067	A
Hexanal	1080	A
2,6,6-Trimethyl-2-vinyltetrahydropyran	1103	A
Isobutanol	1107	A
Isoamyl acetate	1121	A
Ethyl pentanoate	1125	A
β-Pinene	1128	A
2,3-Heptadione	1142	B
β-Myrcene	1143	A
1-Butanol	1149	A
Ethyl crotonate	1152	A
2-Carene	1153	B
α-Terpinene	1154	A
1,4-Dimethylbenzene	1164	B
2-Heptanone	1169	A
Limonene	1170	A
Methyl hexanoate	1170	A
Isoamyl propanoate	1172	B
Ethyl 4-methylpentanoate	1174	B
2-Methylbutyl isobutyrate	1178	B
1,8-Cineole	1182	A
2-Methyl-1-butanol	1194	A
Isoamyl alcohol	1201	A
γ-Terpinene	1203	A
β-Ocimene	1209	A
Ethyl hexanoate	1235	A
1-Pentanol	1236	A
Styrene	1233	A
Isoamyl butanoate	1238	B
***p*-Cymene**	1239	B
Hexyl acetate	1256	A
Terpinolene	1256	A
3-Hydroxy-2-butanoate	1270	A
Ethyl (*E*)-3-hexenoate	1273	B
Ethyl (*Z*)-3-hexenoate	1283	A
(*E*)-3-Hexyl acetate	1284	A

COMPOUND	RETENTION INDICES[1]	IDENTIFICATION[2]
2,2,6-Trimethylcyclohexanone	1288	A
(Z)-3-Hexyl acetate	1292	B
4-Methyl-1-pentanol	1292	A
Propyl hexanoate	1293	B
2-Heptanol	1304	A
Unknown monoterpene	1305	C
3-Methyl-1-pentanol	1310	A
Ethyl heptanoate	1319	A
6-Methyl-5-hepten-2-one	1320	A
Ethyl lactate	1323	A
Ethyl (*E*)-2-hexenoate	1324	A
1-Hexanol	1338	A
Isobutyl hexanoate	1339	B
(*E*)-3-Hexen-1-ol	1347	A
Heptyl acetate	1350	A
3-Ethoxy-1-propanol	1360	A
(*Z*)-3-Hexen-1-ol	1367	A
2-Nonanone	1373	A
Methyl octanoate	1374	A
Nonanal	1375	A
β-Isophorone	1378	A
Ethyl octanoate	1428	A
Unknown monoterpene	1429	C
1-Octen-3-ol	1432	A
1-Heptanol	1433	A
Isopentyl hexanoate	1436	B
6-Methyl-5-hepten-2-ol	1437	B
Furfural	1444	A
Nerol oxide	1446	B
Octyl acetate	1453	A
Acetic acid	1455	A
Nonyl 2-furoate	1457	C
Unknown monoterpene	1459	B
Ethyl 7-octenoate	1462	B
Theaspirane A	1475	B
2-Ethyl-1-hexanol	1483	A
Unknown monoterpene	1484	C
3-Ethyl-4-methylpentanol	1486	B

COMPOUND	RETENTION INDICES[1]	IDENTIFICATION[2]
Geranyl ethyl ether	1493	B
Ethyl 3-hydroxybutanoate	1497	B
2-Nonanol	1501	A
Propyl octanoate	1502	A
Benzaldehyde	1506	A
Vitispirane 1 and 2	1507	B
2-Methyltetrahydrothiophen-3-one	1517	B
Theaspirane B	1518	B
Ethyl nonanoate	1519	A
2,3-Butanediol	1520	A
Propanoic acid	1520	A
Linalool	1526	A
1-Octanol	1535	A
Ethyl 3-(methylthio)propionate	1543	A
Isoamyl lactate	1549	B
Methyl decanoate	1573	A
2-Undecanone	1576	A
Terpinen-4-ol	1578	A
Hotrienol	1593	B
β-Cyclocitral	1600	A
Ethyl 2-furoate	1606	A
γ-Butyrolactone	1607	A
Ethyl methyl succinate	1613	B
Ethyl decanoate	1626	A
Butanoic acid	1633	A
1-Nonanol	1638	A
(*Z*)-β-Farnesene	1640	A
Citronellyl acetate	1641	A
Isoamyl octanoate	1646	A
Ethyl benzoate	1646	A
2-Methyl butanoic acid	1649	B
Diethyl succinate	1660	A
(*Z*)-3-Nonen-1-ol	1663	A
Ethyl 9-decenoate	1685	B
Isovaleric acid	1686	A
α-Terpineol	1686	B
Borneol	1688	B
Ethyl (*E*)-4-decenoate	1689	A

COMPOUND	RETENTION INDICES[1]	IDENTIFICATION[2]
Unknown norisoprenoid	1695	C
3-(Methylthio)-1-propanol	1700	A
Propyl decanoate	1712	B
2,4,4-Trimethyl-3-(3-methylbutyl)cyclohex-2-enone	1717	C
TDN	1731	B
Cabreuva oxide	1732	C
Geranyl acetate	1735	A
(*Z*)-β-Damascenone	1737	B
1-Decanol	1742	A
β-Citronellol	1748	A
Methyl salicylate	1754	A
Diethyl pentanedioate	1759	B
Ethyl phenylacetate	1768	A
Ethyl 4-hydroxybutanoate	1775	B
Nerol	1783	A
Ethyl salicylate	1788	A
2-Phenylethyl acetate	1797	A
(*E*)-β-Damascenone	1802	A
Ethyl dodecanoate	1821	A
(*E*)-Geraniol	1827	A
Unknown norisoprenoid	1830	C
2-Methoxyphenol	1833	B
Hexanoic acid	1837	A
Isoamyl pentadecanoate	1854	C
Benzyl alcohol	1865	A
3-Hydroxy-2,2,4-trimethylpentyl 2-methylpropanoate	1884	B
Ethyl 3-hydroxydodecanoate	1890	B
Ethyl isopentyl succinate	1891	B
α-Calacorene	1892	B
2,6-Ditert-butyl-4-methylphenol	1893	C
2-Phenylethyl alcohol	1895	A
β-Ionone	1915	A
2-Phenylethyl butyrate	1943	A
2-Ethyl hexanoic acid	1950	B
γ-Nonalactone	2010	A
Nerolidol	2021	A
Diethyl malate	2025	A

COMPOUND	RETENTION INDICES[1]	IDENTIFICATION[2]
Amberonne	2034	C
Ethyl tetradecanoate	2045	A
Octanoic acid	2052	A
Ethyl 3-hydroxytridecanoate	2088	C
Ethyl cinnamate	2168	A
Nonanoic acid	2198	A
4-Vinyl guaiacol	2222	A
1,6-Dimethyl-4-(1-methylethyl)naphthalene	2231	B
Thymol	2234	B
Ethyl hexadecenoate	2240	A
Decanoic acid	2287	A
2,4-Di-tert-butylphenol	2311	B
9-Decenoic acid	2343	A
4-Vinylphenol	2355	B
2-Phenylethyl octanoate	2370	B
Ethyl hydrogen succinate	2375	B
Dodecanoic acid	2504	B

[1] RI calculated from retention relative to the retention of a series of n-alkanes (C_8-C_{26}). [2]A, identity confirmed by matching mass spectra and RI with that of authentic standards; B, tentative assignment based upon comparison with mass spectral libraries and published RI; C, unknown compound, possible identity listed if match with mass spectral library is strong.

5.3.6 Grape volatile compound analysis

Normal extraction temperature

Linalool was quantified in berries sampled from the Mv×R131 F2 population by using solid phase microextraction (SPME) in conjunction with gas chromatography-mass spectrometry (GC-MS). Whole grapes were frozen in liquid nitrogen and ground to a powder using an IKA A11 analytical mill (IKA, Staufen, Germany). Then 2.5 g of frozen grape powder, 2.5 g of NaCl and 2.5 mL of 4 mM ZnSO4 were added to a 20 mL crimp-capped amber vial. A 50 µL aliquot of the internal standard, (*Z*)-3-hepten-1-ol (3.4 mg/L, Sigma-Aldrich), was then added to the vial. SPME-GC-MS was carried out using an Agilent 7890A gas chromatograph (Santa Clara, CA, USA) equipped with a Gerstel MP2 auto-sampler (Mülheim an der Ruhr, Germany) and using an Agilent 5975C mass spectrometer. The samples were equilibrated at 40°C for 5 min and then extracted at 40°C for 30 min with agitation at 250 rpm utilising a divinylbenzene-carboxen-polydimethylsiloxane (DVB-CAR-PDMS) fibre (2 cm, 23-gauge, 50/30 µm, Supelco, Bellefonte, PA, USA). Analytes were desorbed in the injector fitted with a 0.75 mm internal diameter SPME injection sleeve (Supelco) in splitless mode for 1 min at 220°C after which the purge was switched on to 50 mL/min. Chromatography was performed using a ZB-Wax column (30 m, 0.25 mm i.d., 0.25 µm film thickness; Phenomenex, Lane Cove, NSW, Australia) using helium as a carrier gas at a constant flow of 1.5 mL/min. After an initial hold at 35°C for 1.5 min, a temperature gradient of 7°C/min from 35°C to 245°C was applied and the oven was then held isothermally for 3.5 min. Mass spectrometry was performed in positive ion mode with a transfer line temperature of 250°C, source temperature of 230°C, quadrupole temperature of 150°C, and

electron impact ionization at 70 eV, which was recorded in scan mode (range, *m/z* 35-350; scan rate, 4.45 scans/s). Chromatograms were analysed for linalool and the internal standard using MassHunter Workstation Version 10.2 (Agilent).

High extraction temperature

In order to extract the less volatile compounds from the berry samples, they were also analysed using a method that uses a high SPME extraction temperature before separation by GC. This method is based on one developed for grapevine leaves (Weingart, et al. 2012) combined with a protocol for using an externally sampled internal standard as described in Wang et al. (2005). Sample preparation was simply the addition of 200 mg of frozen and ground grape berry tissue to a 20 mL amber SPME vial. The samples were then equilibrated at 90°C for 25 min with agitation at 300 rpm. Before sample extraction, the internal standard was adsorbed onto the 2 cm DVB-CAR-PDMS SPME fibre by incubation with agitation in an amber vial containing 20 µL of a standard solution (1.84 g/L d_{13}-hexanol; 0.874 g/L methyl nonanoate; 0.865 g/L d_3-linalool; 2.08 g/L d_7-benzyl alcohol; 4.45 g/L d_{12}-hexanal) in 4 g of vacuum pump fluid (Inland 45; Inland Vacuum Industries, Churchville, NY) at 90°C for 5 min. Samples were then extracted onto the SPME fibre at 90°C for 60 min with agitation at 300 rpm. The compounds were then desorbed from the fibre onto the column via an inlet for 1 min at 220°C in splitless mode. The fibre was cleaned before and after extraction by incubating in a bakeout station for 4 min at 240°C. Chromatography was identical to the 'normal extraction temperature' method immediately above this section as were the mass spectrometer settings. Chromatograms were again analysed for compounds of interesting using MassHunter Workstation Version 10.2 (Agilent).

5.3.7 Statistical analysis

When required, the effects of sample or treatment were tested using ANOVA with IBM SPSS version 23 (IBM; Armonk, NY, USA) and followed by Duncan's posthoc test. When means were not significantly different across the samples (p>0.05), the compounds were eliminated from further analyses. Cluster analyses were performed using all of the volatile compounds and the data were normalised by centring around the mean and scaled by the standard deviation (i.e. converted to a Z score). This allows for trends to be seen in the data without complications associated with the broad range of concentrations observed for the compounds across the samples. Samples were then sorted by hierarchical complete-linkage clustering using Unscrambler X 10.1 (Camo, Oslo, Norway) employing an uncentred correlation distance metric for the standardised data. Principal components analysis (PCA) and partial least squares – discriminant analysis (PLS-DA) of the data were also conducted using The Unscrambler X 10.1.

5.4 Gene mapping and marker development

5.4.1 QTL analysis

Marker linkage for the populations was analysed using JoinMap 4.1 (Van Ooijen 2006). The χ^2 test was used to test each marker for goodness to fit to expected segregation, but distorted markers were initially retained as there was expected to be distortion on linkage group 1 as microvines, with

a mutant *vvgai* allele were selected as progeny to ensure rapid fruit production. Other distorted markers subsequently found to be of low quality were removed from the analysis. For the F1 populations, linkage groups were determined using the independence LOD for a population type 'CP' (outbreeder full-sib family) and the maximum likelihood mapping algorithm under default settings. The resulting consensus maps were used for the QTL analysis. The linkage groups for the Mv×R131 F2 population were determined as above but using the 'F2' population type.

The detection of QTLs related to the concentration of specific volatile compounds in the wines was carried out with MapQTL v6 software (Van Ooijen 2009). As a first step the non-parametric Kruskal-Wallis (KW) rank-sum test method was applied to identify markers significantly associated with the phenotype. Following this multiple QTL mapping (MQM) was conducted using the markers deemed as closest to the QTLs from the KW analysis as cofactors and the default parameters. LOD significance thresholds for $p<0.05$ were calculated using a permutation test analysis employing 1000 permutations at the genome-wide level (Churchill and Doerge 1994).

5.4.2 Genome-wide association studies

Prior to GWAS analysis, the correlation between the volatile compound quantification for both years was observed by plotting one replicate against the other and compounds with R^2 values greater than 0.64 in one vintage were considered for further analysis. When the data were not normally distributed a Box-Cox transformation was applied to the data using online software to calculate the optimum lambda value for the transformation (http://www.wessa.net/rwasp_boxcoxnorm.wasp). A total of 10,510 informative SNPs selected based on the outcome of the filtering process mentioned above were included for the mapping process using the SARDI variety collection and 16,938 for the 1st generation mildew resistant lines. GWAS was carried out in GAPIT version 3 package (Wang and Zhang 2021), in RStudio version 1.4.1717 (R Core Team 2019). To assess the reproducibility of the GWAS, the marker-trait association analysis was carried out using the following models: MLMM (Segura, et al. 2012), FarmCPU (Liu, et al. 2016) and BLINK (Huang, et al. 2019). We accounted for possible errors arising from population structure (Q) and kinship (K) in the GWAS analysis by using principal components (PCs) in the GAPIT analysis. The optimal number of PCs/Covariates was determined using the Bayesian information criterion (BIC) model selection in a two-step process; a preliminary GWAS was performed in which the number of PCs/Covariates was set to 10. The outcome of this analysis revealed that the fourth component explains only ~ 2% of the variation in the population, and components after that explain only ~ 1%. Based on these results, the number of PCs/Covariates was set to 3 and the model selection function in GAPIT was enabled in the final GWAS analysis. The significance threshold of the adjusted *p* value (q value) after performing the false discovery rate (FDR) was set to $p=0.05$. SNPs with an adjusted $p<0.05$ were considered significant.

5.4.3 Marker development

Following the discovery of a single nucleotide polymorphism (SNP) or other DNA change associated with an important phenotype, a simple way to detect the presence or absence of this difference in DNA sequence is to design a marker to identify it in a plant's DNA. In this project it was decided to design what are known as 'cleaved amplified polymorphic sequences (CAPS)' or 'derived CAPS (dCAPS)' markers to our traits of interest. CAPS markers rely on the fact that SNPs can create or

abolish restriction enzyme recognition sites in the DNA. Therefore, oligonucleotide primers can be designed either side of the polymorphic restriction enzyme site so the genomic region can be amplified using PCR and digested to determine if the SNP is present or not. We chose to design primers asymmetrically either side of the SNP and aimed to amplify a region of DNA between 500-800 bp long. For some SNPs, there is no restriction enzyme site altered due to the sequence change. In these cases, we can introduce changes in one of our oligonucleotide primers to create a restriction enzyme site if a SNP is present (or alternatively absent) and these are called dCAPS markers. For dCAPS markers, the primer that introduces the restriction enzyme site is designed to be >30 bp long as this will be difference in band sizes of the PCR products which are difficult to observe if they are too small. Our preferred amplicon length for dCAPS primers was 150-250 bp. The design of both CAPS and dCAPS markers was aided by the use of an online software tool called 'dCAPS Finder 2.0' (http://helix.wustl.edu/dcaps/).

5.5 Candidate gene functional characterisation

5.5.1 Gene expression studies

Total RNA was extracted from grape berries as described by Boss et al. (1996) and purified according to D'Onofrio et al. (2009). First-strand cDNA was synthesised with SuperScript III Reverse Transcriptase (Invitrogen, Massachusetts, USA) as per manufacturer's instructions. Prior to qRT-PCR analysis, the specificity of each amplified fragment was tested by separation on agarose gel, sequencing of purified fragments using Sanger sequencing (AGRF, Adelaide, South Australia) and melt curve analysis. qRT-PCR analysis was carried out as described by Böttcher et al. (2013a). In brief, the Roche LightCycler 480 SYBR Green I Master kit was used for the analysis and a Roche LightCycler 480 system was used to amplify the gene specific fragments with the following program: 5 min at 95°C, 20 s at 95°C, 20 s at 58°C and 20 s at 72°C (45 cycles), 5 min at 72°C, followed by a melt cycle (15 s at 95°C, 45 s at 50°C, continuous heating to 95°C at 0.11°C s-1). The reference gene used was *ACT2* (*actin2*) as reported in Böttcher et al. (2013b) and the expression value of each gene in the cDNAs was normalised to the expression of this reference gene. The significance of any differences between treatments was tested by ANOVA followed by Duncan's post hoc test, using SPSS ver.16.0 (SPSS Inc., Chicago, IL, USA).

5.5.2 Recombinant protein production and assays

NUDX enzymes

Production of recombinant protein

Full-length coding sequences of *VvNUDX1* and *VvNUDX25* were amplified from Riesling berry cDNA and cloned into pDrive cloning vector (QIAGEN, Hilden, Germany) before being inserted into the pET30a (EMD Biosciences Inc, CA, USA) expression vector using the Gibson Assembly Cloning Kit (NEB, Massachusetts, USA) as per manufacture's protocol. BL21 Start (DE3) chemically competent expression cells (Invitrogen, Massachusetts, USA) were transformed with the expression vector harbouring *VvNUDX1* and *VvNUDX25* and single colonies were selected and grown in liquid culture containing kanamycin at 37˚C overnight. A 1 in 40 dilution of the overnight culture was prepared by diluting 25 mL of overnight culture in 500 mL of liquid LB culture containing kanamycin. The culture was

was incubated at 37˚C with shaking. When the bacterial culture density reached an OD600 of ~ 0.5, protein expression was induced by the addition of 0.4 mM IPTG to the culture. Protein induction was carried out overnight at 17˚C with shaking after which the induction culture was centrifuged at 15000 × *g* for 10 minutes at 4˚C. The pellets were resuspended for 1 hour on ice in sodium phosphate buffer; 500 mM sodium chloride (NaCl) and 20 mM of sodium phosphate dibasic (Na_2HPO_4) and sodium phosphate monobasic (NaH_2PO_4), containing 0.2 mg/mL lysozyme, 20 µg/mL DNase I and cOmplete EDTA-free protease inhibitor tablets (1 per 10 mL buffer; Roche, Sydney, Australia). The mixture was sonicated for 15 minutes and centrifuged at 15000g for 20 minutes at 4˚C. The supernatant containing the soluble protein fraction was collected for western blotting and column purification. A His GraviTrap column (GE Healthcare, code: 11003399) was used for purification as per manufacturer's instructions. To elute the bound proteins, 3 mL of sodium buffer containing 50 mM, 100 mM, 200 mM, 300 mM, 400 mM and 500 mM of imidazole were applied to the column and all fractions were collected. SDS-PAGE analysis, western blot and immunodetection were carried out as described in Böttcher et al. (2010). Briefly, 10 µL of each fraction was run on an SDS-PAGE and transferred to a membrane. For immunodetection, a monoclonal antibody raised against poly-histidine (Sigma-Aldrich, St Louis, MO; 1:5000 dilution) subsequently detected by alkaline phosphatase-coupled anti-mouse polyclonal antibody (Promega, Madison, WI; 1:10000 dilution) was used.

Transient expression in tobacco leaves and analysis of glycosylated monoterpenes

Heterologous expression in *Nicotiana benthamiana* leaves using agroinfiltration was carried out as described by Williams et al. (2016). Full-length coding regions of *VvNUDX1* and *VvNUDX25* from Riesling were PCR amplified with appropriate primers and cloned into the pDNOR223 vector (Invitrogen, Massachusetts, USA) using BP clonase. Entry clones were then inserted in the pEarlyGate100 expression vector (Earley, et al. 2006) containing attR1 and attR2 sites by LR reaction. Agrobacterium (strain EHA105) was transformed with the expression constructs using electroporation. Leaves from three individual plants were agroinfiltrated with each construct. An empty vector was used as a negative control. Leaves were harvested seven days after infiltration and kept at -80°C before analysis.

Glycosylated monoterpenes were extracted as described by (Magnard, et al. 2015). One gram of ground infiltrated leaves was used for each extraction and 5 mL of PBS buffer containing 10% (v/v) ethanol was added to the ground leaves and the mixture was incubated at 4˚C for 1 h with rotation. Octanyl-β-glucoside (50 µL of a 10 g/L solution in methanol) was added as an internal standard. The mixture was centrifuged at 5000g for 10 minutes 4˚C and the supernatant was applied to an C18 SPE column (Agilent Bond Elute, part no:12162028B). The column was washed twice with 5 mL of water and the glycosylated monoterpenes were eluted with 2 mL of 90% ethanol. Detection and relative quantification of geranyl glycoside by HPLC-qTOF MS was carried out using an Agilent 1260 infinity II HPLC coupled to a 6530 accurate mass QTOF LC-MS instrument (Agilent Technologies, Santa Clara, USA) as described by Hjelmeland et al. (2015). An authentic geranyl β-D-glucoside reference standard (Carbosynth Ltd, UK) was used to confirm retention time and obtain MS/MS spectra (acquisition range m/z 50-1000). Obtained spectra, MS retention time and exact mass of the reference standard were incorporated into a manually generated compound library (Agilent MassHunter PCDL Manager B.07.00) and used to identify geranyl glucoside peaks in extracted leaf samples. Using the MS data, extracted compound chromatogram (ECC) peaks generated by the Find-

by-Formula algorithm of the Agilent MassHunter Qualitative Analysis software (ver. 10.0) and representing the sum of all ion peaks corresponding to the geranyl glucoside compound, were integrated (Agile-2 integrator) and divided by the ECC peak area of the internal standard.

NUDX enzyme activity assays

Enzyme assays were performed as outlined by Magnard et al. (2015). Geranyl pyrophosphate in triammonium salt form (Sapphire Bioscience, NSW, Australia) was used as substrate. Concentrations of purified NUDX1 and NUDX25 enzymes were measured prior to the reaction using the Qubit Protein BR Assay Kit (Invitrogen) as per manufacturer's instructions. The reaction buffer consisted of 50 mM HEPES pH 8, 5 mM $MgCl_2$, 1 mM $MnCl_2$, 10 mM β-mercaptoethanol and 10% glycerol. Fifty µg of substrate was used in a total reaction of 100 µL and all assays were carried out in triplicate. Heat inactivated VvNUDX1 and VvNUDX25 enzymes were used as negative controls. Reactions were incubated at room temperature for 30 minutes and stopped by addition of 90 µL of ethanol and 10 µL EDTA 100 mM pH 8. All reactions were analysed using an Agilent 1260 infinity II HPLC coupled to a 6530 accurate mass QTOF LC-MS instrument. Substrates were identified in NUDX assay samples by library matching against an authentic standard and their ECC peaks were generated and integrated as described above for geranyl glucoside. As commercial standards for most of the products were not available, their formulas and exact mass were added to a compound database (Agilent MassHunter PCDL Manager B.07.00) and used to tentatively identify, extract, and integrate peaks in the enzyme assays.

(Z)-3:(E)-2 Hexenal isomerase recombinant protein expression and activity assay

Heterologous expression and western blotting were carried out as described for the NUDX enzymes outlined above. Based on the sequencing results of the coding sequences of the *VvHI* genes the following primers were designed with an adaptor sequence on their 5' prime end for amplification and subsequent insertion of the full-length coding sequence of the *VvHI* genes into the pET30a (EMD Biosciences Inc, CA, USA) expression vector using Gibson assembly: 5'-ATGCACCATCATCATCATCATTCTATGGAGTTGAACTTAGCACCAAAGTT-3', 5'-ATGCACCATCATCATCATCATTCTATGGCGTTGAACTTAGCACCAAA-3' and 5'-CGTCGTCGGTACCCAGATCTCTAGTTCTGTGGAGGAACAAGAATGG-3'.

The pET30a vector was

linearised by PCR amplification using the following primers: 5'- AGATCTGGGTACCGACGACG -3' and 5'- AGATATACATATGCACCATCATCATCATCATTCT-3'. After one hour incubation the reaction mixture was used directly for the transformation of the DH5-α electrocompetent cells.

Transformed bacterial cells were plated on Luria broth (LB) media containing kanamycin and incubated at 37°C overnight. Single colonies were selected and grown in a liquid culture containing kanamycin at 37°C. Plasmids were then extracted from the bacterial cells using Isolate II plasmid extraction mini kit (Bioline, Everleigh, Australia) and sequenced using the T7 forward and reverse primers. BL21 Start (DE3) chemically competent expression cells (Invitrogen, Massachusetts, USA) were transformed with pET30a vector containing the coding sequences of *VvHI* genes plated on media containing kanamycin and grown overnight at 37 ˚C. Single colonies were selected and grown in liquid culture containing kanamycin at 37˚C overnight. A 1 in 40 dilution of the overnight culture was prepared by diluting 25 mL of overnight culture in 500 ml of liquid LB culture containing kanamycin. The culture was incubated at 37˚C with shaking. When the bacterial culture density reached an OD600 of ~ 0.500, protein expression was induced by the of addition of 0.4 mM IPTG to

the culture. Protein induction was carried out overnight at 17˚C with shaking.

Purified enzymes were incubated with 50 µM (Z)-3-hexenal 50% in triacetin (Penta Manufacturing, Livingston USA) in MOPS pH 7.5 for 30 minutes in room temperature. (*Z*)-3-hexenol was added as an internal standard. Volatile compounds were extracted twice from the reaction mixture with 1 mL (2 mL in total) 50% pentane 50% hexane. Analysis of the volatile fraction was carried out as described by D'Onofrio et al., (2009). An Agilent 6890N gas chromatograph fitted with a Gerstel MPS2 multipurpose autosampler and a 5973N Mass Selective Detector. The GC was fitted with a 30 m, 0.25 mm internal diameter, 0.25 µm ZB-Wax capillary column (Phenomenex, CA, USA). 1 µL of the volatile fraction was injected into the column. Injection was done in pulse splitless mode with an inlet pressure of 25.0 psi. The carrier gas was helium at a constant flow rate of 1.2 mL/min. The oven was held at an initial temperature of 40˚C for 1 min, and then increased at 10˚C/min to a final temperature of 240˚C and held at the same temperature for 10 min. For mass spectrometry, a source temperature of 230˚C and a quadrupole temperature of 150˚C were applied with an ionising potential at 70 eV and a scan range of 35-350 atomic mass units. Data analysis was carried in Agilent MassHunter Workstation- Quantitative Analysis for GC-MS and LC-MS V 10.2 (Agilent, CA, USA).

6 Results/Discussion

6.1 Mapping traits using biparental crosses

6.1.1 Introduction

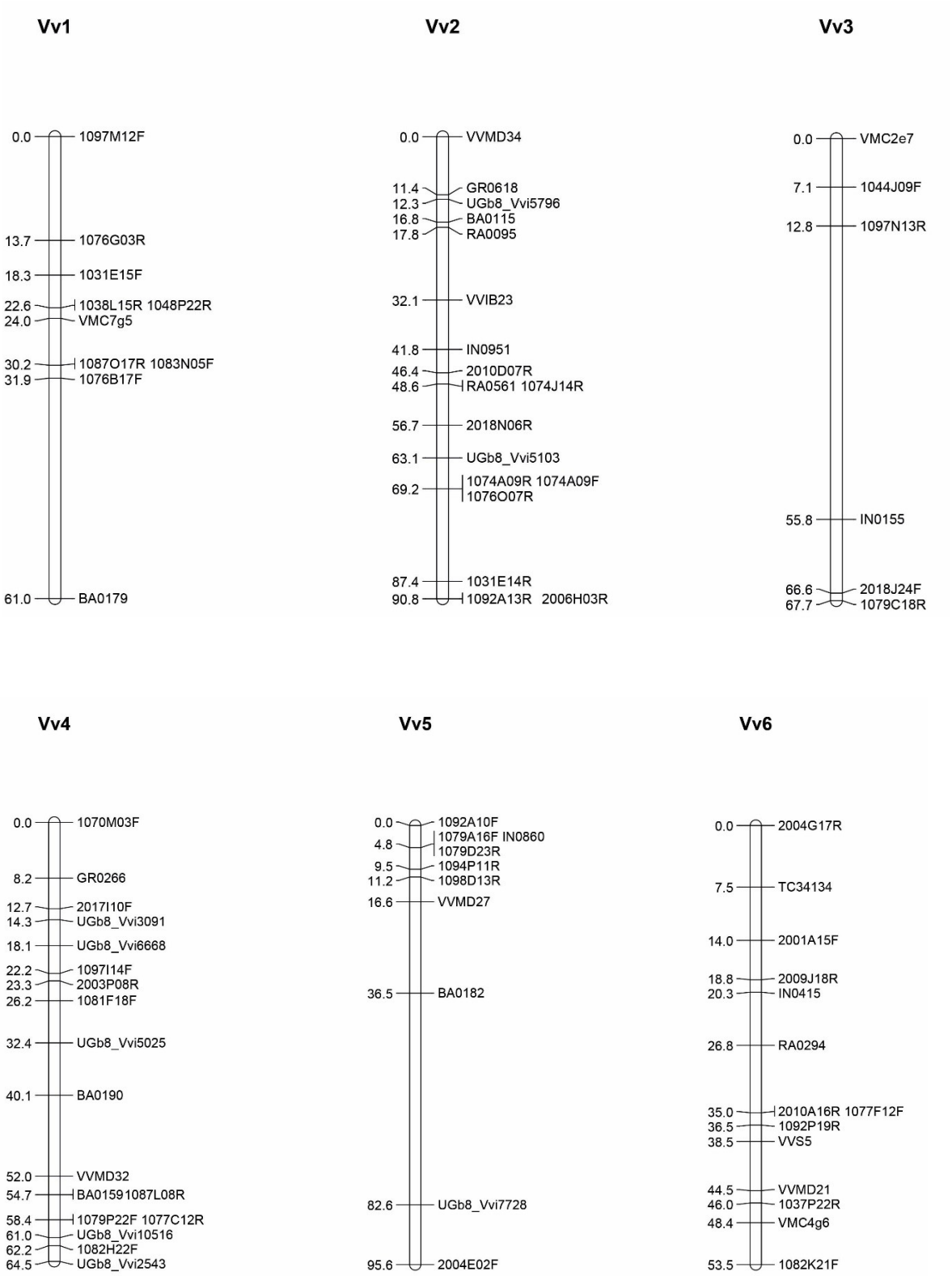
A powerful technique for the development of molecular markers for important crop traits is genetic mapping using biparental crosses. While this is relatively routine in annual crops, the long generation time of perennial crops means that this technique has been used only rarely in grapevine to discover genes associated with consumer-preferred traits (Battilana, et al. 2009, Dunlevy, et al. 2013, Guillaumie, et al. 2013). Furthermore, these studies have been limited to compounds that can be measured in the fruit and they have not attempted to associate markers with the production of certain volatile compounds in wine. Our hypothesis was that if the accumulation of a compound in wine is dependent more on the composition of the fruit than the activity of yeast, then we should be able to map this to a region of the grape genome and therefore generate markers associated with the wine trait. The combination of using the rapid-generation time Mv and micro-scale fermentations allowed this hypothesis to be tested.

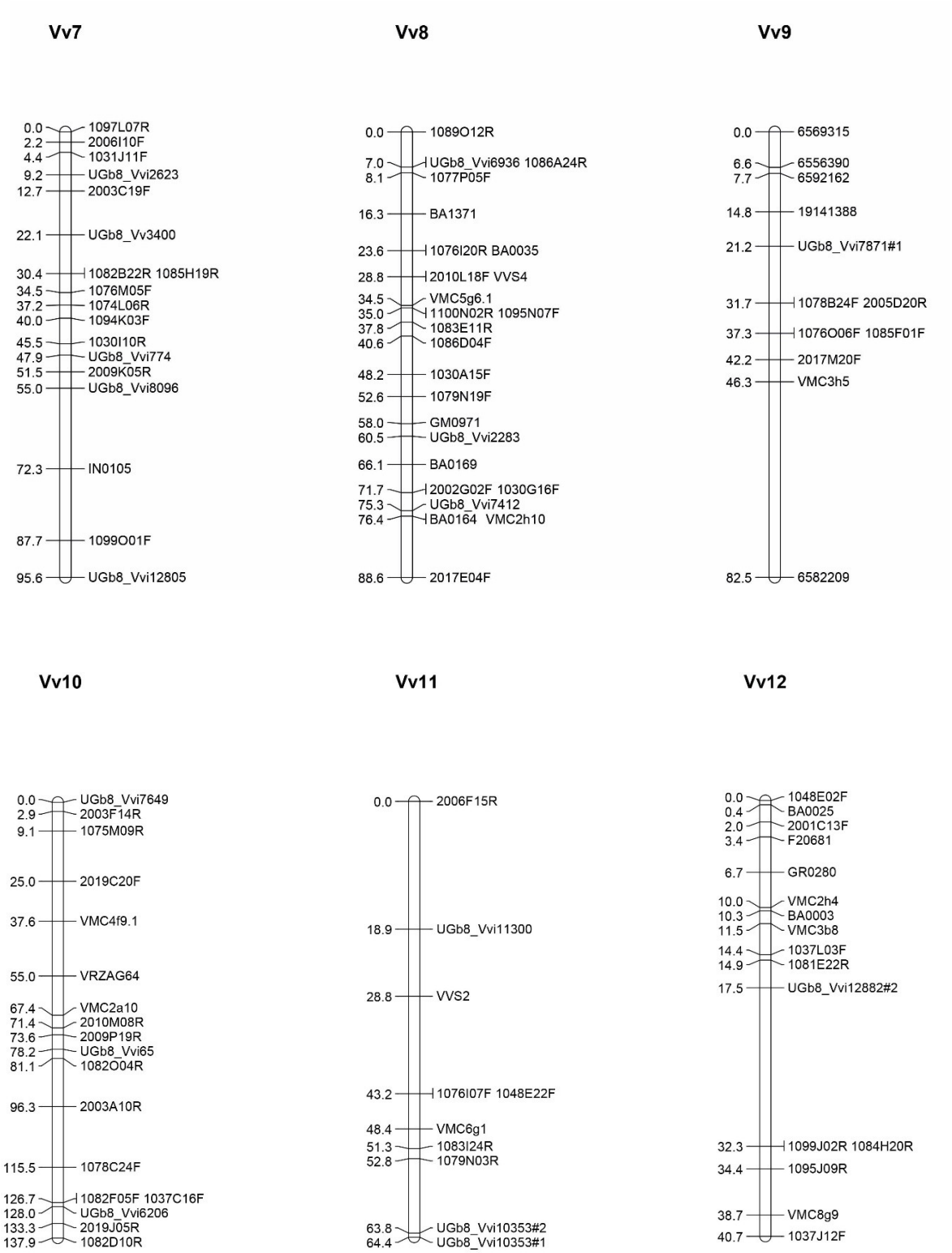
6.1.2 Genetic maps generated from established biparental crosses

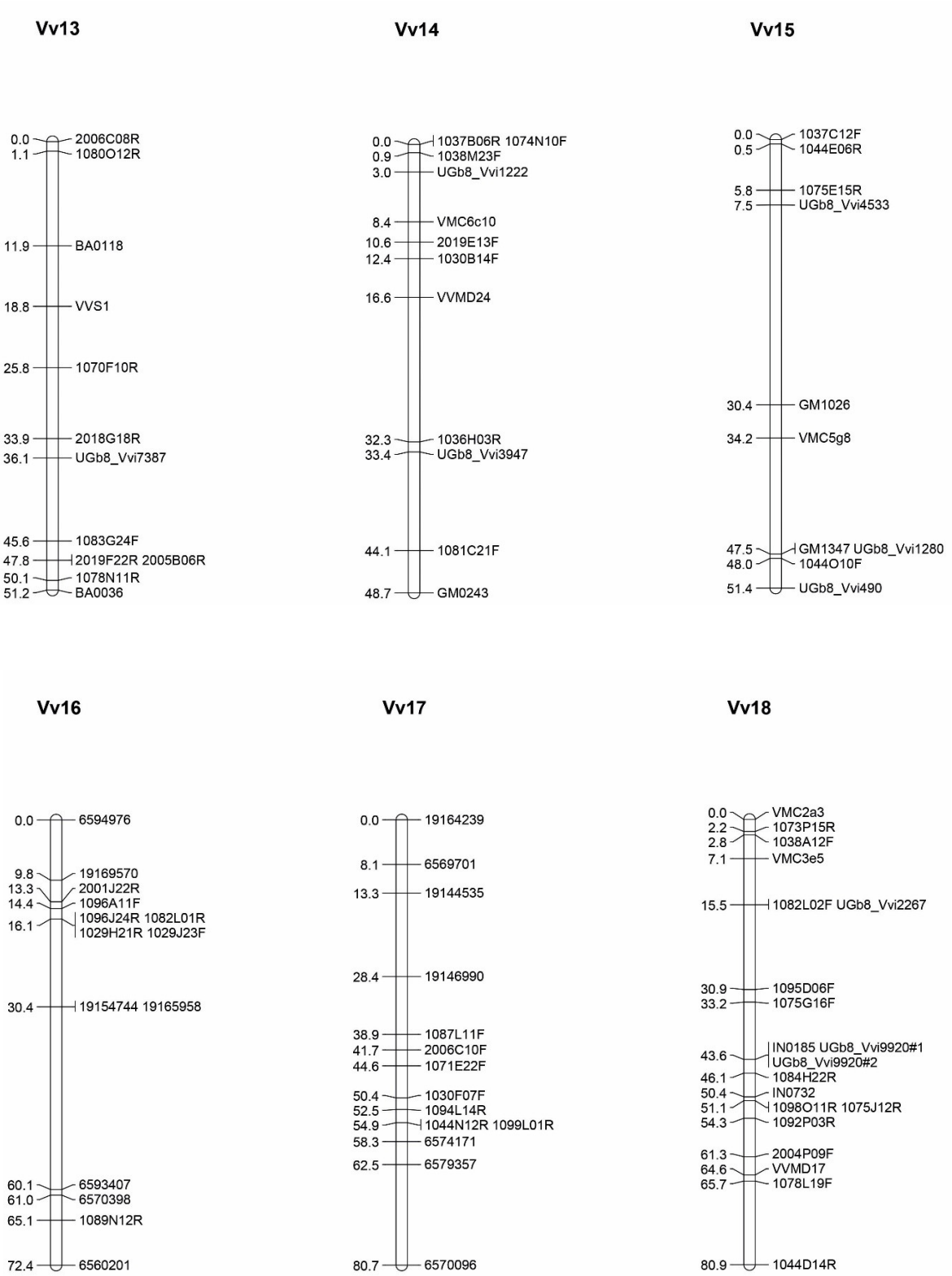
Crosses of microvine × Riesling (Mv × R) and microvine × Cabernet Sauvignon (Mv × CS) had been conducted some years prior in other projects and consisted of ~90 plants for each cross. These were grown under 16 h days at 25–30°C and 8 h nights at 20–25°C in a glasshouse and for the generation of genotypic data, small leaves were harvested, DNA was purified and markers analysed using the methods described above.

Progeny from the Mv × CS cross were genotyped using 251 SNP and 33 SSR markers and the Mv × R progeny using 235 SNP, 37 SSR and 23 dCAPS markers. Using JoinMap 4.1 software, the markers were assembled into 19 linkage groups with the Mv × CS consensus map covering 1371 cM with an average interval length of 5.1 cM (Figure 3). The Mv × R map covered 1367 cM with an average interval length of 4.7 (Figure 4).

The number of markers per linkage group ranged from six to 25 in the Mv × CS map and five to 26 in the Mv × R map. In both cases the linkage group with the least number of markers corresponded to chromosome 3, although the linkage group size was 67.7 cM and 61.6 cM. Linkage group size varied from 40.7 to 137.9 cM in the Mv × CS map and 38.8 to 96.4 cM in the Mv × R map. All of these parameters were similar to other grapevine biparental cross maps generated using the same type of markers (Costantini, et al. 2008).


Vv1
0.0 1097M12F
13.7 1076G03R
18.3 1031E15F
22.6 1038L15R 1048P22R
24.0 VMC7g5
30.2 1087O17R 1083N05F
31.9 1076B17F
61.0 BA0179
Vv2
0.0 VVMD34
11.4 GR0618
12.3 UGb8_Vvi5796
16.8 BA0115
17.8 RA0095
32.1 VVIB23
41.8 IN0951
46.4 2010D07R
48.6 RA0561 1074J14R
56.7 2018N06R
63.1 UGb8_Vvi5103
69.2 1074A09R 1074A09F
1076O07R
87.4 1031E14R
90.8 1092A13R 2006H03R
Vv3
0.0 VMC2e7
7.1 1044J09F
12.8 1097N13R
55.8 IN0155
66.6 2018J24F
67.7 1079C18R
Vv4
0.0 1070M03F
8.2 GR0266
12.7 2017I10F
14.3 UGb8_Vvi3091
18.1 UGb8_Vvi6668
22.2 1097I14F
23.3 2003P08R
26.2 1081F18F
32.4 UGb8_Vvi5025
40.1 BA0190
52.0 VVMD32
54.7 BA01591087L08R
58.4 1079P22F 1077C12R
61.0 UGb8_Vvi10516
62.2 1082H22F
64.5 UGb8_Vvi2543
Vv5
0.0 1092A10F
4.8 1079A16F IN0860
1079D23R
9.5 1094P11R
11.2 1098D13R
16.6 VVMD27
36.5 BA0182
82.6 UGb8_Vvi7728
95.6 2004E02F
Vv6
0.0 2004G17R
7.5 TC34134
14.0 2001A15F
18.8 2009J18R
20.3 IN0415
26.8 RA0294
35.0 2010A16R 1077F12F
36.5 1092P19R
38.5 VVS5
44.5 VVMD21
46.0 1037P22R
48.4 VMC4g6
53.5 1082K21F


Vv7
0.0 1097L07R
2.2 2006I10F
4.4 1031J11F
9.2 UGb8_Vvi2623
12.7 2003C19F
22.1 UGb8_Vv3400
30.4 1082B22R 1085H19R
34.5 1076M05F
37.2 1074L06R
40.0 1094K03F
45.5 1030I10R
47.9 UGb8_Vvi774
51.5 2009K05R
55.0 UGb8_Vvi8096
72.3 IN0105
87.7 1099O01F
95.6 UGb8_Vvi12805
Vv8
0.0 1089O12R
7.0 UGb8_Vvi6936 1086A24R
8.1 1077P05F
16.3 BA1371
23.6 1076I20R BA0035
28.8 2010L18F VVS4
34.5 VMC5g6.1
35.0 1100N02R 1095N07F
37.8 1083E11R
40.6 1086D04F
48.2 1030A15F
52.6 1079N19F
58.0 GM0971
60.5 UGb8_Vvi2283
66.1 BA0169
71.7 2002G02F 1030G16F
75.3 UGb8_Vvi7412
76.4 BA0164 VMC2h10
88.6 2017E04F
Vv9
0.0 6569315
6.6 6556390
7.7 6592162
14.8 19141388
21.2 UGb8_Vvi7871#1
31.7 1078B24F 2005D20R
37.3 1076O06F 1085F01F
42.2 2017M20F
46.3 VMC3h5
82.5 6582209
Vv10
0.0 UGb8_Vvi7649
2.9 2003F14R
9.1 1075M09R
25.0 2019C20F
37.6 VMC4f9.1
55.0 VRZAG64
67.4 VMC2a10
71.4 2010M08R
73.6 2009P19R
78.2 UGb8_Vvi65
81.1 1082O04R
96.3 2003A10R
115.5 1078C24F
126.7 1082F05F 1037C16F
128.0 UGb8_Vvi6206
133.3 2019J05R
137.9 1082D10R
Vv11
0.0 2006F15R
18.9 UGb8_Vvi11300
28.8 VVS2
43.2 1076I07F 1048E22F
48.4 VMC6g1
51.3 1083I24R
52.8 1079N03R
63.8 UGb8_Vvi10353#2
64.4 UGb8_Vvi10353#1
Vv12
0.0 1048E02F
0.4 BA0025
2.0 2001C13F
3.4 F20681
6.7 GR0280
10.0 VMC2h4
10.3 BA0003
11.5 VMC3b8
14.4 1037L03F
14.9 1081E22R
17.5 UGb8_Vvi12882#2
32.3 1099J02R 1084H20R
34.4 1095J09R
38.7 VMC8g9
40.7 1037J12F


Vv13
0.0 2006C08R
1.1 1080O12R
11.9 BA0118
18.8 VVS1
25.8 1070F10R
33.9 2018G18R
36.1 UGb8_Vvi7387
45.6 1083G24F
47.8 2019F22R 2005B06R
50.1 1078N11R
51.2 BA0036
Vv14
0.0 1037B06R 1074N10F
0.9 1038M23F
3.0 UGb8_Vvi1222
8.4 VMC6c10
10.6 2019E13F
12.4 1030B14F
16.6 VVMD24
32.3 1036H03R
33.4 UGb8_Vvi3947
44.1 1081C21F
48.7 GM0243
Vv15
0.0 1037C12F
0.5 1044E06R
5.8 1075E15R
7.5 UGb8_Vvi4533
30.4 GM1026
34.2 VMC5g8
47.5 GM1347 UGb8_Vvi1280
48.0 1044O10F
51.4 UGb8_Vvi490
Vv16
0.0 6594976
9.8 19169570
13.3 2001J22R
14.4 1096A11F
16.1 1096J24R 1082L01R
1029H21R 1029J23F
30.4 19154744 19165958
60.1 6593407
61.0 6570398
65.1 1089N12R
72.4 6560201
Vv17
0.0 19164239
8.1 6569701
13.3 19144535
28.4 19146990
38.9 1087L11F
41.7 2006C10F
44.6 1071E22F
50.4 1030F07F
52.5 1094L14R
54.9 1044N12R 1099L01R
58.3 6574171
62.5 6579357
80.7 6570096
Vv18
0.0 VMC2a3
2.2 1073P15R
2.8 1038A12F
7.1 VMC3e5
15.5 1082L02F UGb8_Vvi2267
30.9 1095D06F
33.2 1075G16F
43.6 IN0185 UGb8_Vvi9920#1
UGb8_Vvi9920#2
46.1 1084H22R
50.4 IN0732
51.1 1098O11R 1075J12R
54.3 1092P03R
61.3 2004P09F
64.6 VVMD17
65.7 1078L19F
80.9 1044D14R

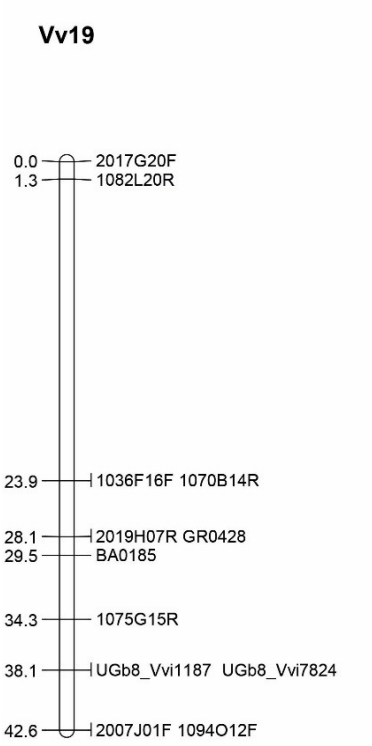
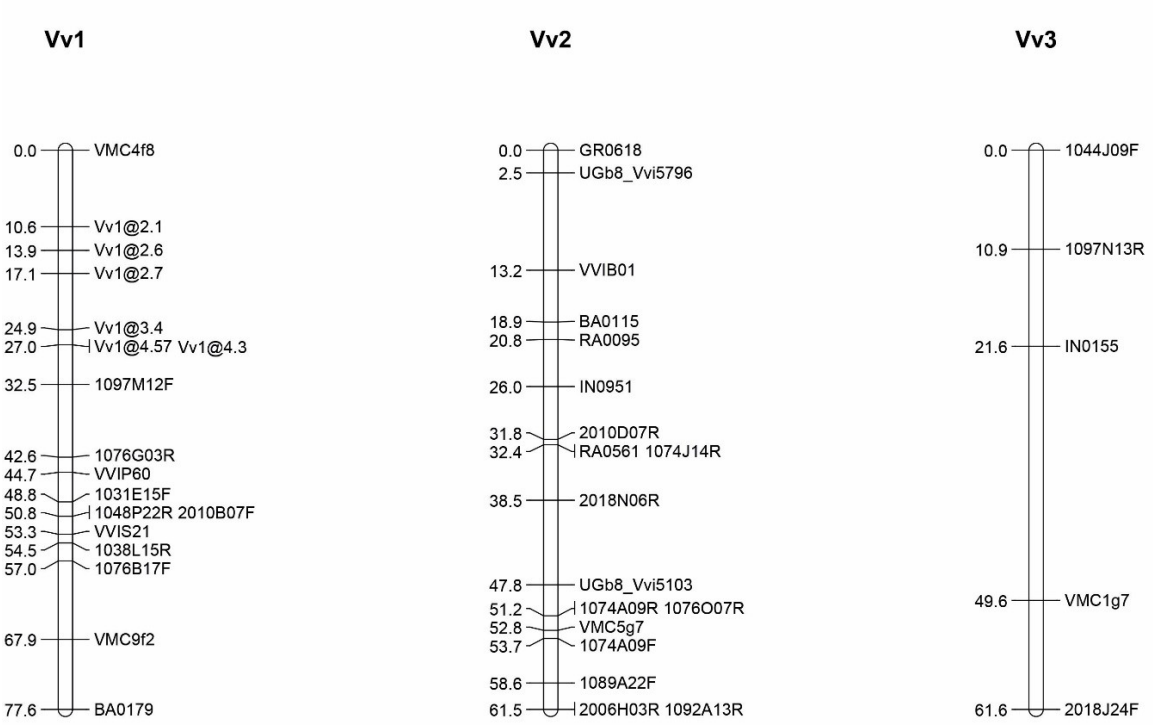
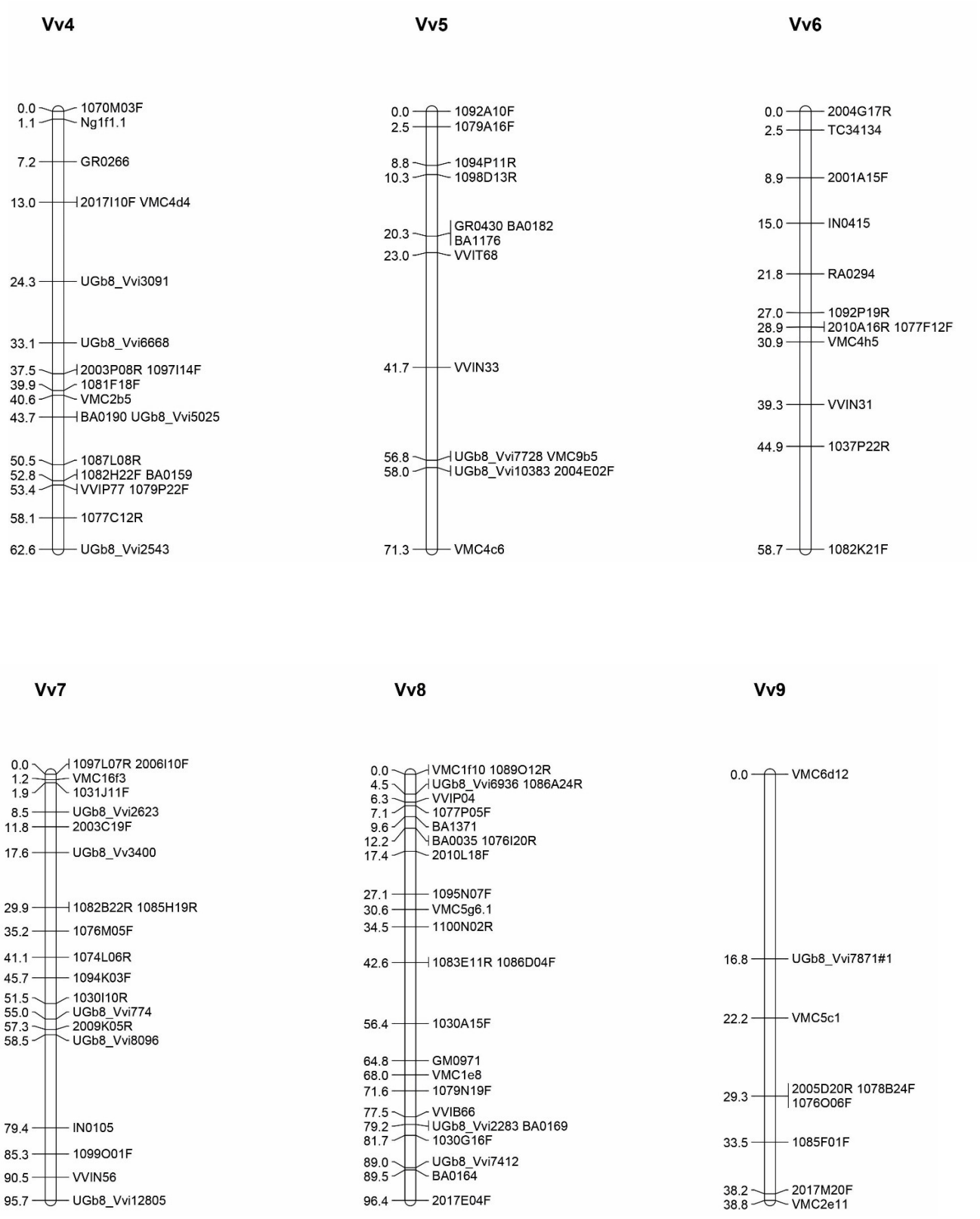
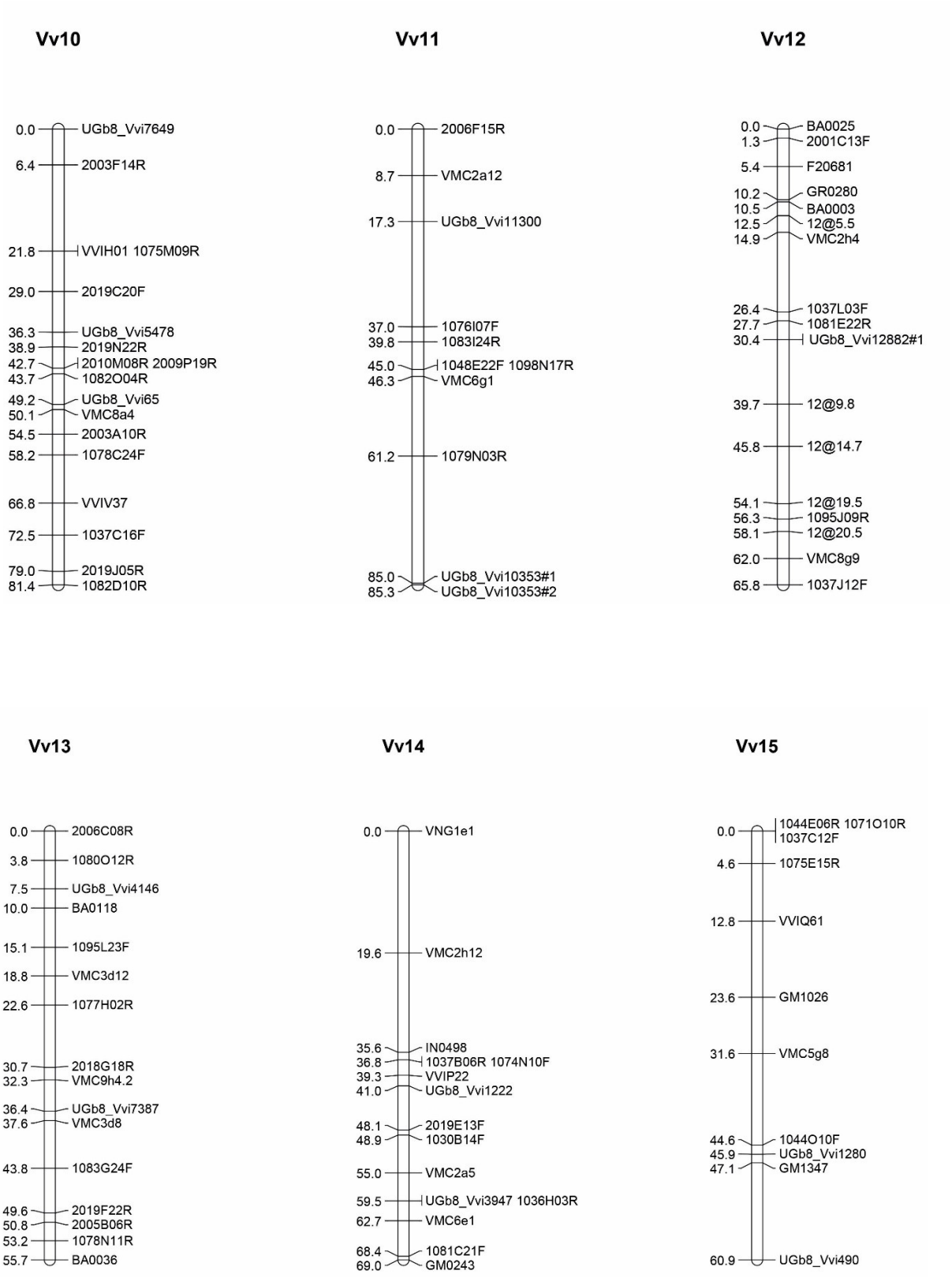


Figure 3 Linkage map of *Vitis vinifera* Microvine × Cabernet Sauvignon. Linkage groups are numbered as originally assigned in Adam-Blondon et al. (2004) and only the consensus maps are shown. Distances of markers from the top are indicated on the left in cM.




Vv4
0.0 1070M03F
1.1 Ng1f1.1
7.2 GR0266
13.0 2017I10F VMC4d4
24.3 UGb8_Vvi3091
33.1 UGb8_Vvi6668
37.5 2003P08R 1097I14F
39.9 1081F18F
40.6 VMC2b5
43.7 BA0190 UGb8_Vvi5025
50.5 1087L08R
52.8 1082H22F BA0159
53.4 VVIP77 1079P22F
58.1 1077C12R
62.6 UGb8_Vvi2543
Vv5
0.0 1092A10F
2.5 1079A16F
8.8 1094P11R
10.3 1098D13R
20.3 GR0430 BA0182
BA1176
23.0 VVIT68
41.7 VVIN33
56.8 UGb8_Vvi7728 VMC9b5
58.0 UGb8_Vvi10383 2004E02F
71.3 VMC4c6
Vv6
0.0 2004G17R
2.5 TC34134
8.9 2001A15F
15.0 IN0415
21.8 RA0294
27.0 1092P19R
28.9 2010A16R 1077F12F
30.9 VMC4h5
39.3 VVIN31
44.9 1037P22R
58.7 1082K21F
Vv7
0.0 1097L07R 2006I10F
1.2 VMC16f3
1.9 1031J11F
8.5 UGb8_Vvi2623
11.8 2003C19F
17.6 UGb8_Vv3400
29.9 1082B22R 1085H19R
35.2 1076M05F
41.1 1074L06R
45.7 1094K03F
51.5 1030I10R
55.0 UGb8_Vvi774
57.3 2009K05R
58.5 UGb8_Vvi8096
79.4 IN0105
85.3 1099O01F
90.5 VVIN56
95.7 UGb8_Vvi12805
Vv8
0.0 VMC1f10 1089O12R
4.5 UGb8_Vvi6936 1086A24R
6.3 VVIP04
7.1 1077P05F
9.6 BA1371
12.2 BA0035 1076I20R
17.4 2010L18F
27.1 1095N07F
30.6 VMC5g6.1
34.5 1100N02R
42.6 1083E11R 1086D04F
56.4 1030A15F
64.8 GM0971
68.0 VMC1e8
71.6 1079N19F
77.5 VVIB66
79.2 UGb8_Vvi2283 BA0169
81.7 1030G16F
89.0 UGb8_Vvi7412
89.5 BA0164
96.4 2017E04F
Vv9
0.0 VMC6d12
16.8 UGb8_Vvi7871#1
22.2 VMC5c1
29.3 2005D20R 1078B24F
1076O06F
33.5 1085F01F
38.2 2017M20F
38.8 VMC2e11


Vv10
0.0 UGb8_Vvi7649
6.4 2003F14R
21.8 VVIH01 1075M09R
29.0 2019C20F
36.3 UGb8_Vvi5478
38.9 2019N22R
42.7 2010M08R 2009P19R
43.7 1082O04R
49.2 UGb8_Vvi65
50.1 VMC8a4
54.5 2003A10R
58.2 1078C24F
66.8 VVIV37
72.5 1037C16F
79.0 2019J05R
81.4 1082D10R
Vv11
0.0 2006F15R
8.7 VMC2a12
17.3 UGb8_Vvi11300
37.0 1076I07F
39.8 1083I24R
45.0 1048E22F 1098N17R
46.3 VMC6g1
61.2 1079N03R
85.0 UGb8_Vvi10353#1
85.3 UGb8_Vvi10353#2
Vv12
0.0 BA0025
1.3 2001C13F
5.4 F20681
10.2 GR0280
10.5 BA0003
12.5 12@5.5
14.9 VMC2h4
26.4 1037L03F
27.7 1081E22R
30.4 UGb8_Vvi12882#1
39.7 12@9.8
45.8 12@14.7
54.1 12@19.5
56.3 1095J09R
58.1 12@20.5
62.0 VMC8g9
65.8 1037J12F
Vv13
0.0 2006C08R
3.8 1080O12R
7.5 UGb8_Vvi4146
10.0 BA0118
15.1 1095L23F
18.8 VMC3d12
22.6 1077H02R
30.7 2018G18R
32.3 VMC9h4.2
36.4 UGb8_Vvi7387
37.6 VMC3d8
43.8 1083G24F
49.6 2019F22R
50.8 2005B06R
53.2 1078N11R
55.7 BA0036
Vv14
0.0 VNG1e1
19.6 VMC2h12
35.6 IN0498
36.8 1037B06R 1074N10F
39.3 VVIP22
41.0 UGb8_Vvi1222
48.1 2019E13F
48.9 1030B14F
55.0 VMC2a5
59.5 UGb8_Vvi3947 1036H03R
62.7 VMC6e1
68.4 1081C21F
69.0 GM0243
Vv15
0.0 1044E06R 1071O10R
1037C12F
4.6 1075E15R
12.8 VVIQ61
23.6 GM1026
31.6 VMC5g8
44.6 1044O10F
45.9 UGb8_Vvi1280
47.1 GM1347
60.9 UGb8_Vvi490

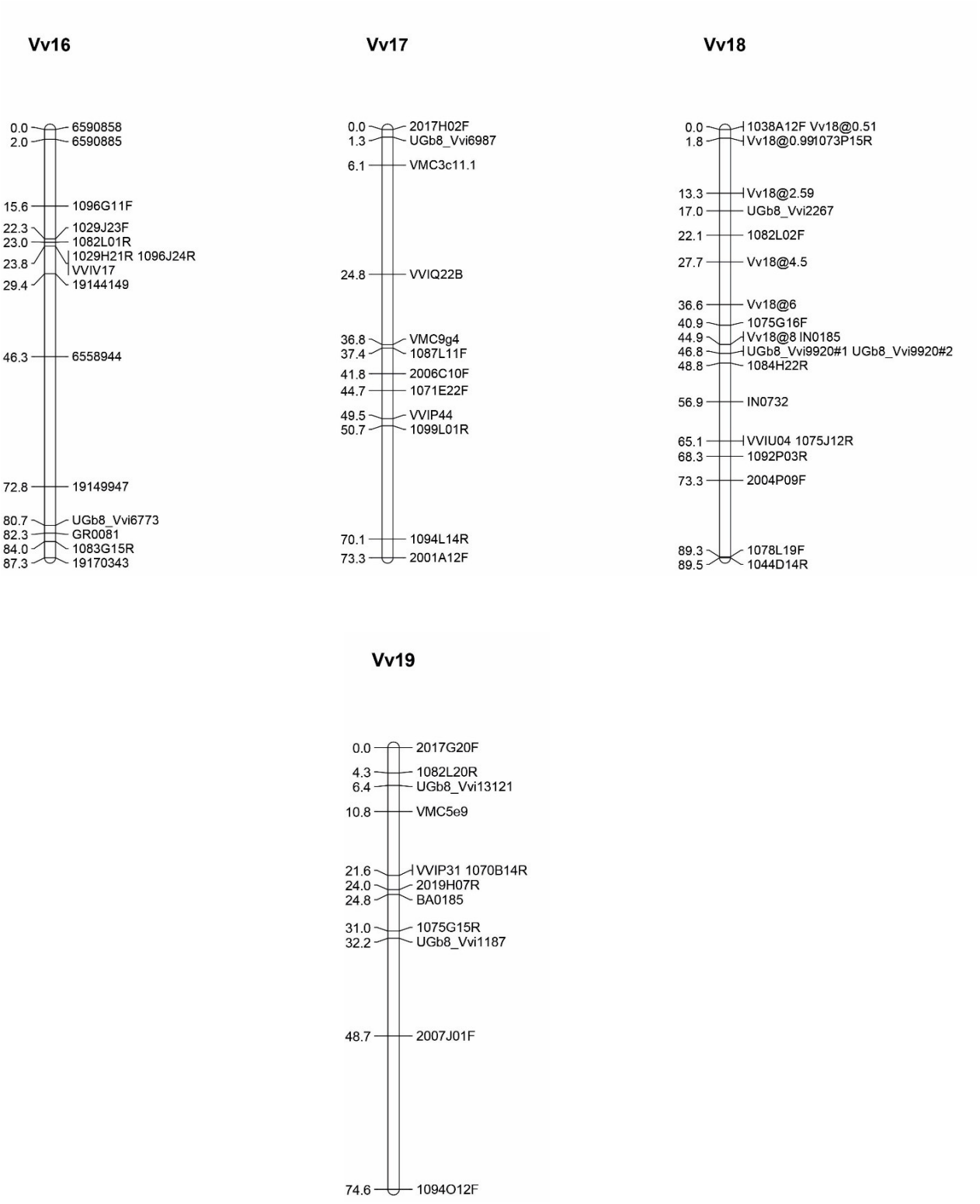


Figure 4 Linkage map of *Vitis vinifera* Microvine × Riesling. Linkage groups are numbered as originally assigned in Adam-Blondon et al. (2004) and only the consensus maps are shown. Distances of markers from the top are indicated on the left in cM.

6.1.3 QTL mapping of wine composition traits in the biparental crosses

Introduction

QTL mapping basically brings together phenotype and genotype to look for significant associations. Markers can then be made to the QTLs that best describe the variation in the trait so that it can be selected for or against in future breeding populations. In this project, the phenotypic data used were the abundance of volatile compounds in micro-scale wines produced from the F1 progeny of crosses between the microvine and either Cabernet Sauvignon or Riesling. The consensus maps described in 6.1.2 were used for the QTL analysis using MapQTL software. The initial compounds targeted were those thought to play a significant role in the sensory properties of wines.

(*Z*)-3-Hexenol

Introduction

One class of compounds thought to impart 'green' sensory characters on wins is the C_6 alcohols. These are derived from the degradation of linoleic and linolenic acid via the lipoxygenase pathway (Figure 5). It catalyses the release of C_6 volatile compounds from unsaturated fatty acids when most plant tissues are damaged and is responsible for the distinctive odour of freshly mown grass. As such, most of the C6 compounds are described as having a 'green' or 'cut grass' aroma, which is often seen as being undesirable in red wines.

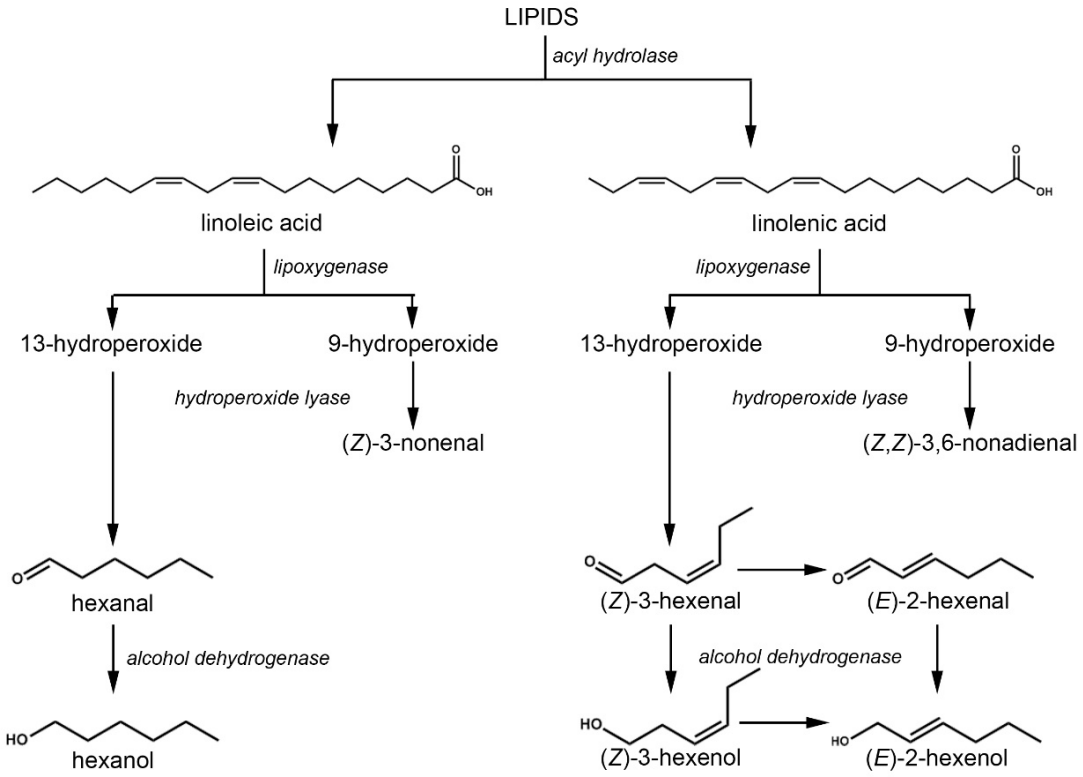


Figure 5 The lipoxygenase pathway

Commonly, all three of the alcohols at the end of this pathway, namely hexanol, (*Z*)-3-hexenol and (*E*)-2-hexanol are found in all wines. Another C_6 alcohol, (*E*)-3-hexenol, is also found in most wines, and presumably derives from non-enzymatic isomerisation of the unsaturated C_6 alcohols. They have distinct differences in their odour thresholds with the values calculated in water being: hexanol = 2500 µg/L; (*E*)-3-hexenol = 1550 µg/L, (*E*)-2-hexenol = 400 µg/L and (*Z*)-3-hexenol = 70 µg/L (Leffingwell & Associates; http://www.leffingwell.com). Therefore, (Z)-3-hexenol is the most potent of these compounds and finding any genetic variation that directs the lipoxygenase pathway down other production threads could reduce green characters in wines.

Data analysis

One of the first checks made to ensure that the accumulation of this compound is determined predominantly by genetic differences in the progeny and not by yeast or fermentation effects, was to plot the abundance measured in one biological and fermentation replicate versus the other. Our hypothesis was that if yeast are mainly driving the production of this compound, the replicates would not correlate with each other and we would expect the ranking of the progeny in terms of compound abundance to not be stable across replicates. A good example of this type of compound is 2,3-butanediol, formed from glucose by *Saccharomyces spp.* during fermentation. When the abundances of replicate 1 were plotted against replicate 2, there appeared to be no relationship between the two, suggesting that there was no grape genetic component contributing to this trait (Figure 6).

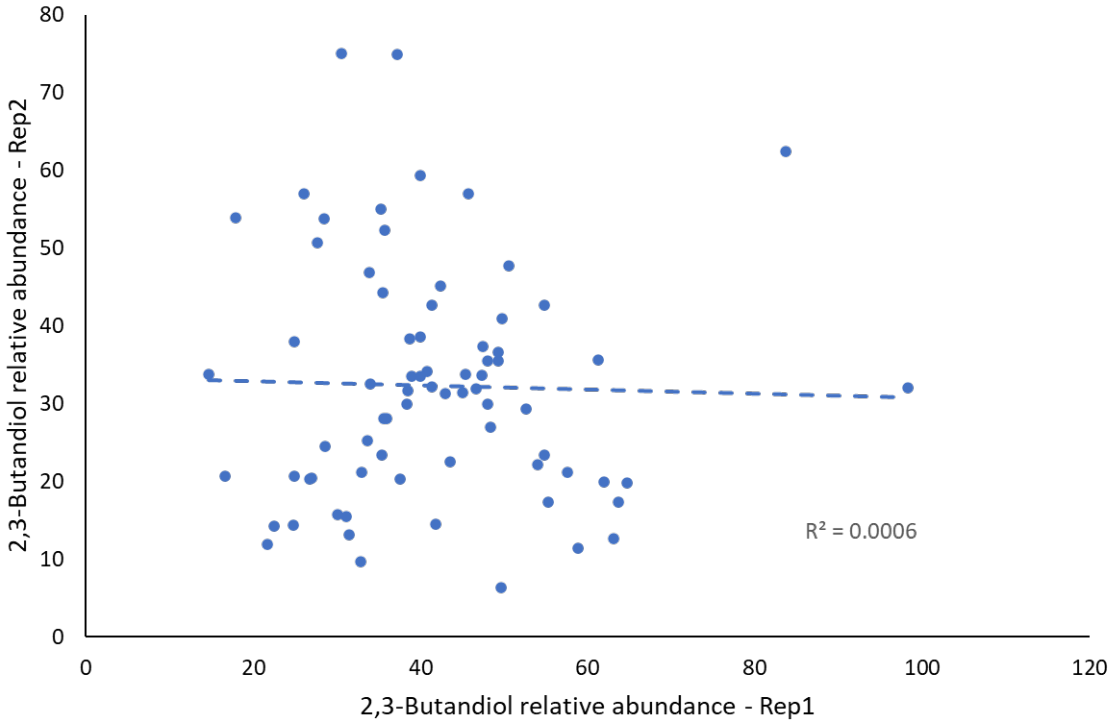


Figure 6 Plot of abundances of 2,3-butanediol in fermentation replicate 1 versus fermentation replicate 2.

Contrast this to the result when (*Z*)-3-hexenol abundances from wines produced in replicates 1 and 2 were plotted against each other for the two F1 populations (Figure 7). Clearly the replicates were

closely related to each other with very high R^2 values, suggesting that the abundances were likely to be controlled predominantly by the composition of the grape material. Another interesting thing to note about the (*Z*)-3-hexenol abundances was the range of values, with the Mv × R progeny having up to a ~18-fold difference between the highest and lowest progeny, whereas the Mv × CS progeny varied by up to ~60-fold when comparing the (Z)-3-hexenol abundances.

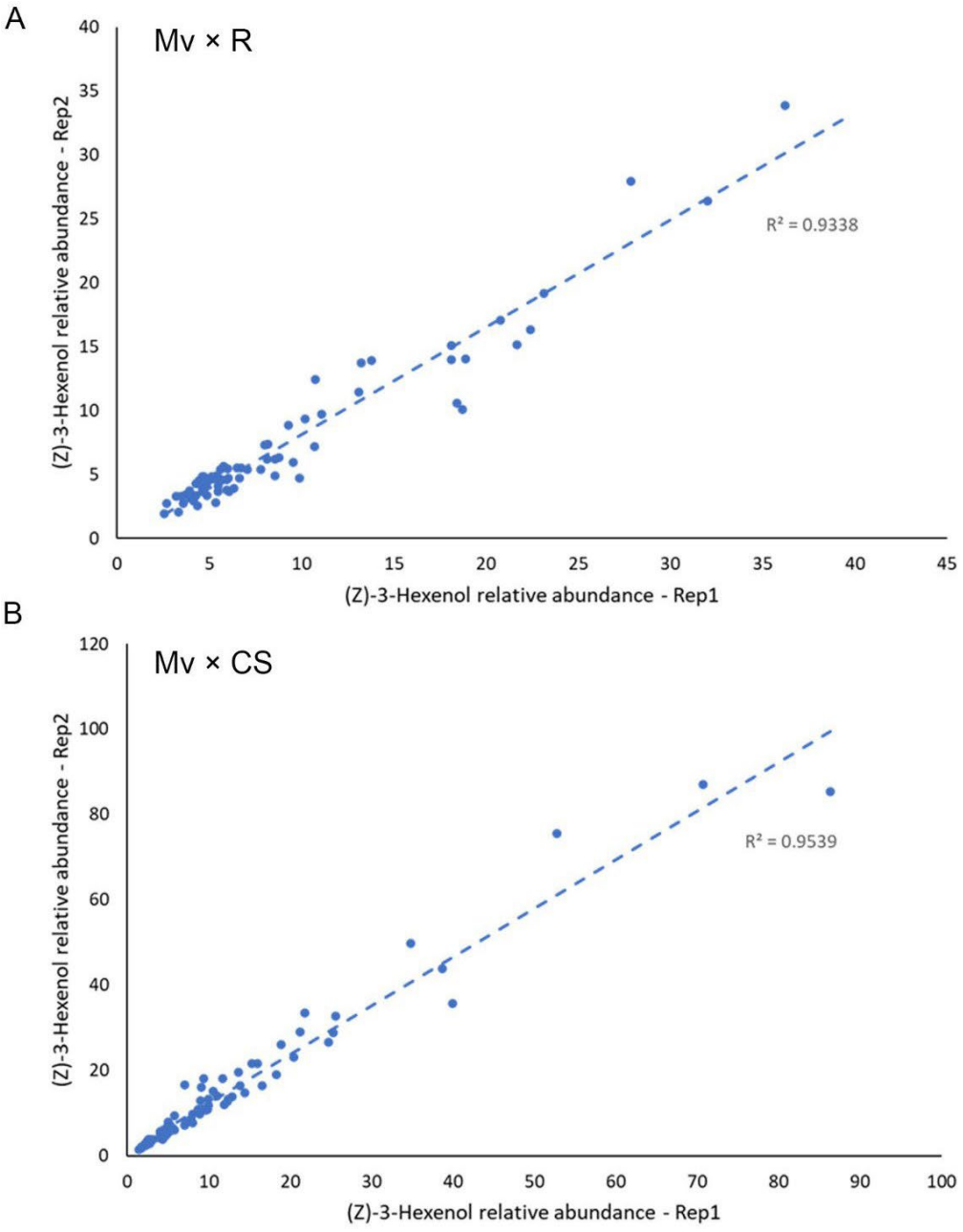


Figure 7 (Z)-3-Hexenol abundances in both fermentation reps for the Mv × CS and Mv × R populations plotted against each other.

QTL analysis

Using the maps produced in 6.1.2 and the data described above, QTL mapping was conducted using MapQTL software as outlined in 5.4.1. The first analysis conducted was a non-parametric mapping method using the Kruskal-Wallis rank sum test. This method looks for a significant association between individual markers at the phenotype being tested and is a preliminary test to look for linkage groups with potential QTLs before conducting multiple-QTL models (MQM) mapping. Analysis of the Mv × R cross revealed a number of markers with significant associations with the phenotype at the $p<0.0001$ level in both replicates (Table 4). These were located predominantly on chromosome 12, towards the bottom of the linkage group, and were flanked by markers with significant associations, but at lower levels. One marker on chromosome 7 was also significant, but there were no highly significant flanking markers.

Table 4 Linkage group, position, loci and significance values for the top ten markers from a Kruskal-Wallace analysis of the (*Z*)-3-hexenol wine phenotype in the Mv × R progeny.

GROUP	POSITION (cM)	MARKER	SIGNIFICANCE	
			REP1	REP2
Vv12	52.2	1099J02R	$p<0.0001$	$p<0.0001$
Vv12	53.2	VMC4f3.1	$p<0.0001$	$p<0.0001$
Vv12	56.1	1095J09R	$p<0.0001$	$p<0.0001$
Vv12	65.0	VMC8g9	$p<0.0001$	$p<0.0001$
Vv7	16.4	IN0105	$p<0.0001$	$p<0.005$
Vv12	27.8	1081E22R	$p<0.0005$	$p<0.001$
Vv12	30.5	UGb8_Vvi12882#1	$p<0.0005$	$p<0.0005$
Vv12	44.5	12@14.8	$p<0.0005$	$p<0.0001$
Vv12	59.9	12@19.0	$p<0.0005$	$p<0.005$
Vv12	26.5	1037L03F	$p<0.001$	$p<0.005$

The association between the markers and the phenotype can be visualised by box plots and this is illustrated in Figure 8 Box plots for two highly significant markers in chromosome 12 showing the relationship between phenotype and genotype in replicate 1 in the Mv × R cross. The 1099J02R marker segregated into two genotypes in the progeny and when the plant had the 'lm' combination of alleles from the parents more (*Z*)-3-hexenol was present in the micro-scale wines compared to when the 'll' combination of alleles was inherited. The VMC4f3.1 marker segregated in the progeny following a pattern that suggested there were three alleles in the parents that were both heterozygous (i.e., 'ef × eg' segregation). The data shows that the 'fg' combination of alleles in this cross resulted in a higher abundance of (*Z*)-3-hexenol in the wines. This suggests that two recessive alleles, one from each parent, were coming together in some of the progeny to raise (*Z*)-3-hexenol concentrations.

In order to confirm the location of this QTL and to test if it is specific for this cross, the Mv × CS cross was also analysed using a non-parametric Kruskal-Wallace test to look for significant marker-trait associations. Only seven markers had significance levels better than $p < 0.01$ in at least one

replicate (Table 5). These markers were all located on chromosome 12 with four of the seven markers also being significant markers found in the Mv × R cross (Table 4).

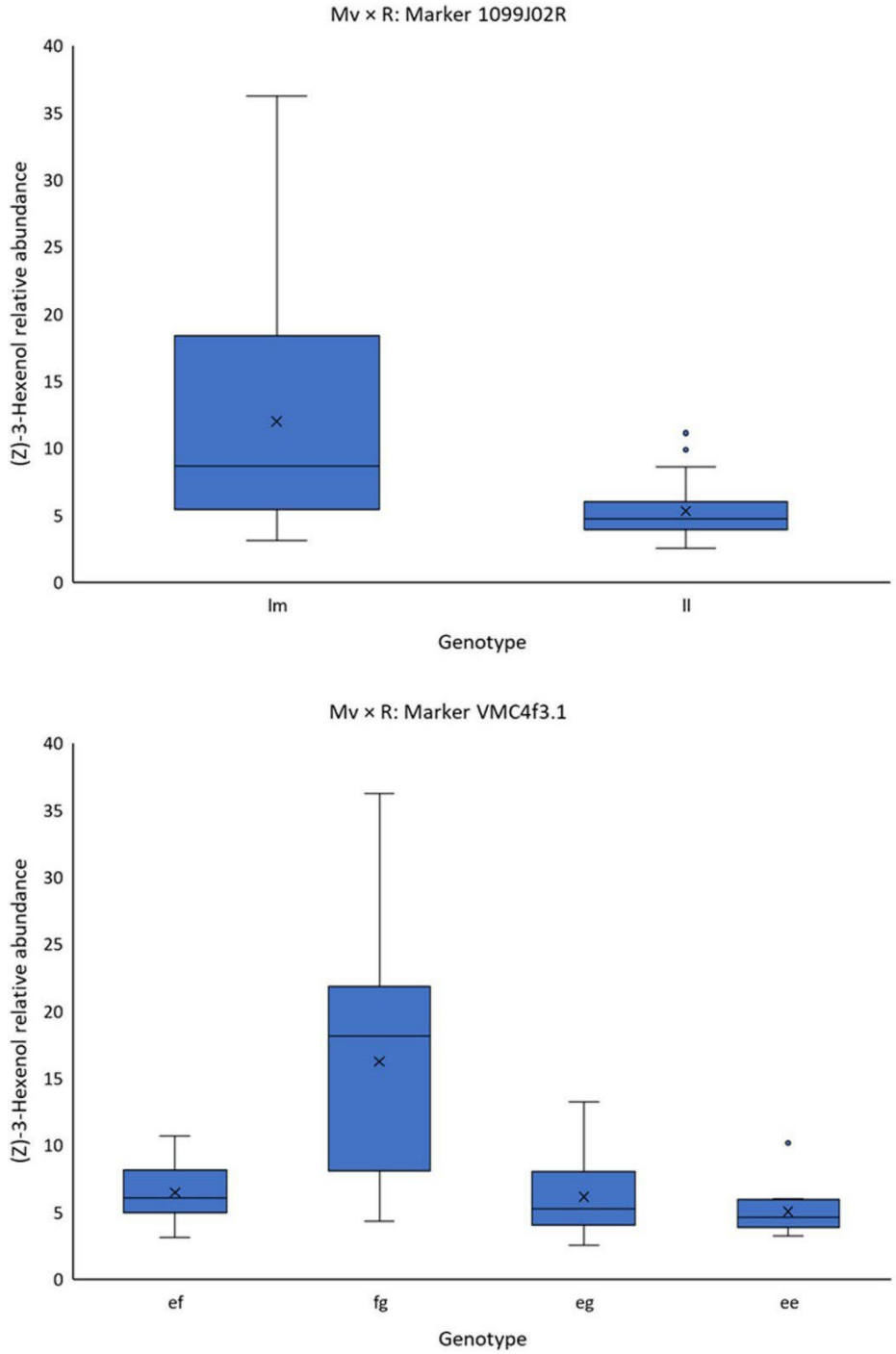


Figure 8 Box plots for two highly significant markers in chromosome 12 showing the relationship between phenotype and genotype in replicate 1 in the Mv × R cross.

Box plots of the markers again allowed the visualisation of the relationship between the markers and the concentrations of (Z)-3-hexenol in the wines (Figure 9). The 1099J02R marker was heterozygous in the microvine and homozygous in Cabernet Sauvignon, and the 'm' allele from the

microvine was again associated with higher concentrations of (Z)-3-hexenol in the wines. Most of the other significant markers were of the 'lm × ll' type and showed similar associations as 1099J02R.

Table 5 Linkage group, position, loci and significance values for the top seven markers from a Kruskal-Wallace analysis of the (Z)-3-hexenol wine phenotype in the Mv × CS progeny.

GROUP	POSITION (cM)	MARKER	SIGNIFICANCE	
			REP1	REP2
Vv12	35.8	1084H20R	$p<0.0001$	$p<0.0001$
Vv12	35.8	1099J02R	$p<0.0001$	$p<0.0001$
Vv12	43.9	1037J12F	$p<0.0001$	$p<0.0005$
Vv12	38.0	1095J09R	$p<0.0005$	$p<0.0005$
Vv12	49.1	VMC8g9	$p<0.001$	$p<0.005$
Vv12	9.3	GR0280	$p<0.01$	$p<0.05$
Vv12	18.2	1081E22R	$p<0.01$	$p<0.05$

Therefore, the genotype/phenotype association of the VMC8g9 marker was plotted instead to explore any contribution from the Cabernet Sauvignon genome (Figure 9). The results suggest that the Cabernet Sauvignon alleles may not vary in their ability to influence the concentration of (Z)-3-hexenol in the wines. Segregation of the VMC8g9 marker suggests that both parents are heterozygous with there being four different alleles, hence an 'ab × cd' marker type. The progeny that inherit the 'a' allele from the microvine tend to have higher (Z)-3-hexenol concentrations in the wine compared to those that inherit 'b' (Figure 9). Conversely, the inheritance of 'c' or 'd' from Cabernet Sauvignon does not seem to greatly affect (Z)-3-hexenol concentrations.

Following the Kruskal-Wallace analysis, MQM mapping was conducted to de-lineate the genome region underlying any significant QTLs related to wine (Z)-3-hexenol concentration. After optimisation of the models by testing the effect of using combinations of markers as co-factors, the MQM mapping in both crosses identified a single significant QTL on chromosome 12 that aligned with the results from the Kruskal-Wallace analysis.

Looking at the Mv × R cross, there was a peak present near the bottom of chromosome 12 in both replicates (Figure 10). In replicate one, the highest LOD score (12.7) occurred at the 1095J09R marker located at ~56.3 cM from the top of the chromosome, and the LOD at this position was well over the $p < 0.05$ LOD threshold of 4.7. In replicate two, the highest LOD score was slightly less than in replicate one, being 11.6 (LOD threshold $p < 0.05 = 5.1$) and was located at the '12@17.9' marker, which was 54.1 cM from the top of chromosome 12. The physical position of the peak on chromosome 12 corresponded to an interval between 17.6-19.0 Mbp from the top of the chromosome when using the 12x v2 build of the Pinot noir PN40024 genome. Both markers were predicted to explain about 52% of the total variation in the (*Z*)-3-hexenol concentration in the wines.

With regards to the Mv × CS cross, the peak was found in an overlapping part of chromosome 12 (Figure 11). In both replicates, the nearest markers to the peak LOD were 1099J02R and 1084H20R, which are both located 32.3 cM from the top of chromosome 12 in this map. Highest LOD scores were 11 and 7.9 for replicate 1 and 2 respectively and the $p < 0.05$ LOD thresholds were 5.0 and 5.2 for each replicate. The explained variance for these QTLs was less than that for the Mv × R cross, being 37% and 29% for replicate 1 and 2 of the Mv × CS cross respectively. It should be remembered

that the cM position of markers will change depending on the population and relative crossover frequency, so although these are different between the two crosses, the physical position of the QTL in the Mv × CS cross falls between ~17.4 and 18.3 Mbp from the top of chromosome 12 using the Pinot noir PN40024 12x v2 build.

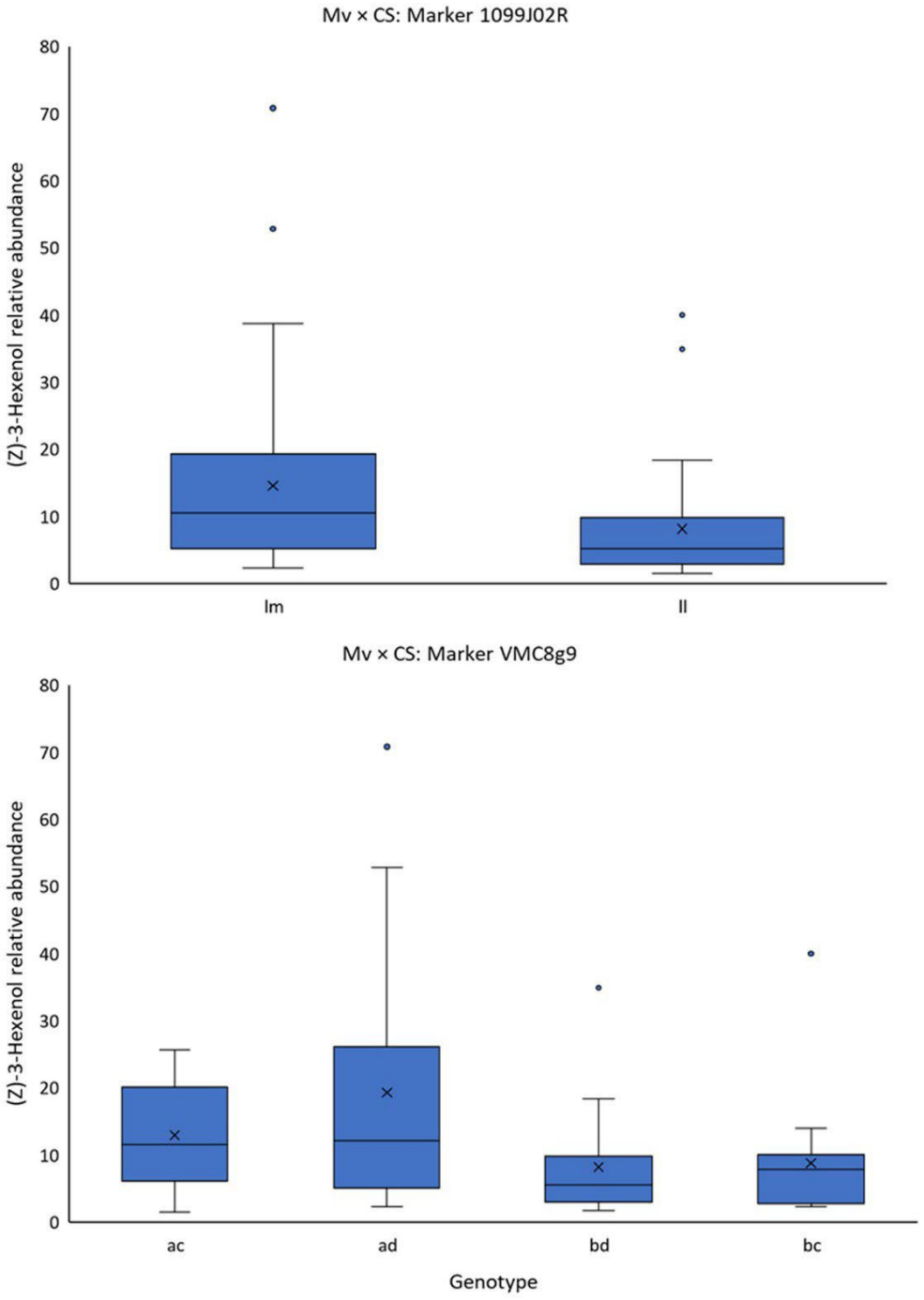


Figure 9 Box plots for two significant markers in chromosome 12 showing the relationship between phenotype and genotype in replicate 1 of the Mv × CS cross.

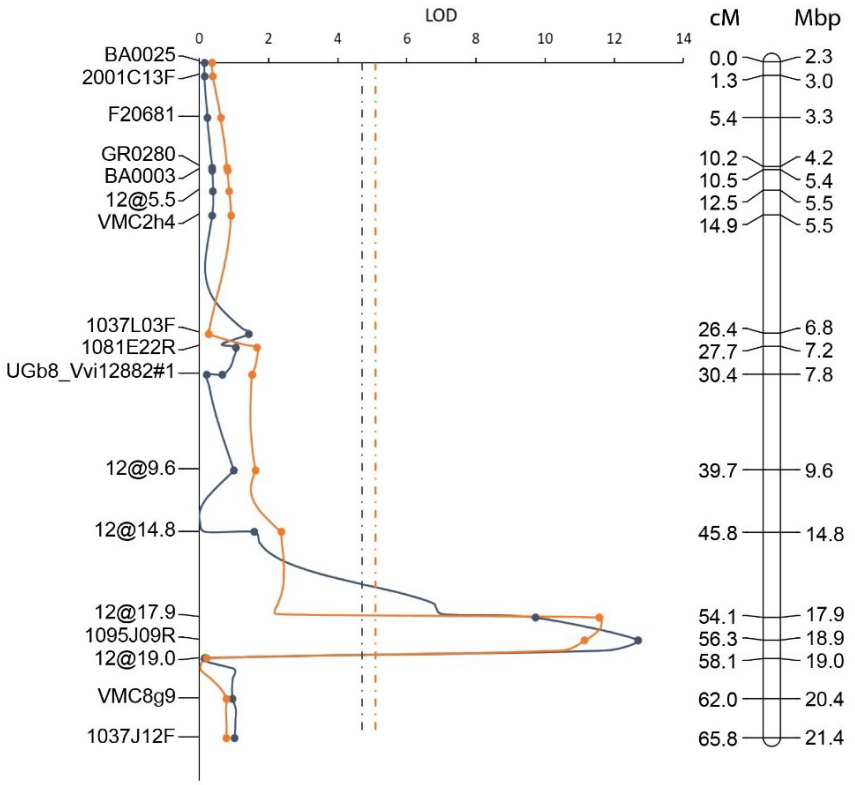


Figure 10 Schematic of the chromosome 12, wine (Z)-3-hexenol QTL mapping interval in Mv × R. Shows logarithm of odds (LOD) score against marker on the left and map units (in cM) and physical position (in Mbp) on the right. The vertical dashed lines represent the genome-wide significance threshold ($p < 0.05$). Data from replicate 1 is plotted in grey and data from replicate 2 in orange.

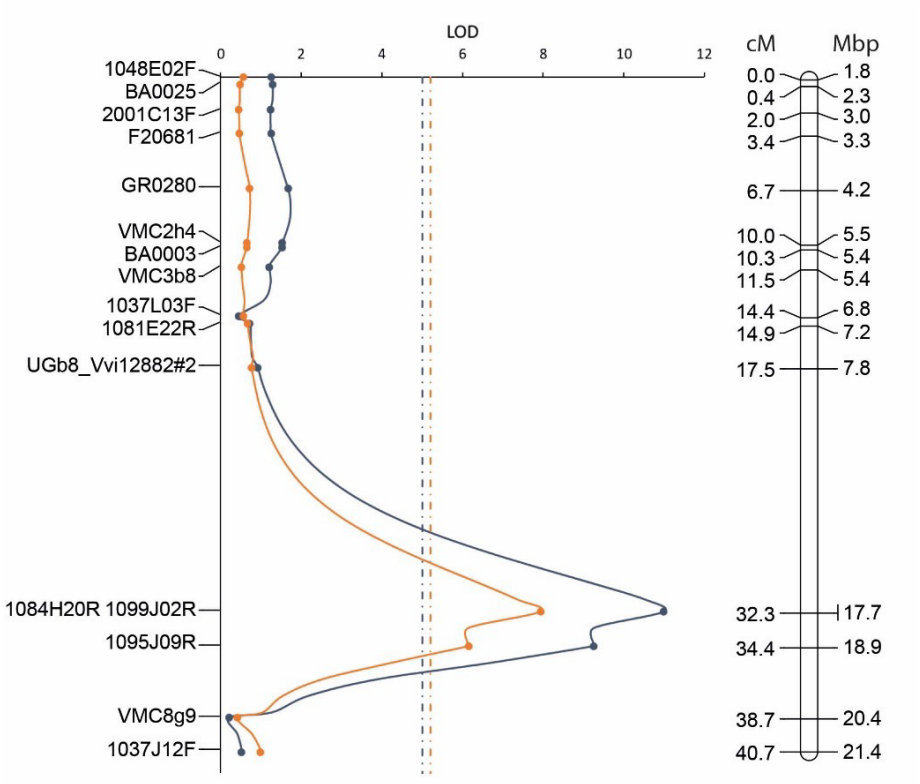


Figure 11 Schematic of the chromosome 12, wine (Z)-3-hexenol QTL mapping interval in Mv × CS. Shows logarithm of odds (LOD) score against marker on the left and map units (in cM) and physical position (in Mbp) on the right. The

vertical dashed lines represent the genome-wide significance threshold ($p < 0.05$). Data from replicate 1 are plotted in grey and data from replicate 2 in orange.

In conclusion, the QTL mapping of the wine (Z)-3-hexenol concentration trait in the two biparental crosses identified a single locus on chromosome 12 describing 30-60 % of the total variation in this phenotype and covering a region from ~17.4 to 19.0 Mbp. This region of the genome was interrogated further for the presence of candidate genes and the development of markers and this work is described in other results sections of this report.

Acetate esters

Introduction

While the C_6 alcohols are associated with undesirable 'green' sensory characters in wine, esters are assumed to contribute to 'fruit' attributes and therefore have a more positive effect on wine quality (Sumby, et al. 2010). There are a large number of different esters present in wine, although the majority can be grouped into two main subsets, the ethyl esters and acetate esters. Ester production is catalysed by enzymes in yeast during fermentation. Ethyl esters result from the condensation of ethanol and an acyl-CoA; acetate esters are produced when various alcohols are condensed with acetyl-CoA (Figure 11).

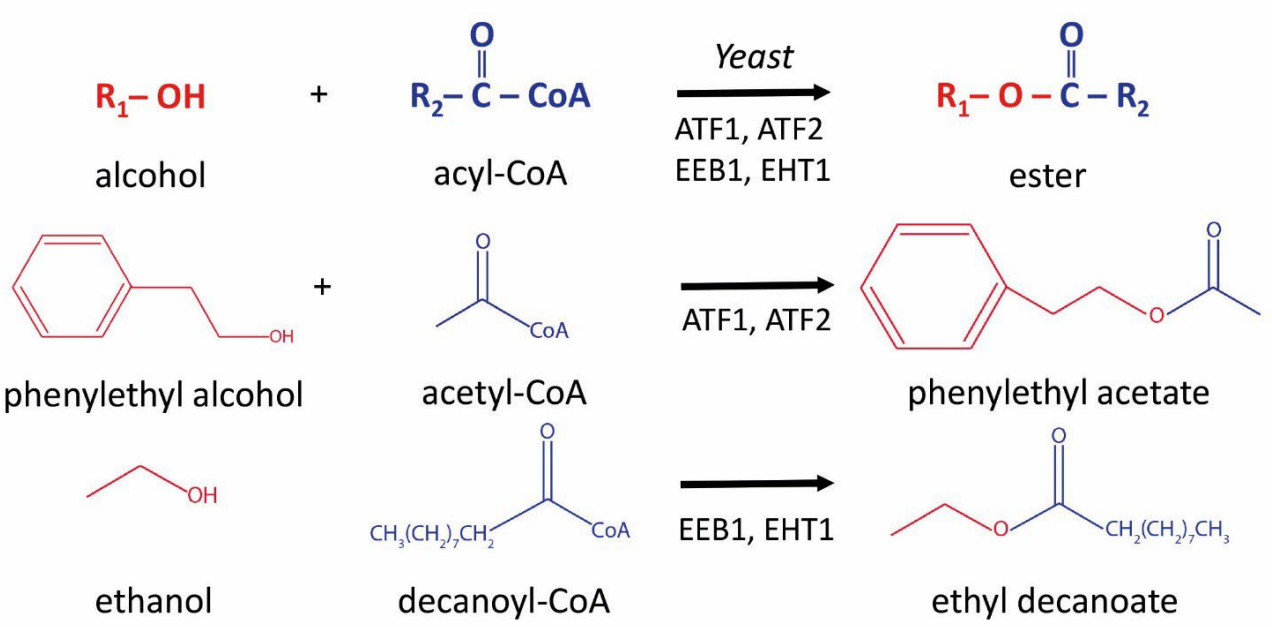


Figure 12 Chemical equations showing generic ester formation at the top with an example of the production of an acetate ester and ethyl ester beneath.

The ethyl esters commonly have fruit-based descriptors, while some of the acetate esters can have floral and herbaceous attributes, such as phenylethyl acetate ('rose') and hexyl acetate ('herbaceous') (Sumby, et al. 2010). Nevertheless, increased phenylethyl acetate has been linked to higher 'fruity' characters in wine (Viana, et al. 2009) and acetate esters could modify the amount and type of fruit aroma perceived in an ester mixture even if present in concentrations below their threshold (Lytra, et al. 2013). Despite the yeast-based production of these compounds during

fermentation, we have previously shown that this is dependent on the composition of the fruit, either via the presence of precursors or other factors that contribute to the activity of the yeast enzymes during fermentation (Boss, et al. 2015, Dennis, et al. 2012, Keyzers and Boss 2010).

Data analysis

One of the first checks of the data analysis, as mentioned above, is to plot replicates against each other. This gives us an idea as to whether the trait is likely influenced by grape composition or not – with those compounds showing good correlations being hypothesised as having a strong influence of grape composition on wine concentration. The three most abundant acetate esters in the wines were phenylethyl acetate, hexyl acetate and isoamyl acetate. Plotting the relative abundances of one replicate against the other showed that phenylethyl acetate showed a good correlation, suggesting that fruit composition is playing a significant role in the trait (Figure 13). Hexyl acetate also showed a good correlation between the two replicates in the Mv × R progeny but showed more variation in the Mv × CS progeny (Figure 13). Isoamyl acetate showed the least correlation between the replicates from both crosses (Figure 13).

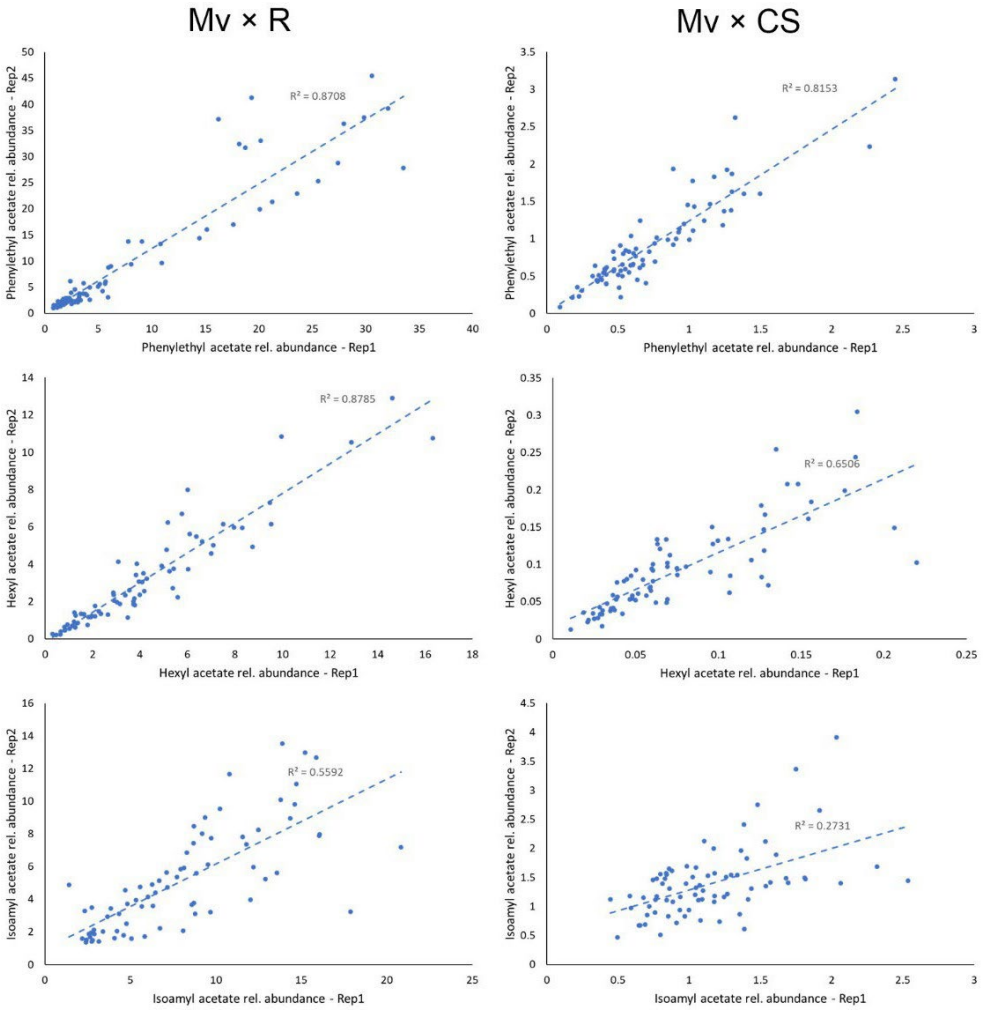


Figure 13 Relative abundances of three acetate esters in both fermentation replicates for the Mv × CS and Mv × R populations plotted against each other.

When looking at the spread of the data, the difference between the lowest and highest relative abundances for phenylethyl acetate represented a 46-fold difference in the Mv × R cross and 37-fold in the Mv × CS cross. Hexyl acetate was more variable in the Mv × R cross with a 64-fold difference between the wine with the highest and lowest relative abundance than the Mv × CS cross which had a 23-fold difference. Finally, isoamyl acetate was less variable with 15- and eight-fold changes between highest and lowest relative abundances in the Mv × R and MV × CS progeny respectively.

QTL analysis

It was a surprise that the wine acetate ester concentrations looked like they were highly influenced by fruit composition given that they are predominantly produced by yeast during fermentation. Nevertheless, the data analysis suggested that they may be targets for QTL mapping. Because of the closer relationship between the replicates in both of the biparental crosses compared to the R^2 values for the comparisons of the replicates for the other acetate esters, phenylethyl acetate abundance was chosen as the initial trait to be mapped.

The initial Kruskal-Wallace non-parametric test supported the hypothesis that the fruit was influencing the production of phenylethyl acetate in the wines. A number of markers in the Mv × R cross had significant associations with the trait at the $p<0.0001$ level in both replicates (Table 6). These were all located at the top of chromosome 18, although some of the markers lower on the chromosome were only significant at the $p<0.0005$ level in the second replicate.

Table 6 Linkage group, position, loci and significance values for the top eight markers from a Kruskal-Wallace analysis of the phenylethyl acetate wine abundance phenotype in the Mv × R progeny.

GROUP	POSITION (cM)	MARKER	SIGNIFICANCE	
			REP1	REP2
Vv18	0.0	1038A12F	$p<0.0001$	$p<0.0001$
Vv18	0.0	Vv18@0.51	$p<0.0001$	$p<0.0001$
Vv18	1.9	1073P15R	$p<0.0001$	$p<0.0001$
Vv18	13.4	UGb8_Vvi2267	$p<0.0001$	$p<0.0001$
Vv18	18.3	1082L02F	$p<0.0001$	$p<0.0001$
Vv18	24.6	Vv18@4.5	$p<0.0001$	$p<0.0001$
Vv18	33.5	Vv18@6	$p<0.0001$	$p<0.0005$
Vv18	38.0	1075G16F	$p<0.0001$	$p<0.0005$

Plotting the phenylethyl acetate abundance for each 1038A12F genotype showed a distinct effect of certain alleles upon this trait (Figure 14). The marker was heterozygous in both parents and can be denoted as showing an 'hk × hk' type segregation. The combination of 'kk' alleles results in higher concentrations of phenylethyl acetate in the wines made from those progeny, whereas plants carrying the 'hh' alleles had lower phenylethyl acetate concentrations (Figure 14). The heterozygous 'hk' progeny mainly have low phenylethyl acetate concentrations, although some outliers are apparent at the top of the box plot (Figure 14). It is possible that the origin of the 'k' allele is having an effect on the trait. Most of the other markers in this region also segregated 'hk × hk', except

1082L02F, which segregated as an 'lm × ll' type marker. In this case the presence of the 'm' allele resulted in lower wine phenylethyl acetate concentrations in the progeny (Figure 14).

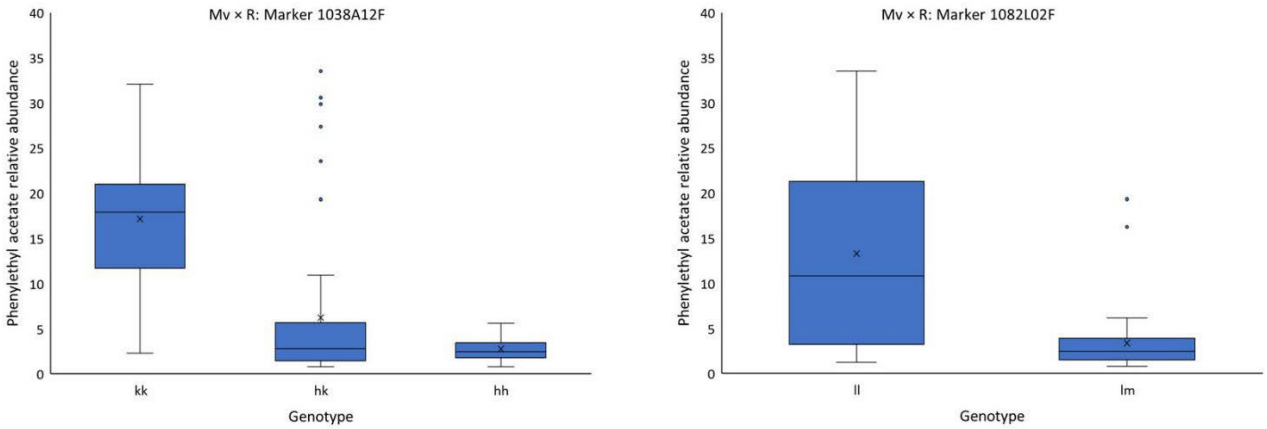


Figure 14 Box plots for two significant markers in chromosome 18 showing the relationship between phenotype and genotype in replicate 1 of the Mv × R cross.

This trait was also mapped in the Mv × CS population to test if the effect of grape genotype on phenylethyl acetate production was also present. Kruskal-Wallace analysis of the trait:marker associations showed a similar pattern to that seen in the Mv × R cross. Several markers at the top of chromosome 18 were significant at the $p<0.001$ level (Table 7), which is not as strong as the association seen in the Mv × R population. A marker on chromosome 8 and another on chromosome 11 were also significant at $p<0.01$ but did not have significant markers surrounding their position on their respective chromosomes.

Table 7 Linkage group, position, loci and significance values for the top seven markers from a Kruskal-Wallace analysis of the phenylethyl acetate wine abundance phenotype in the Mv × CS progeny.

GROUP	POSITION (cM)	MARKER	SIGNIFICANCE	
			REP1	REP2
Vv18	12.5	UGb8_Vvi2267	$p<0.0001$	$p<0.0001$
Vv18	3.4	1073P15R	$p<0.0005$	$p<0.0001$
Vv18	0.0	VMC3e5	$p<0.001$	$p<0.001$
Vv8	76.4	VMC2h10	$p<0.005$	$p<0.01$
Vv11	18.9	UGb8_Vvi11300	$p<0.01$	ns
Vv18	12.5	1082L02F	$p<0.01$	$p<0.005$
Vv18	30.2	1074G16F	$p<0.05$	$p<0.01$

Plotting two of these markers to show the relationship between the genotype and phenotype suggests that the production of high concentrations of phenylethyl acetate is a recessive trait in this population. This is best shown in the UGb8_Vvi2267 marker, where the combination of 'kk' alleles resulted in higher phenylethyl acetate in the wine compared to the concentrations observed in

those progeny with 'hk' and 'hh' allelic combinations (Figure 15). That is, the presence of one dominant 'h' allele resulted in lower acetate ester production during fermentation. Marker VMC3e5 segregated 'nn × np' and the presence of the 'p' allele from Cabernet Sauvignon resulted in lower concentrations of phenylethyl acetate than when the 'n' allele was present in the progeny (Figure 15).

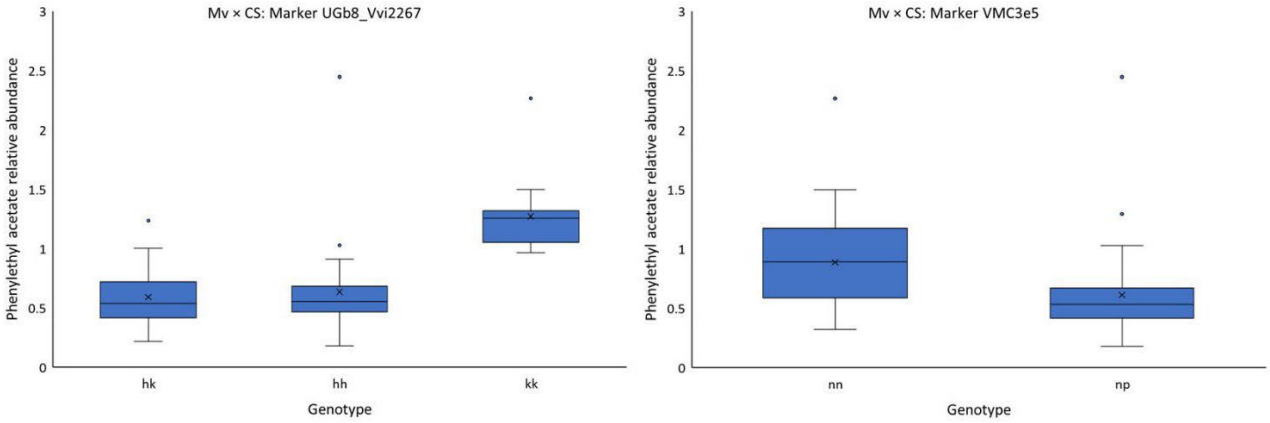


Figure 15 Box plots for two significant markers in chromosome 18 showing the relationship between phenotype and genotype in replicate 1 of the Mv × CS cross.

QTL mapping using the MQM methodology was conducted and combinations of markers as co-factors tested until an optimum model was reached. In the Mv × R cross, a single significant peak was present at the top of chromosome 18 in both replicates (Figure 16). The highest LOD score in replicate 1 was 16.76 and this was located just down from the 1082L02F marker, which is found 18.3 cM from the top of the chromosome. This LOD score was much greater than the threshold LOD for $p < 0.05$ which was 4.7 for this QTL analysis. The LOD threshold for replicate 2 was 4.4, and this was exceeded again by markers in the same QTL region identified in replicate 1 (Figure 16). The peak LOD (12.3) was observed just downstream of 1082L02F again. The physical position of this peak on chromosome 18 between the markers Vv18@2.5.9 and Vv18@4.5 corresponds to 2.6-4.5 Mbp from the top of the chromosome using the 12x v2 build of the Pinot noir PN40024 genome. The peak in replicate 1 described 62% of the total variation in the phenotype and 54% of the variation in replicate 2.

In the Mv × CS cross, a single QTL was also identified on chromosome 18 (Figure 17), overlapping the one discovered in the Mv × R cross. The nearest markers to the highest LOD were 1082L02F and UGb8_Vvi2267, which both mapped 12.5 cM from the top of the chromosome in this cross. In replicate 1, the peak LOD was 9.3 at these markers, much higher than the $p < 0.05$ LOD threshold of 4.5, and in replicate 2, the LOD peak was higher, being 10.2 just upstream from the markers at 12.5 cM with the LOD threshold being 4.6. The explained variance at the peak of the QTL were 43% and 46% for replicate 1 and 2 respectively. The physical position of this peak on chromosome 18, taken as between markers 1038A12F and 1095D06F, is from 0.4 Mbp and 6.1 Mbp from the top of chromosome 18 using the 12x v2 build of the Pinot noir PN40024 genome. So, the overlapping

region of the QTL is that mapped in the Mv × R cross, from 2.6-4.5 Mbp. Candidate genes and marker development is discussed further in other sections of this report.

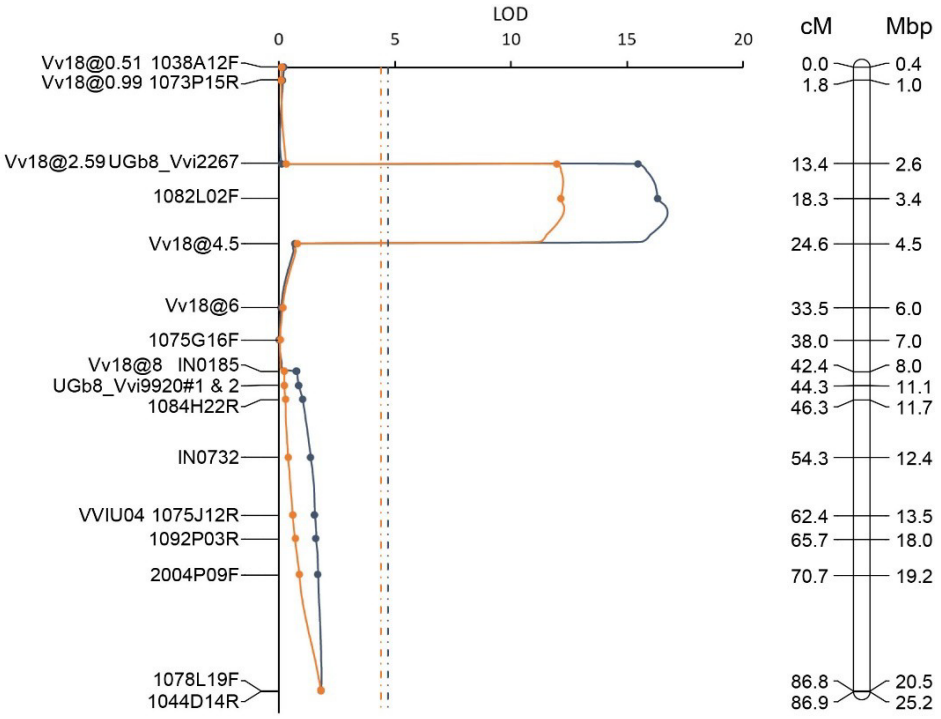


Figure 16 Schematic of the chromosome 18, wine phenylethyl acetate QTL mapping interval in Mv × R. Shows logarithm of odds (LOD) score against marker on the left and map units (in cM) and physical position (in Mbp) on the right. The vertical dashed lines represent the genome-wide significance threshold ($p < 0.05$). Data from replicate 1 is plotted in grey and data from replicate 2 in orange.

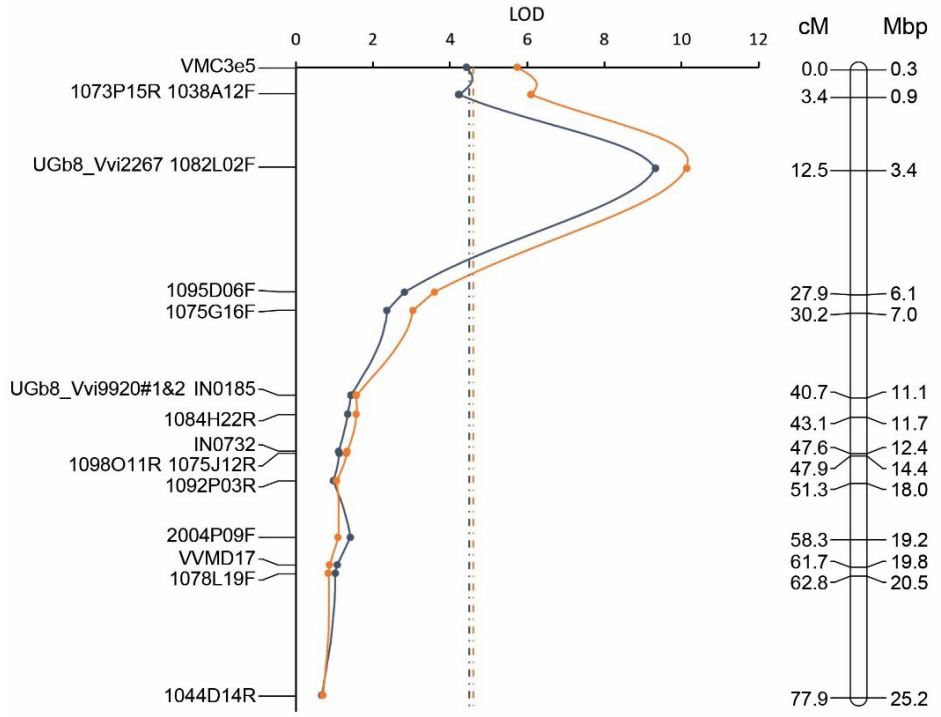


Figure 17 Schematic of the chromosome 18, wine phenylethyl acetate QTL mapping interval in Mv × CS. Shows logarithm of odds (LOD) score against marker on the left and map units (in cM) and physical position (in Mbp) on the

right. The vertical dashed lines represent the genome-wide significance threshold ($p < 0.05$). Data from replicate 1 is plotted in grey and data from replicate 2 in orange.

After the success of the phenylethyl acetate QTL mapping, the process was repeated with hexyl acetate as the correlation between the replicates was relatively good (Figure 13) suggesting a grape compositional influence on the abundance of this compound in wine. The Kruskal-Wallace non-parametric analysis of the marker:phenotype relationship in the Mv × R cross suggested that the same area of the genome that was mapped for phenylethyl acetate was also responsible for hexyl acetate abundance. Highly significant associations between the concentration of hexyl acetate in the wines and markers were identified at the top of chromosome 18 using both biological replicates (Table 8). A marker on chromosome 8 was also associated with the phenotype at the p<0.005 level.

Table 8 Linkage group, position, loci and significance values for the top ten markers from a Kruskal-Wallace analysis of the hexyl acetate wine abundance phenotype in the Mv × R progeny.

GROUP	POSITION (cM)	MARKER	SIGNIFICANCE	
			REP1	REP2
Vv18	18.3	1082L02F	p<0.0001	p<0.0001
Vv18	13.4	UGb8_Vvi2267	p<0.0001	p<0.0005
Vv18	0.0	1038A12F	p<0.0001	p<0.001
Vv18	1.9	1073P15R	p<0.0001	p<0.001
Vv18	24.6	Vv18@4.5	p<0.0001	p<0.005
Vv18	0.0	Vv18@0.51	p<0.0005	p<0.005
Vv18	33.5	Vv18@6	p<0.001	p<0.05
Vv18	38.0	1075G16F	p<0.005	p<0.05
Vv18	44.9	IN0185	p<0.005	p<0.05
Vv8	30.6	VMC5g6.1	p<0.005	p<0.005

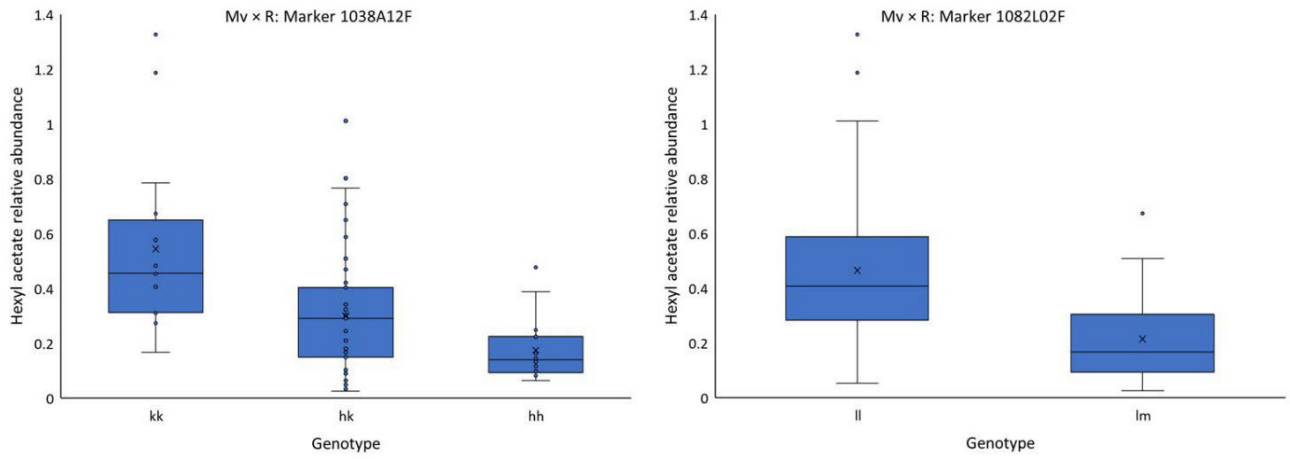


Figure 18 Box plots for two significant markers in chromosome 18 showing the relationship between phenotype and genotype in replicate 1 of the Mv × R cross.

Visualising the marker:trait association using box plot again shows that there is a clear trend for certain genotypes to be associated with higher hexyl acetate concentrations in the wine (Figure 18). In this case the 1038A12F marker shows more of a semi-dominant phenotype, with the heterozygotes (hk) being intermediate between those plants carrying homozygous genotypes. The 'kk' genotype is again associated with higher hexyl acetate concentrations. The 1082L02F marker again segregates into those with higher hexyl acetate concentrations ('ll' genotype) and those that have lower hexyl acetate concentrations ('lm' genotype).

To try and confirm the observations from the Mv × R F1 population, the hexyl acetate trait was also mapped in the Mv × CS F1 progeny. The association on markers with hexyl acetate concentration wasn't as convincing in this population with only a few markers from chromosome 18 showing a significant marker:trait association (Table 9). While the 1073P15R marker was significantly associated with the phenotype at the $p<0.0001$ level in the second replicate, it was less so in replicate one and only two other surrounding markers showed a relatively strong association.

Table 9 Linkage group, position, loci and significance values for the top seven markers from a Kruskal-Wallace analysis of the hexyl acetate wine abundance phenotype in the Mv × CS progeny.

GROUP	POSITION (cM)	MARKER	SIGNIFICANCE	
			REP1	REP2
Vv18	3.4	1073P15R	$p<0.005$	$p<0.0001$
Vv18	12.5	UGb8_Vvi2267	$p<0.005$	$p<0.001$
Vv18	3.4	1038A12F	$p<0.05$	$p<0.005$
Vv3	7.1	1044J09F	$p<0.005$	$p<0.01$
Vv10	126.7	1082F05F	$p<0.05$	$p<0.01$
Vv9	15.7	1085F01F	$p<0.05$	$p<0.01$
Vv6	7.5	lm34134	$p<0.01$	$p<0.01$

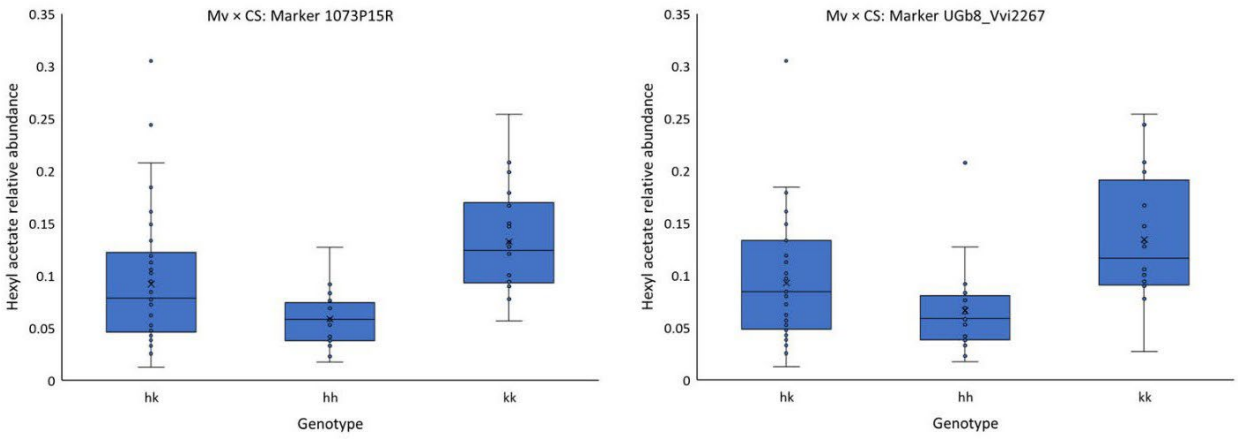


Figure 19 Box plots for two significant markers in chromosome 18 showing the relationship between phenotype and genotype in replicate 2 of the Mv × CS cross.

The box plots of the best two markers and the hexyl acetate concentrations in replicate 2, where the most significant associations were observed, showed that the progeny with the 'kk' genotype tended to have more hexyl acetate in the wines (Figure 19). However, the relationship was not as clear as that observed for phenylethyl acetate (Figure 15). This weaker relationship was confirmed when the MQM mapping of this trait failed to identify any significant QTLs in this population (data not shown).

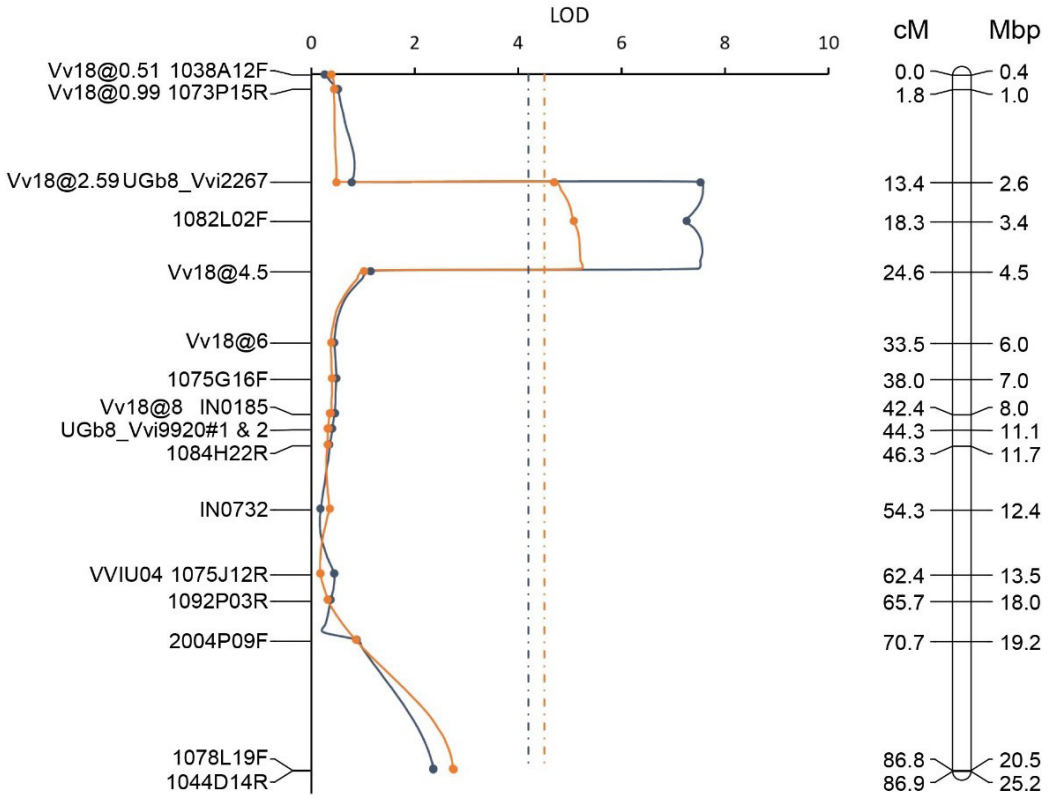


Figure 20 Schematic of the chromosome 18, wine hexyl acetate QTL mapping interval in Mv × R. Shows logarithm of odds (LOD) score against marker on the left and map units (in cM) and physical position (in Mbp) on the right. The vertical dashed lines represent the genome-wide significance threshold ($p < 0.05$). Data from replicate 1 is plotted in grey and data from replicate 2 in orange.

MQM mapping of the trait in the Mv × R cross, however, did identify a single significant QTL using both data replicates. Like for phenylethyl acetate, hexyl acetate concentration in the wines was associated with a region of the grape genome near the top of chromosome 18 for both biological replicates (Figure 20). The highest LOD score in replicate 1 was 7.58 just down from the UGb8_Vvi2267 marker, which is found 13.4 cM from the top of the chromosome. The threshold LOD for $p < 0.05$ was 4.2 for replicate 1 of this QTL analysis and 4.4 for replicate 2 (Figure 16). The peak LOD in for the replicate 2 data was 5.22 and was observed just upstream of the Vv18@4.5 marker. The physical position of these peaks on chromosome 18, between the markers Vv18@2.59 and Vv18@4.5, corresponds to 2.6-4.5 Mbp from the top of the chromosome using the 12x v2 build of the Pinot noir PN40024 genome. The peak in replicate 1 described 31% of the total variation in the phenotype and 28% of the variation in replicate 2.

The third acetate ester analysed was isoamyl acetate, although it showed the weakest correlation between the biological replicates (Figure 13). Nevertheless, the analysis of the Mv × R population with regards to phenylethyl acetate and hexyl acetate suggested that there is a region of chromosome 18 that is associated with the production of acetate esters in general in wines and if isoamyl acetate follows this trend it would strengthen this hypothesis.

The initial Kruskal-Wallace analysis gave good evidence that this hypothesis is correct. There were several markers again at the top of chromosome 18 that showed a significant association with the concentration of isoamyl acetate in the wines (Table 10). However, this relationship was much stronger in the data from biological replicate 1 than in the second replicate, which is likely to be a consequence of the lack of correlation seen between the replicates (Figure 13).

Table 10 Linkage group, position, loci and significance values for the top seven markers from a Kruskal-Wallace analysis of the isoamyl acetate wine abundance phenotype in the Mv × R progeny.

GROUP	POSITION (cM)	MARKER	SIGNIFICANCE	
			REP1	REP2
Vv18	0.0	1038A12F	$p<0.0001$	$p<0.01$
Vv18	1.9	1073P15R	$p<0.0001$	$p<0.05$
Vv18	13.4	UGb8_Vvi2267	$p<0.0001$	$p<0.005$
Vv18	18.3	1082L02F	$p<0.0001$	$p<0.005$
Vv18	24.6	Vv18@4.5	$p<0.0001$	$p<0.01$
Vv18	0.0	Vv18@0.51	$p<0.0005$	$p<0.05$
Vv18	33.5	Vv18@6	$p<0.0005$	$p<0.05$
Vv18	38.0	1075G16F	$p<0.0005$	$p<0.01$
Vv2	20.8	RA0095	$p<0.01$	$p<0.01$

The difference between the replicates was also evident when the trait:genotype relationships were observed using boxplots. First, when plotting the 1038A12F marker, which segregates 'hk × hk', against isoamyl acetate concentration in the wine showed that the progeny with the 'kk' genotype are more likely to produce more isoamyl acetate that those with the 'hk' or 'hh' genotype (Figure 21). However, the separation in isoamyl acetate concentration between the three genotypes was more apparent in replicate 1 then in replicate 2 (Figure 21). Similarly, the 1082L02F marker indicated that those progeny inheriting the 'm' allele from Mv were more likely to have lower isoamyl acetate concentrations in the wines compared to those with the 'll' genotype, with the effect more evident in replicate 1 than replicate 2 (Figure 22).

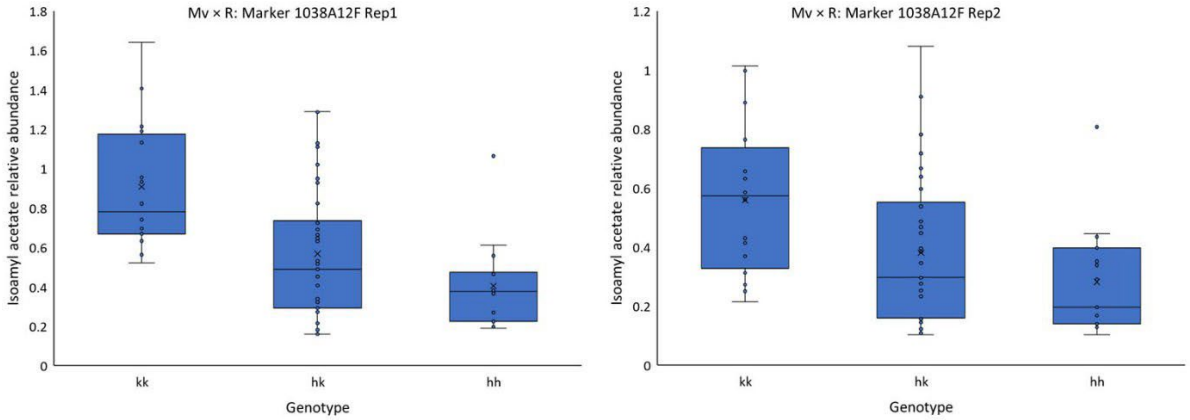


Figure 21 Box plots for the 1038A12F marker on chromosome 18 showing the relationship between phenotype and genotype in replicates 1 and 2 of the Mv × R cross.

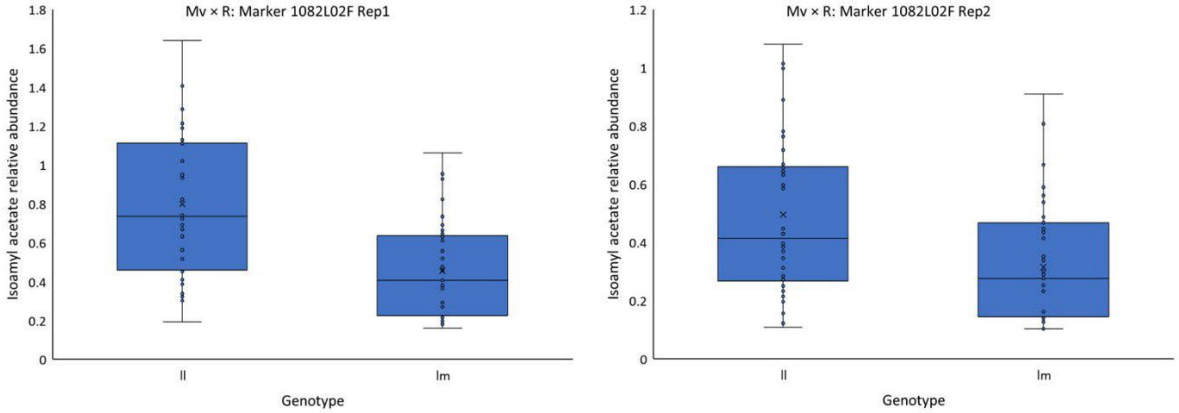


Figure 22 Box plots for the 1082L02F marker on chromosome 18 showing the relationship between phenotype and genotype in replicates 1 and 2 of the Mv × R cross.

Kruskal-Wallace analysis of the isoamyl acetate concentrations in the wines made from both biological replicates of the Mv × CS progeny showed that there were no markers that had significant associations with the phenotype at $p<0.01$. This was confirmed with MQM analysis, which failed to detect any QTLs with LODs above the $p<0.05$ threshold (data not shown).

MQM QTL analysis of the Mv × R population was successful in detecting a significant QTL in the grapevine genome associated with isoamyl acetate production, but only in the data obtained from replicate 1 (Figure 23). The position of the QTL was identical to that detected for phenylethyl acetate and hexyl acetate, between the Vv18@2.59 and Vv18@4.5 markers at the top of chromosome 18. The peak LOD in replicate 1 was 8.54 and was located downstream of the 1082L02F marker. This was higher than the genome-wide LOD threshold at $p<0.05$ which was calculated as being 4.4 for this dataset. This peak was also accounting for 39% of the variation in the isoamyl acetate concentrations in the wines made from the progeny. Analysis of replicate 2 failed to identify a significant QTL associated with isoamyl acetate concentrations as the highest LODs were detected on chromosome 18, but failed to exceed the genome-wide $p<0.05$ threshold (Figure 23).

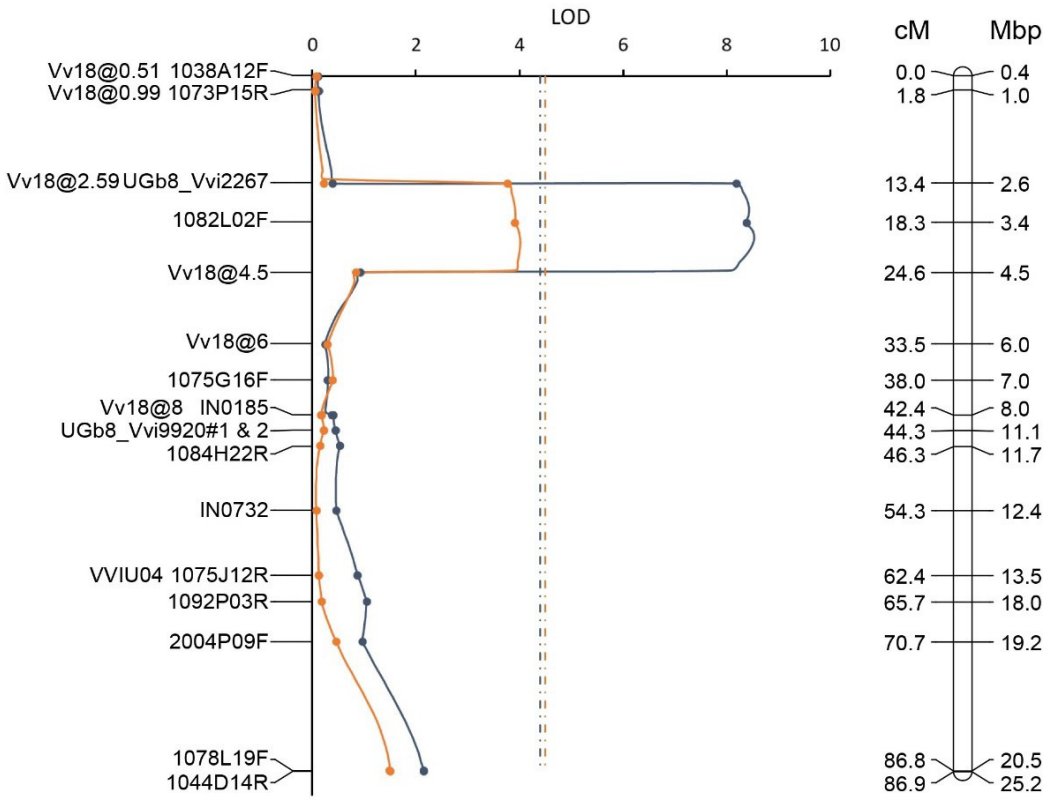


Figure 23 Schematic of the chromosome 18, wine isoamyl acetate QTL mapping interval in Mv × R. Shows logarithm of odds (LOD) score against marker on the left and map units (in cM) and physical position (in Mbp) on the right. The vertical dashed lines represent the genome-wide significance threshold ($p < 0.05$). Data from replicate 1 is plotted in grey and data from replicate 2 in orange.

In conclusion, we were surprised that the concentration of acetate esters in wine was associated with a distinct section of the grapevine genome given the role yeast plays in the production of these compounds during fermentation. Nevertheless, the evidence is convincing that grape composition plays a role in this phenotype given that the same region was identified in two populations for phenylethyl acetate and this QTL was also associated with hexyl acetate and isoamyl acetate concentrations in the Mv × R progeny. The QTL region that overlapped in the Mv × CS and Mv × R maps corresponded to the region between the Vv18@2.59 and Vv18@4.5 markers which is between 2.59 and 4.5 Mbp from the top of chromosome 18. Gene lists from that region and potential candidate genes are discussed later in this report.

Linalool (at moderate 'Riesling-like' concentrations)

Introduction

The targets already presented in the sections above are associated with 'green' and 'fruity' based characters in wine, so linalool was also chosen as a target that is known to contribute to varietal 'floral' or 'citrus' wine sensory attributes. Linalool is a monoterpene that is produced in chloroplasts (Luan and Wüst 2002) from five-carbon isoprene units condensed into the ten-carbon

precursor of most monoterpenes, geranyl diphosphate (Figure 24). Concentrations of linalool and other monoterpenes in grape berries have been used to classify varieties as 'Muscat', 'non-Muscat aromatic' and 'neutral' (Robinson, et al. 2014).

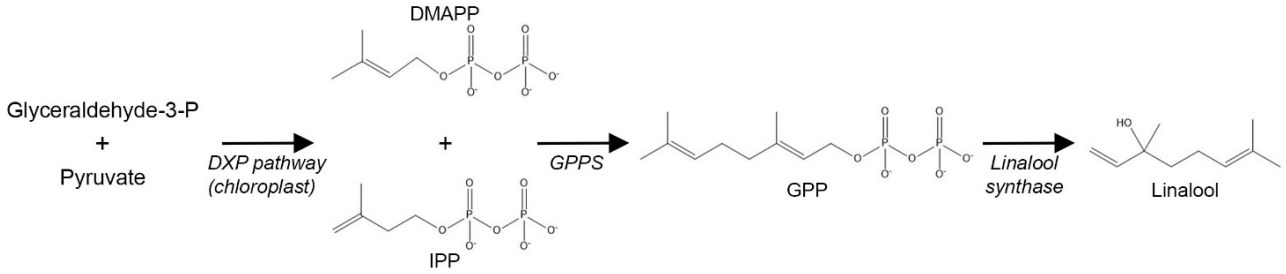


Figure 24 Simplified schematic of the pathway to linalool synthesis in grape berries. DMAPP = dimethylallyl pyrophosphate; DXP = 1-deoxy-D-xylulose 5-phosphate ; GPP = geranyl diphosphate; GPPS = geranyl diphosphate synthase ; IPP = isopentenyl diphosphate.

While the genetic mechanism behind the production of high concentrations of linalool (and other monoterpenes) that creates the 'Muscat' character in grapes and wines is known (Battilana, et al. 2009, Battilana, et al. 2011), the control of moderate monoterpene production, typical of non-Muscat aromatic varieties such as Riesling, is not clear (Emanuelli, et al. 2010). A population of plants from a Mv × R biparental cross had been produced for internal research purposes at CSIRO and it was decided to assess the production of linalool in the progeny to see if segregation of this trait was occurring in these plants.

Data analysis

As with the examination of the volatile compounds in the previous sections above, the first test conducted on the data was to compare replicates by plotting them against each other. While it is known that linalool is grape-derived, it does reveal if the trait will be difficult to map if environmental variation dominates the genetic variation. While the R^2 value was close to 1, indicating that the relationship between the replicates was close, a closer inspection of the graph shows that the slope of the trendline is driven by an outlier that had considerably higher concentrations of linalool in the wines compared to the other progeny (Figure 25). Plotting the individual linalool concentrations for the progeny and parents in replicate 1 shows that the outlier (plant #131) produced similar concentrations of linalool as the Riesling parent (Figure 26). The other plants produced wines with low concentrations of linalool, however, these wines all hade higher abundances (~0.1-0.5) than the microvine control (0.06).

These data suggested that the trait was not segregating in a Mendelian manner. The extreme distortion of segregation was puzzling, but an explanation was hypothesised whereby the fact that we had selected only microvines in the population may have skewed the inheritance of genes linked to the *Vvgai1* mutated allele of the *VvGAI1* gene. Therefore, if an inhibitor of monoterpene production closely linked to the *Vvgai1* allele exists, this would have been inherited in most of the progeny and resulted in the segregation observed in the Mv × R population.

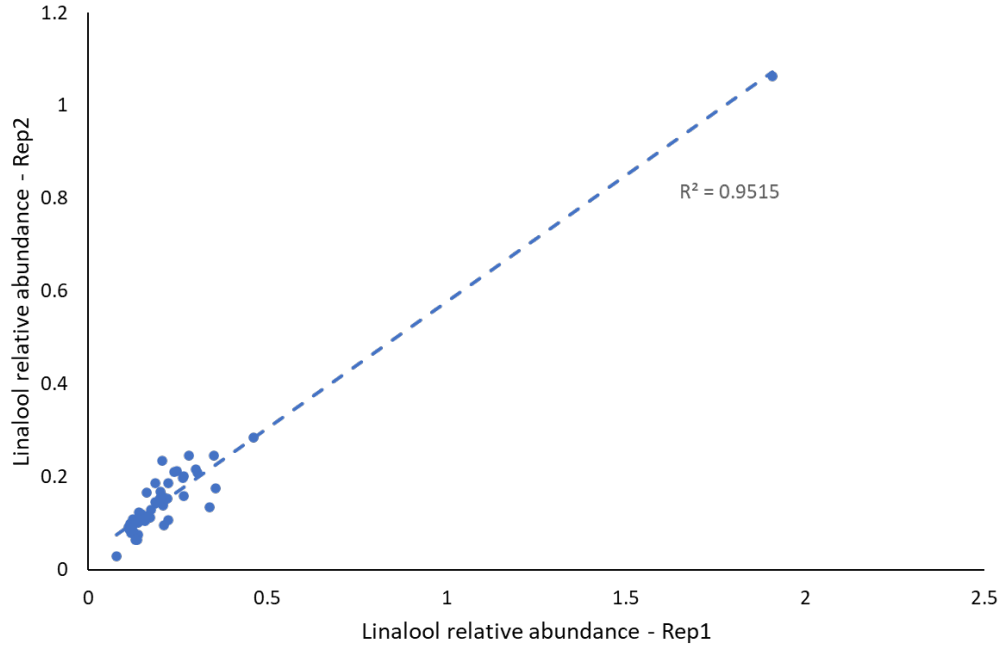


Figure 25 Relative abundances of linalool in both fermentation replicates for the Mv × R population plotted against each other.

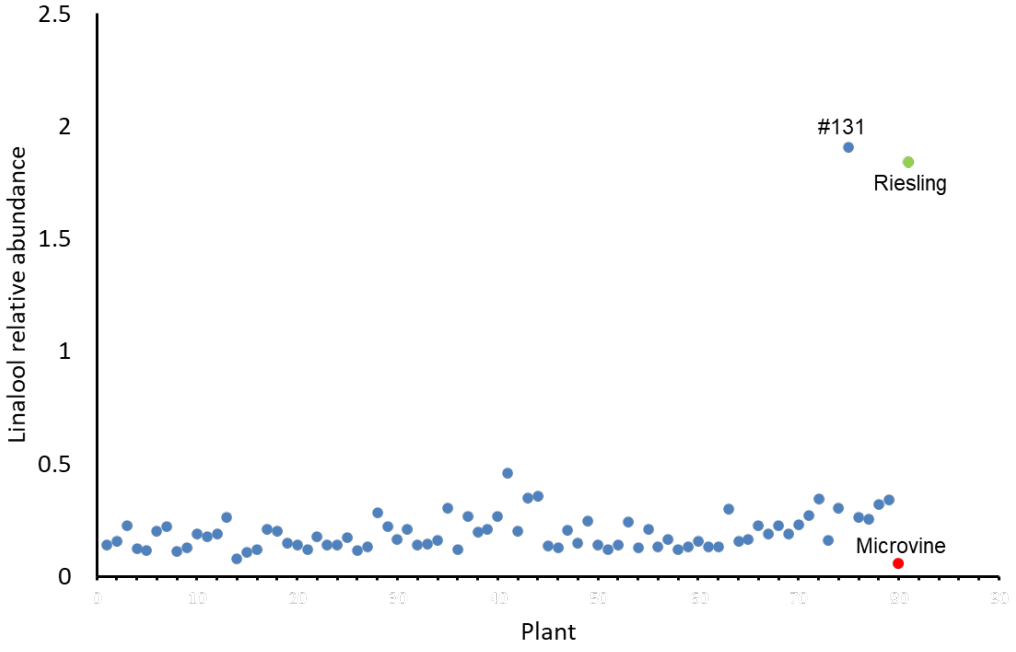


Figure 26 Relative abundances of linalool in fermentation replicate 1 for the Mv × R population and the Mv and R parents.

Marker		VMC4f8	VvChrom1@2.1	Chrom1@2.6	Chrom1@2.7	VvChrom1@3.4	VvChrom1@4.3	VvChrom1@4.57	VvCh1@4.62a	VvCh1@4.62b	VvGAI
Segregation		<efxeg>	<lmxll>	<lmxll>	<lmxll>	<lmxll>	<lmxll>	<lmxll>	<lmxll>	<lmxll>	<lmxll>
Individual	Linalool	631500	2118688	2577064	2727772	3444338	4327716	4571182	4620465	4624618	4867000
1	0.138	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
2	0.158	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
4	0.226						ll				lm
5	0.125	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
6	0.116	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
7	0.203	ef	ll	ll	lm	lm	ll	ll	ll	ll	lm
8	0.224	fg	ll	ll	lm	lm	ll	ll	ll	ll	lm
9	0.113	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
10	0.127	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
11	0.188						ll				lm
12	0.177	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
14	0.189	fg	ll	ll	lm	lm	ll	ll	ll	ll	lm
15	0.264	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
16	0.079	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
17	0.106	fg	ll	ll	ll	lm	ll	ll	ll	ll	lm
18	0.118	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
19	0.210	ee	lm	lm	ll	lm	ll	ll	ll	lll	lm
20	0.201	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
23	0.148	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
25	0.141	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
27	0.120	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
28	0.176	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
29	0.140	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
30	0.140	ef	lm	lm	ll	lm	ll	ll	ll	ll	lm
32	0.173	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
33	0.115	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
34	0.132	ef	lm	lm	ll	lm	ll	ll	ll	ll	lm
35	0.282	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
36	0.222	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
37	0.165	fg	lm	lm	ll	lm	ll	ll	ll	ll	lm
38	0.211	ef	lm	lm	ll	lm	ll	ll	ll	ll	lm
39	0.142	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
41	0.142	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
43	0.161	ef	ll	ll	lm	lm	ll	ll	ll	ll	lm
44	0.305	ef	ll	ll	ll	lm	ll	ll	ll	ll	lm
45	0.119	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
46	0.268	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
48	0.197	ee	lm	lm	ll	lm	ll	ll		ll	lm
50	0.210	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
51	0.267	ef	ll	ll	lm	lm	ll	ll	ll	ll	lm
52	0.461	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
53	0.201	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
54	0.351	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
58	0.355	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
60	0.137	ef	lm	lm	ll	lm	ll	ll	ll	ll	lm
61	0.127	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
65	0.207	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
66	0.149	fg	ll	ll	lm	ll	lm	lm	lm	lm	lm
67	0.248	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
68	0.142	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
69	0.118	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
71	0.142	fg	lm	lm	ll	lm	ll	ll	ll	ll	lm
72	0.241	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
73	0.127	fg	ll	ll	lm	lm	ll	ll	ll	ll	lm
75	0.210	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
76	0.130	ef	lm	lm	ll	lm	ll	ll	ll	ll	lm
77	0.163	ef	lm	lm	ll	lm	ll	ll	ll	ll	lm
79	0.119	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
80	0.133	fg	ll	lm	ll	lm	ll	ll	ll	ll	lm
82	0.158	fg	ll	lm	ll	lm	ll	ll	ll	ll	lm
85	0.133	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
86	0.134	ee	lm	ll	ll	lm	ll	ll	ll	ll	lm
90	0.301	fg	ll	ll	lm	ll	ll	ll	ll	ll	lm
91	0.158	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
97	0.166	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
100	0.225	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
101	0.188	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
102	m	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
105	m	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
109	0.226	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
111	0.190	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
113	0.229	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
117	0.269	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
122	m	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
124	0.344	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
126	0.162	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
127	m	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
128	0.304	ee	lm	lm	ll	lm	ll	ll	ll	ll	lm
131	1.908	ef	ll	ll	lm	lm	lm	lm	ll	ll	lm
132	0.264	fg	lm	lm	ll	lm	ll	ll	ll	ll	lm
133	0.256	eg	lm	lm	ll	lm	ll	ll	ll	ll	lm
135	0.319		lm		ll	lm	ll				lm
138	0.339	fg	ll	ll	lm	ll	lm	lm	lm	lm	lm
L1	0.058	ef	lm	lm	lm	lm	lm	lm	lm	lm	lm
Riesling	1.838	eg	ll	ll	ll	ll	ll	ll	ll	ll	ll

Figure 27 Genotypes of the Mv × R progeny in the region upstream of the *VvGAI1* locus. Marker name, segregation pattern and position from the top of chromosome 1 are listed at the top. Non-crossover genotypes are highlighted in salmon and crossovers in green, with the *VvGAI1* locus in yellow.

QTL analysis

To test the hypothesis about an inhibitor of linalool synthesis being linked to the *Vvgai1* gene some fine mapping was conducted in the region upstream of this gene on chromosome 1. A series of markers were designed to distinguish between the two microvine alleles of genes between the *VvGAI1* and an SSR marker near the top of the chromosome (VMC4f8). These were then tested on the progeny to see if any of the plants had crossovers close to the dwarfing locus (*VvGAI1*) as rare crossover events in this region might explain the presence of a single plant in the population with 'Riesling-like' concentrations of linalool. Interestingly, the #131 plant did have a crossover event relatively close to the *VvGAI1* locus, being located somewhere between ~4.57 and ~4.62 Mbp from the top of chromosome 1 (Figure 27). This crossover could have removed the microvine allele of a gene that encodes an inhibitor of monoterpene production. The explanation for the phenotype of #131, however, would appear to be more complex as there are two other progeny (#66 and #138) that also have crossovers closer to the *VvGAI1* locus but do not produce 'Riesling-like' concentrations of linalool. The simplest explanation for this is that there is another locus segregating in this population that activates the production of linalool. This gene would be heterozygous in Riesling, and while #131 has inherited the allele that induces higher linalool production, neither #66 nor #138 contains this copy of the gene and hence the inhibitor locus from the microvine is irrelevant as those plants do not have the capacity to produce 'Riesling-like' concentrations of linalool.

The hypothesis about the influence of this area of chromosome 1 on the production was tested by conducting a QTL analysis of the trait. Interestingly, the Kruskal-Wallace analysis failed to identify any markers with a significant association with the phenotype at the $p<0.005$ level. When the MQM QTL analysis was conducted, a single QTL was detected on chromosome 1 (Figure 28). The sharp peak corresponds to the Vv1@4.57 marker located ~300 kb upstream of the *VvGAI1* locus and accounted for 43% and 36% of the variation in linalool concentrations in replicate 1 and 2 respectively. For both replicates, the LOD at the Vv1@4.57 marker was higher than the threshold LOD for $p<0.05$; peak LOD in replicate 1 = 27 and threshold LOD = 11, peak LOD in replicate 1 = 18 and threshold LOD = 12.

As a result of these analyses, we decided to pursue two further lines of enquiry with regards to this trait. First, we examined the genes in the crossover region that we took as being between 3.44 and 4.62 Mbp from the top of chromosome 1 to see if any had the potential to be involved in the production of linalool. Any likely candidate genes would be characterised to further explore their function in the suppression of linalool production and this work is described in a later section of this report. Second, the genotypes of #66 and #138 (crossovers close to *VvGAI1* but low linalool production) suggest that there is a promoter of linalool production heterozygous in the Riesling parent. Given that the #131 plant appears to have not inherited the inhibitor close to *VvGAI1*, but has inherited the promoter from Riesling, we would expect this to segregate in an F2 population derived from #131. Hence seed was collected from #131 berries and a new population of microvines generated to test this hypothesis.

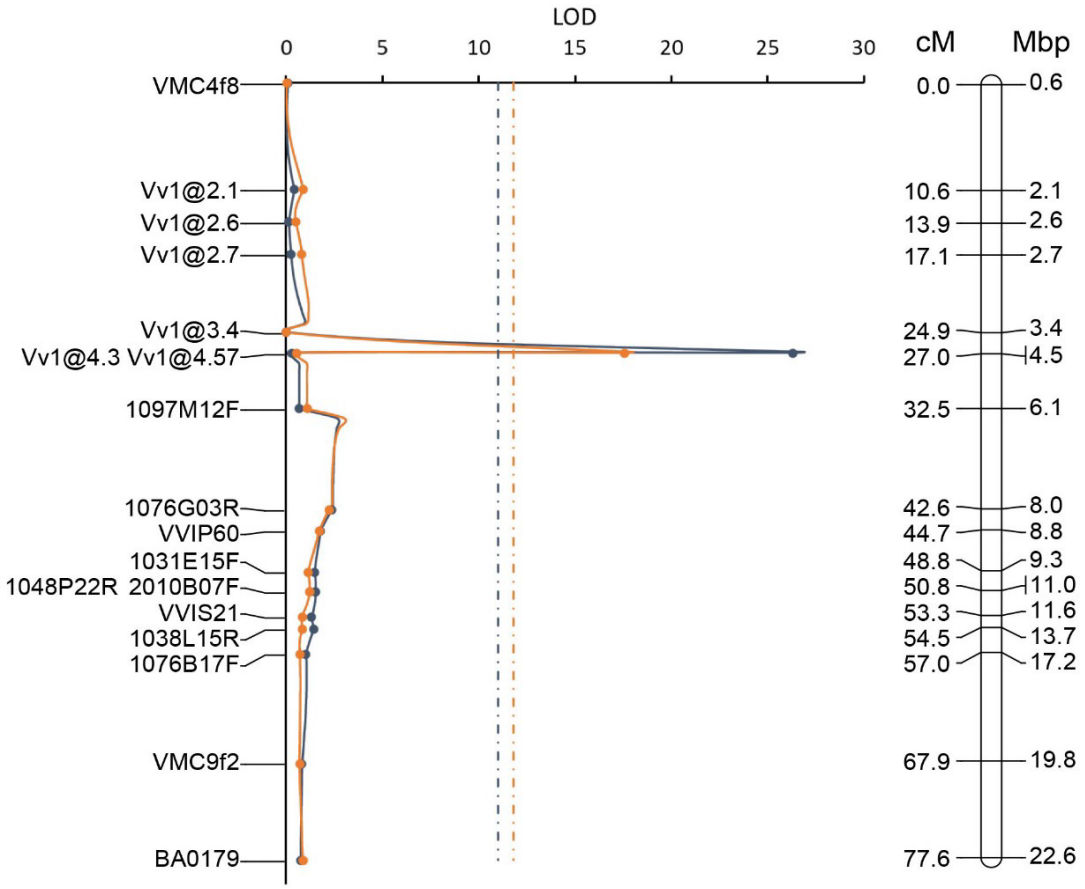


Figure 28 Schematic of the chromosome 1, linalool QTL mapping interval in Mv × R. Shows logarithm of odds (LOD) score against marker on the left and map units (in cM) and physical position (in Mbp) on the right. The vertical dashed lines represent the genome-wide significance threshold ($p < 0.05$). Data from replicate 1 is plotted in grey and data from replicate 2 in orange.

β-Damascenone

Introduction

β-Damascenone is an interesting volatile component of wine that has been shown to enhance fruity characteristics in wine, even when it is present in concentrations below its sensory threshold (Pineau, et al. 2007). The importance of β-damascenone to wine aroma is such that it has been included in a list of compounds measured to estimate grape quality (Cleary, et al. 2015). It is thought to originate from the degradation of the carotenoid neoxanthin into the C_{13}-norisoprenoid known as grasshopper ketone (Figure 29). Grasshopper ketone is then reduced, forming an allenic triol which is then glycosylated into a number of intermediate compounds that can undergo rearrangements in the acidic conditions of wine and be converted into β-damascenone (Winterhalter and Gök 2013). Given its grape origins, we thought it would be a good target for mapping in a suitable population and so the two biparental populations were analysed to try and detect QTLs associated with β-damascenone accumulation in wine.

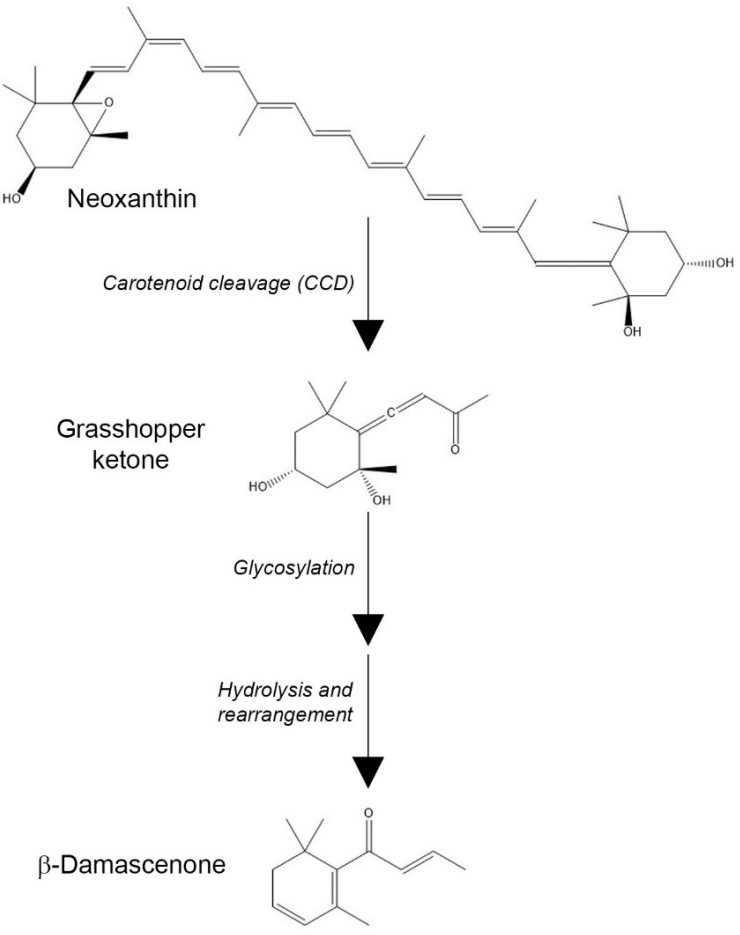


Figure 29 Schematic diagram of the proposed pathway to β-damascenone accumulation in wine. CCD = carotenoid cleavage dioxygenase.

Data analysis

β-Damascenone was quantified in the micro-scale wines produced from the progeny of both the Mv × R and Mv × CS crosses. It was assumed that the amount of free β-damascenone in the wines would be a representative proportion of the potential pool of bound β-damascenone in the grapes, given that the winemaking variables are kept constant. A comparison of the two winemaking replicates would give us an indication if this assumption was merited. The data showed that the winemaking replicates for the Mv × R cross were not as consistent as those in the Mv × CS cross (Figure 30).

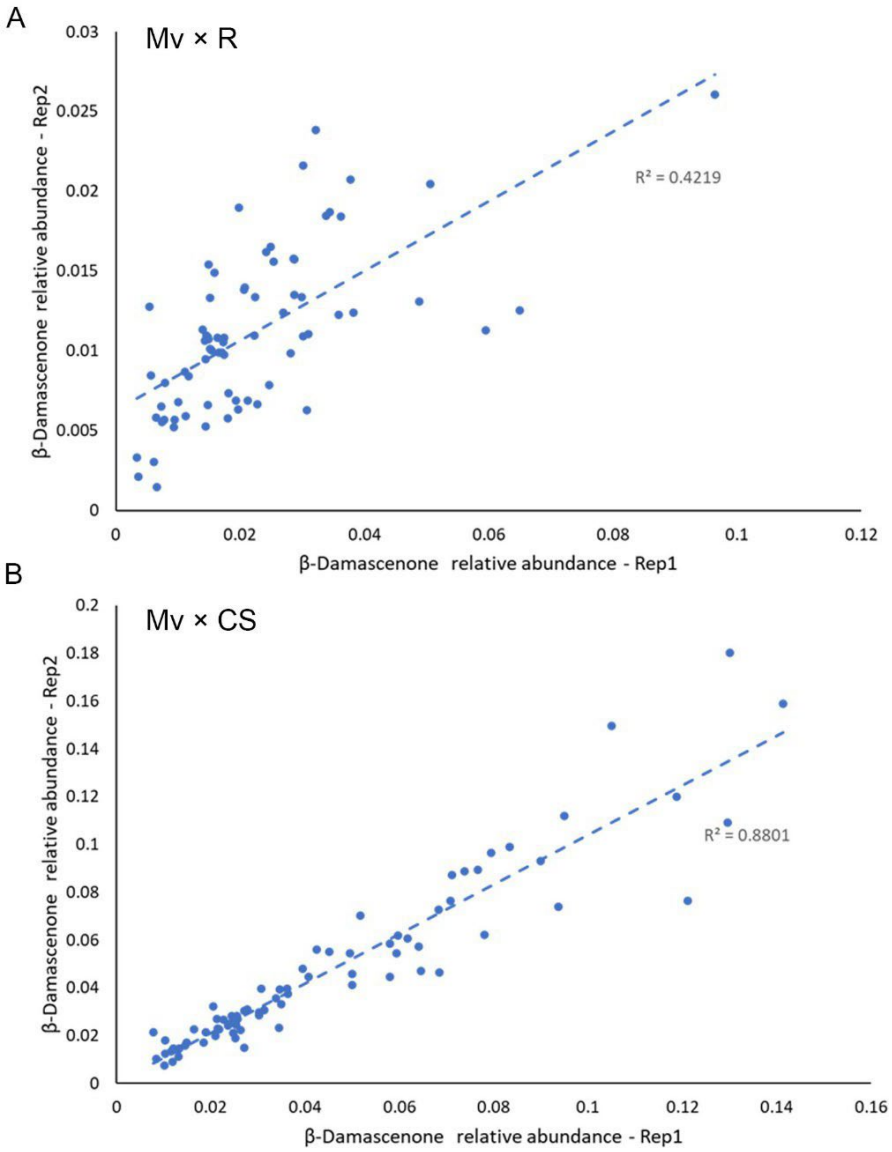


Figure 30 Relative abundances of β-damascenone in both fermentation replicates for the Mv × CS and Mv × R populations plotted against each other.

The main difference between the replicates in the Mv × R progeny was that the β-damascenone abundances in the second replicate were generally lower than those observed in replicate 1 (Figure

30). Nevertheless, it was decided to analyse both sets of data for associated QTLs, even though the Mv × CS data were more promising given the close relationship between the replicates.

QTL analysis

The first analysis conducted was the non-parametric test for marker:trait association using the Kruskal-Wallace rank-sum test. The Mv × R cross only had one marker with a significance level of at least 0.005 in either replicate and this was the VVIP60 marker on chromosome 1 which was only significant at this level in replicate 1. An examination of QTLs using MQM was conducted for this cross, but no region was found to be significantly associated with the trait. When the Mv × CS cross was analysed using the Kruskal-Wallace test, the VVIB23 marker on chromosome 2 was significant at the $p < 0.0005$ level in both replicates and no other marker was significant at a level lower than $p < 0.005$ in either replicate.

The MQM QTL analysis of the β-damascenone concentrations in the Mv × CS progeny resulted in the detection of a single significant QTL on chromosome 2 in both replicates (Figure 31).

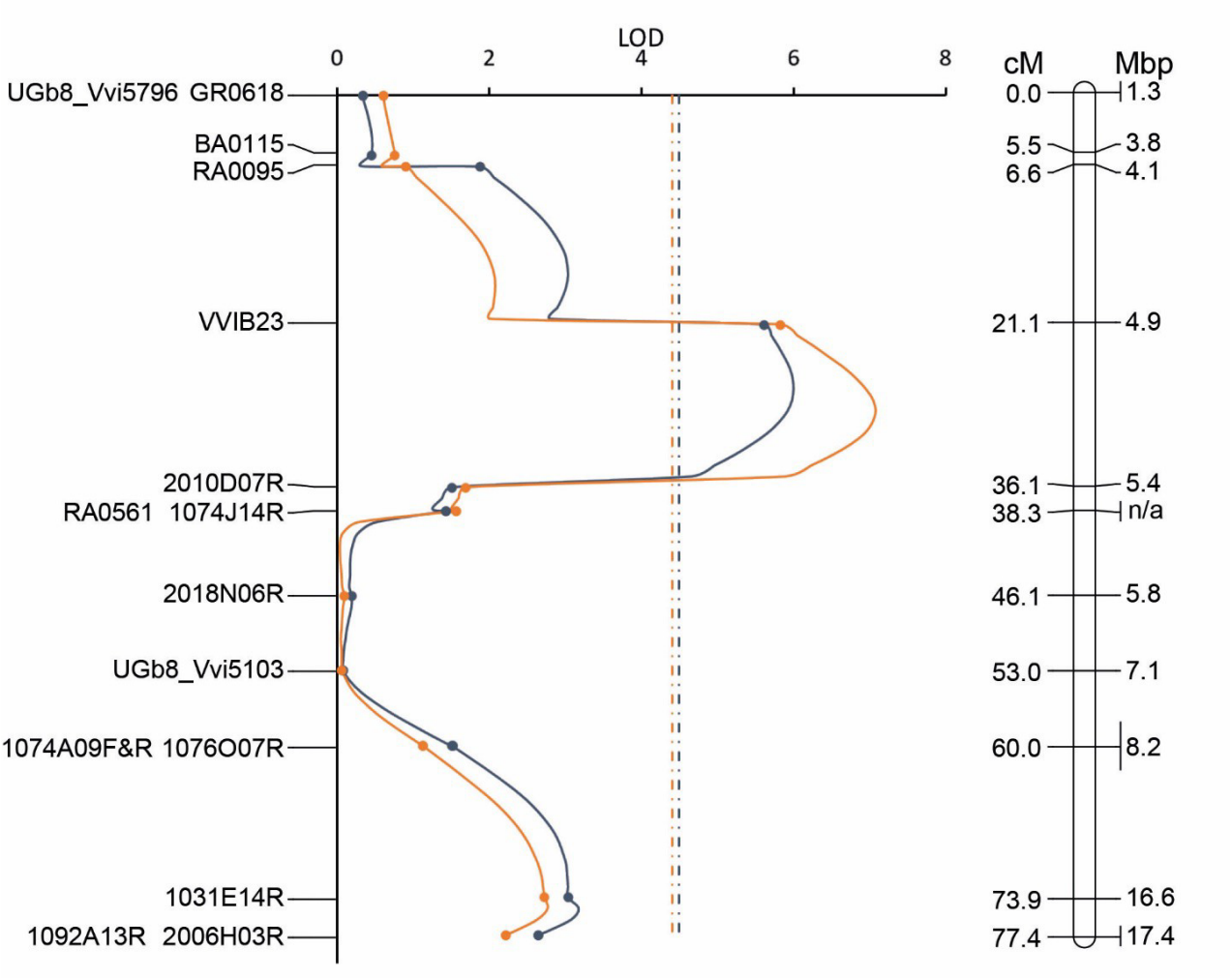


Figure 31 Schematic of the chromosome 2, β-damascenone QTL mapping interval in Mv × CS. Shows logarithm of odds (LOD) score against marker on the left and map units (in cM) and physical position (in Mbp) on the right. The vertical dashed lines represent the genome-wide significance threshold ($p < 0.05$). Data from replicate 1 are plotted in grey and data from replicate 2 in orange.

The QTL was positioned between the markers VVIB23 and 2010D07R in the upper half of chromosome 2. The peak LOD in replicate 1 was 6 and was located midway between VVIB23 and 2010D07R. This was higher than the genome-wide LOD threshold at $p<0.05$ which was calculated as being 4.5 for this dataset. This peak was also accounting for 31% of the variation in the β-damascenone concentrations in the wines made from the progeny. The LOD threshold ($p<0.05$) for replicate 2 was 4.4 and the peak LOD was higher than this being 7.1. Again, this was located midway between the two markers listed above and, in this dataset, represented 35% of the variation in β-damascenone concentrations.

It was encouraging to locate a QTL in the Mv × CS cross that was associated with the production of this important C13-norisoprenoid. Unfortunately, this was not supported by the mapping in the Mv × R cross. The comparison of the replication of the phenotype may reveal why this mapping was less successful in the Mv × R cross. The graph in Figure 30 shows that the winemaking replicates in this cross were less correlated than those from the Mv × CS cross. This suggests that the variation in the data is caused more by the winemaking process in the Mv × R progeny and that the genetic variation in β-damascenone biosynthesis in this population is not as great as in the Mv × CS progeny. The region on chromosome 2 between the VVIB23 and 2010D07R markers falls between 4.9 and 5.4 Mbp from the top of the chromosome and gene lists from the region and potential candidate genes are discussed later in this report.

6.2 Mapping traits using genome-wide association studies

6.2.1 Introduction

Genetic mapping can be summarised as a method of finding the causal relationship between genetic polymorphisms and phenotypic traits. Traditionally this has been conducted using mapping populations that segregate for the trait of interest as demonstrated in section 6.1 above. The concept of doing such genetics in grapevine has been improved in recent years through the discovery of the rapid cycling microvine, which makes such experiments possible in the three year timeframe of a normal research grant. This research group has previous used this approach to determine the gene responsible for methoxypyrazine biosynthesis in berries (Dunlevy, et al. 2013). However, it does have its limitations. When analysing the progeny of the biparental cross we are only capturing the allelic diversity that is present in the parents, and while grapes are highly heterozygous, we may not discover important phenotypic variants present in the broad range of grape varieties. Second, the resolution of the map is reliant on recombination frequency. This can be improved by increasing population size, but with grape being a woody perennial, space is often a limiting factor. Again, the microvine helps in this regard, but we cannot expect to plant hundreds or thousands of plants as is done with annual crops. Fortunately, the availability of germplasm collections enables an alternative approach to be used in parallel with the biparental crosses.

Genome-wide association studies (GWAS) represent a complementary tool that can be used to support biparental crosses and act as a way of discovering novel genotype-phenotype relationships in a crop (for a review see Korte and Farlow 2013). The technique is based on the availability of large single-nucleotide polymorphism (SNP) datasets from the plants of interest in addition to the phenotypic data used to make the associations. Fortunately, the availability of cheap, accurate next generation sequencing technologies has enabled such studies to become achievable in a range of organisms. While GWAS can overcome the major limitations of the biparental studies outlined above, it has its own limitations. The main ones involve the presence of rare variants in the population having a large effect on the phenotype and the converse of common variants having only a small effect on the phenotype. The other main issue is the effect of linkage disequilibrium (LD) on the number of markers required to avoid missing causal SNPs. If LD decays quickly, more markers are required to get a resolution that will cover the genome and unfortunately in grapevine, the LD decay has been calculated as being rapid (Marrano, et al. 2017). Nevertheless, the potential to identify novel SNPs associated with a consumer trait in parallel with the biparental mapping justifies taking this approach.

6.2.2 Data analysis

Genotypic data

Genotypic data from the germplasm collection was obtained from Diversity Arrays Technology (Canberra, Australia) using a methodology known as DArTseq. In brief, genomic DNA from each plant is digested with two restriction enzymes, one a frequent cutter and the other a rare cutter. Adapters are ligated to the ends of the resulting DNA fragments, one being specific to the frequent cutter and another to the rare cutter. Next generation sequencing is then conducted using a specific adaptor to the rare cutter and the resulting short sequences mapped to the

genome and examined for polymorphisms when compared to the PN40024 genome sequence. In total, 112026 SNP markers were obtained.

The SNP markers were then filtered based on call rate (> 0.95), average repeatability (> 0.995) and read depth (lower= 5, upper= 60). Finally, SNPs with a minor allele frequency < 0.05 were excluded from further analysis. This resulted in a final dataset of 10,510 informative SNPs that were included for the mapping process.

To check for any underlying genetic structure in the variety collection, the figure SNP dataset was used to generate a PCA plot describing the variation among the DNA samples analysed (Figure 32). As can be seen in both plots, there was a relatively random spread of samples across the three PCs with no distinct groups of large numbers of samples forming structure in the data. Unexpectedly, the four 'Pinot' samples (noir, gris, blanc and Meunier) cluster on top of each other (circled in Figure 32), and there was separation of the French-American hybrids sampled from the collection negatively on PC2. Nevertheless, the results suggest that no special data treatment is required to avoid any underlying structure in the SNP analysis.

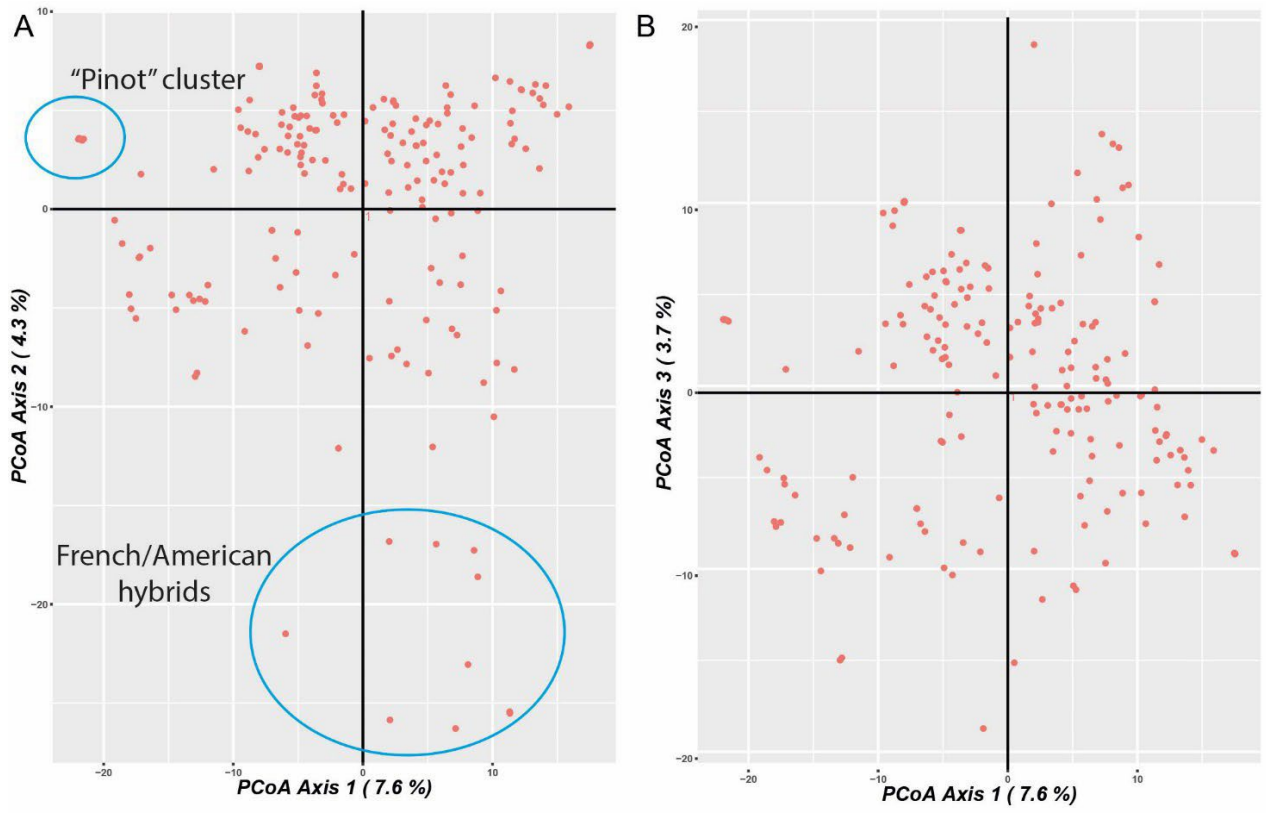


Figure 32 PCA plots of PC1 vs PC2 (A) and PC1 vs PC3 (B) of the varieties genotyped for GWAS.

Phenotypic data

Micro-scale wines were made from biological replicates taken from the germplasm collection at the SARDI research station in Nuriootpa. A majority of black wine grapes were sampled in 2019 and white varieties in 2020. The volatile composition of the wines was analysed using SPME GC/MS and

53 compounds targeted for quantification based on factors such as their potential to be influenced by grape composition and the ability to detect them in a GC-MS scan. One of the first checks of the data was to compare the two replicates within each vintage (Table 11) as this helps determine if the compound is influenced by the fermentation variables more than the fruit composition as discussed previously. This acted as a screen for the subsequent GWAS analysis, as generally, those with low correlations were not used any further.

Table 11 List of compounds analysed in the micro- scale wines for the GWAS study with correlation (R^2) between replicates sampled in each vintage.

Compound	Correlation between replicates (R^2)	
	2019	2020
Benzaldehyde	0.63	0.74
Benzyl alcohol	0.85	0.74
Butanol	0.08	0.16
γ-Butyrolactone	0.20	0.29
2-Carene	0.82	0.74
1,8-Cineole	0.27	0.01
β-Citronellol	0.89	0.93
p-Cymene	0.81	0.62
β-Damascenone	0.76	0.28
Diethyl pentanedioate	0.48	0.74
Diethyl succinate	0.23	0.81
3-Ethoxy-1-propanol	0.66	0.69
2-Ethyl-1-hexanol	0.05	0.04
Ethyl (*E*)-3-hexenoate	0.10	0.24
Ethyl (Z)-3-hexenoate	0.25	0.42
Ethyl 2-methylbutanoate	0.37	0.53
Ethyl 3-hydroxybutanoate	0.54	0.70
Ethyl 3-hydroxytridecanoate	0.18	0.39
Ethyl 3-methylbutyl succinate	0.69	0.81
Ethyl 4-hydroxybutanoate	0.13	0.64
Ethyl 9-decenoate	0.49	0.38
Ethyl decanoate	0.37	0.27
Ethyl heptanoate	0.03	0.12
Ethyl hexanoate	0.07	0.32
Ethyl isovalerate	0.40	0.55
Ethyl octanoate	0.29	0.56
Ethyl pentanoate	0.02	0.01
Ethyl phenylacetate	0.36	0.05

Compound	Correlation between replicates (R^2)	
	2019	2020
2-Heptanol	0.07	0.64
1-Hexanol	0.75	0.66
(*E*)-3-Hexenol	0.62	0.69
(*Z*)-3-Hexenol	0.83	0.96
Hexanoic acid	0.29	0.25
Hexyl acetate	0.14	0.30
Hotrienol	0.84	0.80
Isoamyl acetate	0.01	0.24
Isoamyl alcohol	0.15	0.61
Linalool	0.99	0.79
2-Methyl butanoic acid	0.29	0.50
3-Methyl-1-pentanol	0.52	0.89
4-Methyl-1-pentanol	0.45	0.69
2-Methyl-1-propanol	0.078	0.44
3-(Methylthio)-1-propanol	0.52	0.92
Methyl hexanoate	0.49	0.01
2-Nonanol	0.79	0.51
2-Nonanone	0.35	0.41
1-Octanol	0.72	0.61
Octanoic acid	0.05	0.07
Phenylethyl acetate	0.73	0.68
Phenylethyl alcohol	0.16	0.61
Propanoic acid	0.09	0.15
α-Terpinene	0.69	0.32
α-Terpineol	0.89	0.44

Data were also tested for their normality. While the GWAS methodologies are thought to be robust against datasets that deviate from normality, results may be improved by normalising data before analysis. Additionally, while the data were being analysed, the opportunity was taken to look for varieties that were outliers with regard to the production of specific wine volatile compounds. Outliers were identified if the values were at least three times the interquartile range (Q3-Q1) greater than the Q3 value or at least three times the interquartile range less than the Q1 value (Table 12). These may be interesting varieties to use for future crosses designed to explore the production of important grape-derived wine volatile compounds. We have excluded the compounds that had low repeatability between the replicates and hence were not considered for GWAS and have indicated the monoterpenes that are found in very high concentrations in 'Muscat' varieties.

Table 12 List of compounds analysed by GWAS, indication of the result of the normality test (Shapiro-Wilk) and list of varieties with outliers in each year. The variety in brackets and italicised was an outlier because it had much lower concentrations; the rest were outliers due to high concentrations.

Compound	Normality	Major outliers	
		2019	2020
Benzaldehyde	N	Criolla negra, Malbec, Clairette, Ugetha, Zinfandel	Perle, Doradillo, Enhresfelser
Benzyl alcohol	N	Americano, Auldana #3 (Mission)	Perle, Orange Muscat, Semillon, Morio Muscat
2-Carene	N	Monoterpene	Monoterpene
β-Citronellol	N	Monoterpene	Monoterpene
p-Cymene	N	Monoterpene	Monoterpene
β-Damascenone	Y	Grec rose, Ugetha	-
Diethyl pentanedioate	N	-	Viognier, Sauvignonasse, Seneca, Verdelet, Verdelho
Diethyl succinate	Y	-	-
3-Ethoxy-1-propanol	N	-	-
Ethyl 3-hydroxybutanoate	N	Touriga	Seneca
Ethyl 3-methylbutyl succinate	N	Kadarka, Pinot Meunier, Muscat a petite grains rouge, Carignan, Pinot noir	Viognier
Ethyl 4-hydroxybutanoate	N	-	-
Ethyl 9-decenoate	Y	Americano	-
Ethyl isovalerate	Y	-	-
Ethyl octanoate	Y	-	Viognier
1-Hexanol	Y	-	-
(*E*)-3-Hexenol	Y	-	-
(*Z*)-3-Hexenol	N	Egiodola, Auldana #1, Arinarnoa	Harslevelu, Dourado, Furmint, Liliorila
Hotrienol	N	Monoterpene	Monoterpene
Isoamyl alcohol	N	-	Viognier, Orange Muscat, Semillon, *(Kekneylu)*
Linalool	N	Monoterpene	Monoterpene
4-Methyl-1-pentanol	Y	-	-
3-(Methylthio)-1-propanol	N	-	Rolle, Verdelet, Cabernet Sauvignon
2-Nonanol	N	Muscat a petite grains rouge, Carignan, Flora, Malbec, Montils, Touriga, Couderc noir, Grec rose, Tempranillo, Royalty	-
2-Nonanone	N	Flora, Freisa, Grec rose, Tempranillo	-
1-Octanol	N	Grec rose, Tempranillo	0.61

Compound	Normality	Major outliers	
		2019	2020
Phenylethyl acetate	N	-	Verdelet, Perle
Phenylethyl alcohol	Y	-	-

A majority of compounds targeted for GWAS were not normally distributed (Table 12). This was often due to the presence of outliers that skewed the data, mostly to the higher end of the distribution. Some varieties were outliers for several compounds (e.g. Grec rose and Viognier), which raises the question of a systemic error with the volatile compound analysis, and in these cases the chromatograms were reanalysed to confirm the findings. Of main interest were the varieties that had considerably higher concentrations of some of major compounds of interest, namely, β-damascenone, phenylethyl acetate and (*Z*)-3-hexenol. In the 2019 vintage, two varieties were classified as major outliers with regards to their β-damascenone content (Figure 90), Grec rose and Ugetha, although there were five other minor outliers. Phenylethyl acetate concentrations were very much greater in Verdelet and Perle in the 2020 vintage (Figure 90). Several varieties were major outliers with regards to (*Z*)-3-hexenol concentrations in both vintages (Figure 90).

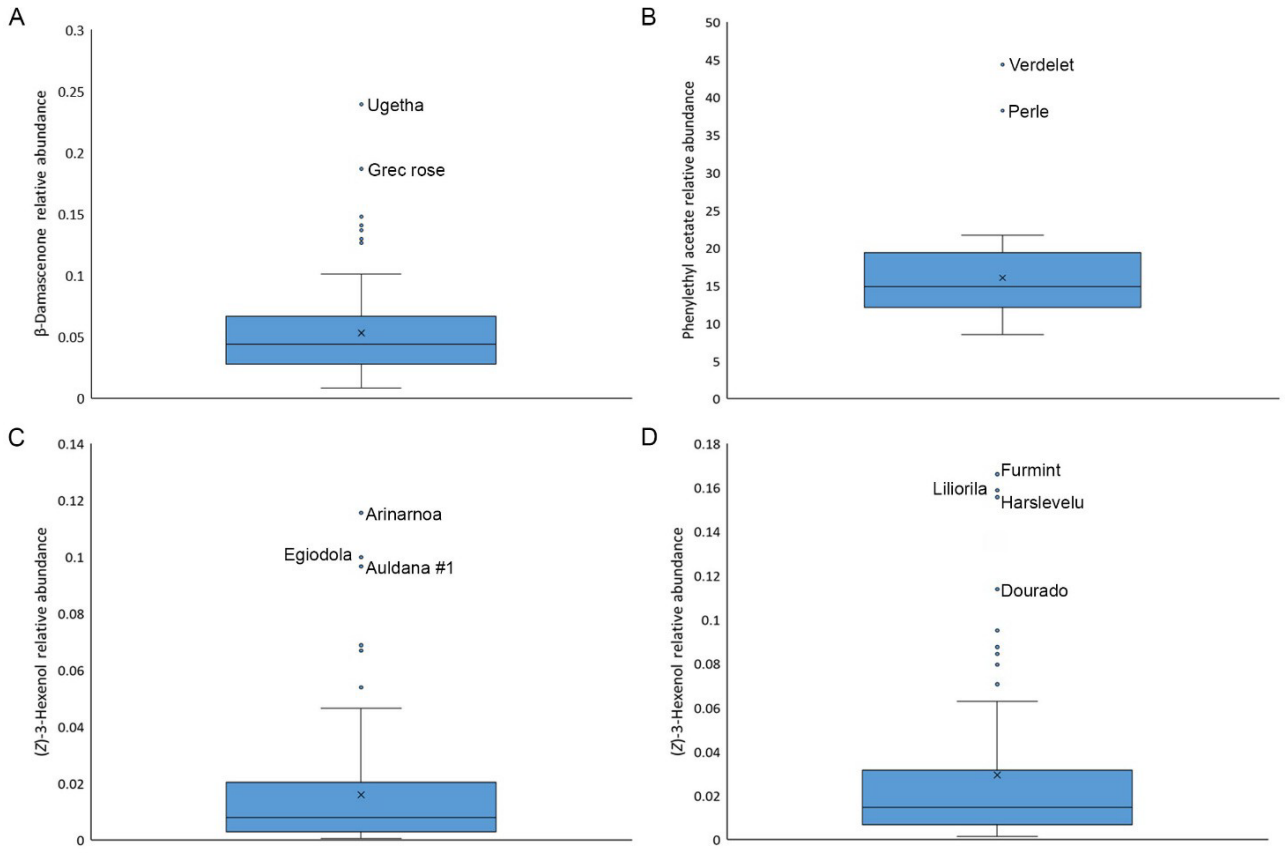


Figure 33 Box plots of relative abundance data from some important targets naming major outliers (larger or smaller than 3 times the interquartile range). The graphs show the mean (x), median as the horizontal line in the box, the upper and lower quartiles, whiskers indicate the smallest and largest values within 1.5 times the interquartile range and other dots represent minor outliers (between 1.5 and 3 times the interquartile range). A = β-damascenone from 2019; B = phenylethyl acetate from 2020; C = (*Z*)-3-hexenol from 2019; D = (*Z*)-3-hexenol from 2020.

GWAS

Introduction

With the genotypic and phenotypic data collected, the next step in the process is the GWAS. Analysis was conducted as outlined in the methods section. As a point of clarification, several GWAS models were used for each compound and normalisation of the volatile compound quantitative data was conducted in some cases depending on the observations made during data analysis or the GWAS itself. The different GWAS models represent a range of statistical techniques used to assess SNP:phenotype associations and we refer those interested to an excellent recent review for more information (Tibbs Cortes, et al. 2021). In short, linear models such as mixed linear model (MLM) and multiple loci mixed linear model (MLMM) account for underling structure that might exist in the population, but they are quite conservative. These methods lack the statistical power of more recent multilocus models such as fixed and random model circulating probability unification (FarmCPU) and Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) (Huang, et al. 2019). However, the methods ask different questions about the genotype:phenotype relationship, and so it is normal practice to run various models for each data set. Due to the lack of any obvious structure in our population data, we have focused on MLMM and BLINK results in this report as the single locus analytical methods such as MLM are less powerful and their ability to account for genetic relatedness among subsets of samples is not essential for our dataset.

The phenotypic data were also treated in different ways during the GWAS. As for most statistical tests, normality of the data is assumed, but in essence, transforming non-normally distributed data should increase the power of the analysis but not render analysis on untransformed data invalid. In fact, in some cases where sample size is sufficient and the phenotypic variation is large, normalisation is discouraged as it increases false discovery rate (Goh and Yap 2009). Nevertheless, given that many of our datasets are not normally distributed (Table 12), Box-Cox and log transformations were used for phenotypic data from certain volatile compounds to try and improve the power of the analysis. We also found that some phenotypes were mapping to the colour locus on chromosome 2. For example, 1-hexanol mapped to two markers on chromosome 2 immediately downstream of the *VvMYBA1* gene responsible for skin colour (Figure 34). This would suggest that skin colour has an influence on this phenotype in the wine. However, the sampling regime we used resulted in predominantly red-skinned varieties being sampled in 2019 and mainly white-skinned varieties in 2020. Therefore, if there was a major vintage effect on the concentration of certain wine volatiles, this may result in the trait being mapped to the colour locus. Plotting the 1-hexanol results from 2019 and 2020 shows that there was a difference in the spread of results from each vintage (Figure 35), confirming that the mapping result (Figure 34) may be due to a vintage effect. So, unfortunately, if berry skin colour was having an effect on volatile composition, we are unable to distinguish this from a vintage effect with the sampling regime used. For compounds where we suspected there was a significant vintage effect, another type of normalisation called rank-based inverse normalisation was used within each vintage to account for the vintage effect before proceeding to GWAS.

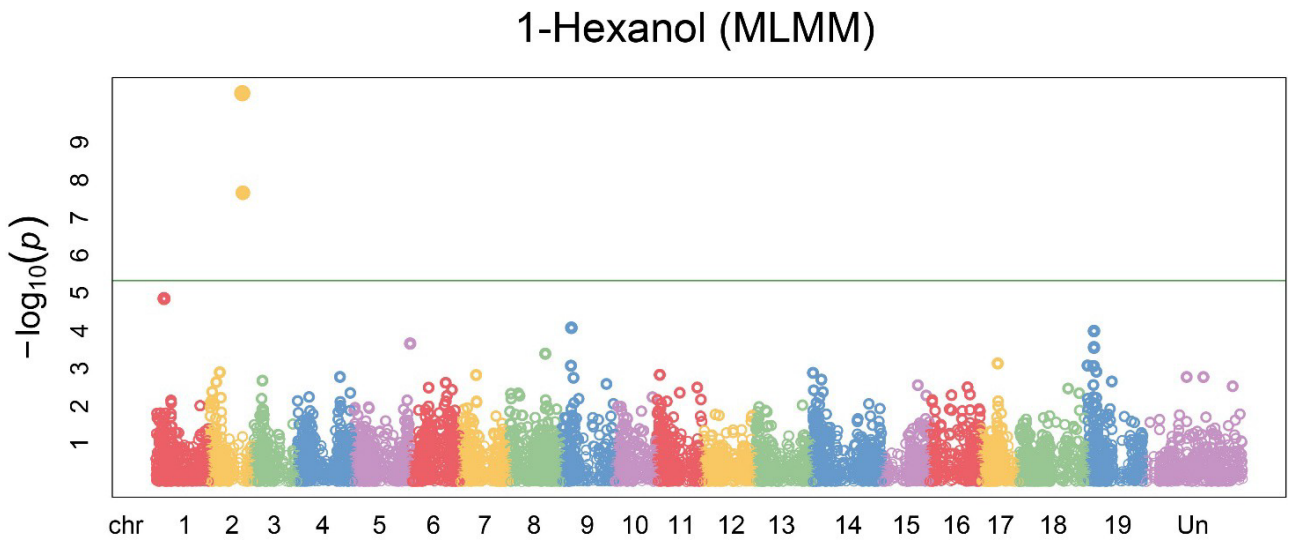


Figure 34 GWAS of 1-hexanol in the germplasm collection using the MLMM model. The horizontal line represents the significance threshold and markers grouped by chromosomes with unmapped markers indicated by 'Un'.

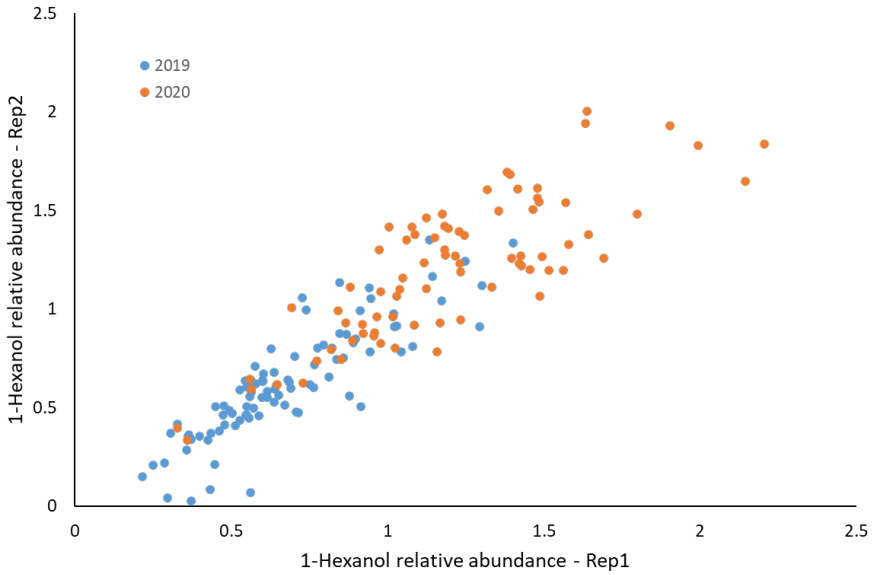


Figure 35 Plot of 1-hexanol abundances in replicate wines produced from various varieties from the germplasm collection with blue data points from the 2019 vintage and orange from the 2020 vintage.

Benzaldehyde

Benzaldehyde is responsible for 'bitter almond' sensory characteristics in wine. Benzaldehyde abundances in the small-scale wines did appear to show a vintage effect, but as there were some samples with extreme phenotypes, it was decided to conduct the analyses without normalisation as the presence of high phenotype variation can increase the power of GWAS. Interestingly, even with this vintage effect the mapping revealed the presence of QTLs on chromosomes 4 and 16 with two statistical models (Figure 36). The two SNPs on chromosome 16 are located near the bottom of the chromosome, one at 16.8 Mbp and the other at 17.8 Mbp. The SNP at 16.8 Mbp does fall in a gene, but the protein product is uncharacterised and has no strong homology to genes of known function.

It is predicted to have a transmembrane domain. The SNP falls in the coding region of a proline-tRNA ligase gene. While it is questionable whether this gene itself functions to control of benzaldehyde abundances in wine, it could be that genes linked to this SNP are involved, although as discussed above, linkage disequilibrium rapidly decays in grapevine. Just upstream of the SNP are a number of stilbene synthase genes and, given that a both stilbenes and benzaldehyde biosynthesis share a proposed precursor, it is possible that a SNP in one of these genes could be responsible for the trait. The SNP on chromosome 4 is located in a gene encoding a peptide chain release factor from the chloroplast and this protein functions to terminate polypeptide synthesis. A nearby gene with potential function in the production of benzaldehyde or its precursors encodes a dehydrogenase/reductase enzyme. These SNPs have the potential to be used as markers to predict progeny with a greater probability of producing wine with high or low benzaldehyde concentrations.

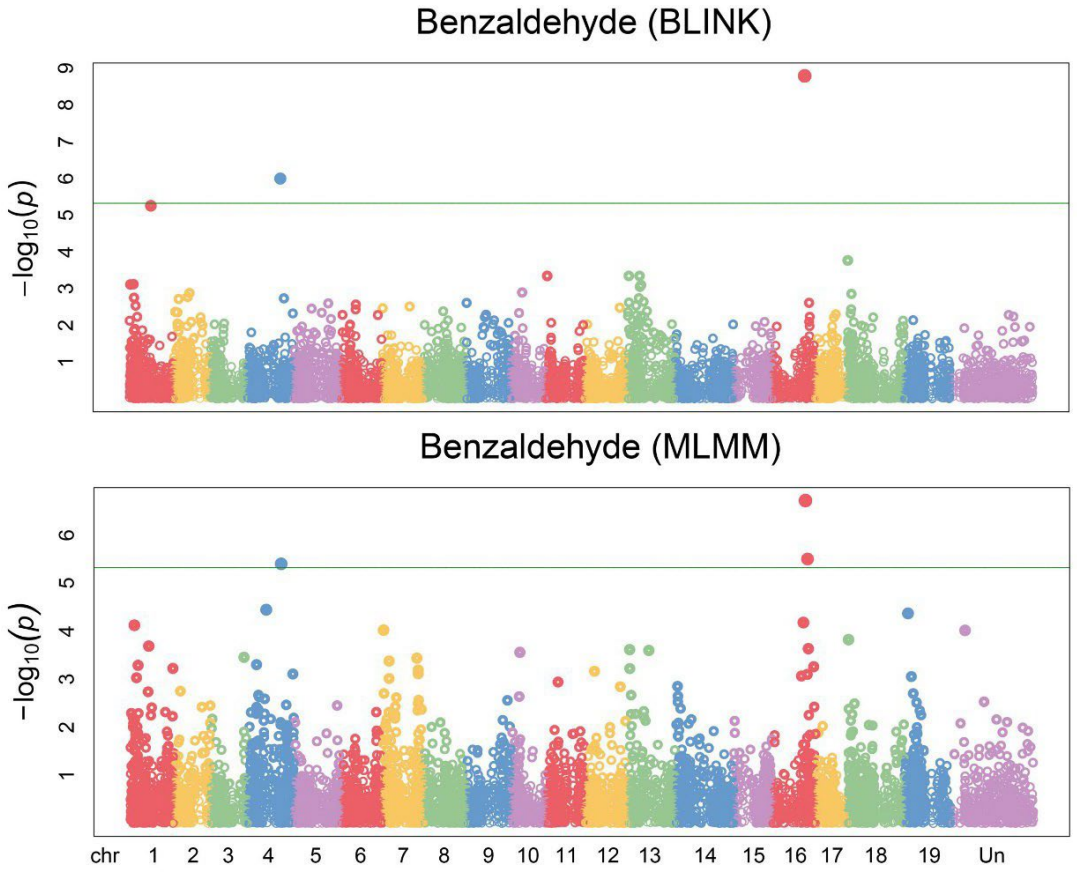


Figure 36 GWAS of benzaldehyde in the germplasm collection using the BLINK and MLMM models. The horizontal line represents the significance threshold and markers grouped by chromosomes with unmapped markers indicated by 'Un'.

Benzyl alcohol

Benzyl alcohol is described as having 'berry' and 'cherry' sensory notes (Genovese, et al. 2015) and is known to change significantly in grapes during the ripening process (Kalua and Boss 2010). There was an obvious vintage effect with regards to the production of this compound in the wine samples, and so the rank-based inverse normal transformation was used on the data before the GWAS. In all the models tested there were no SNPs that exceeded the significance threshold at the 0.05 level. Two SNPs, one on chromosome 14 and another on chromosome 17 were significant at the 0.1 level

when the general linear model (GLM) was used, but no obvious candidate genes were in the vicinity of either SNP in the grape genome.

2-Carene

The monoterpene 2-carene has been described as having a citrus/turpentine like sensory character and is more abundant in 'Muscat' varieties. When using non-normalised data, the GWAS models did not give consistent results and so Box-Cox transformed data were applied before more modelling was attempted. This gave consistent results when using the various models, although the increased power of BLINK resulted in more SNPs being above the significance threshold than the MLMM model (Figure 37).

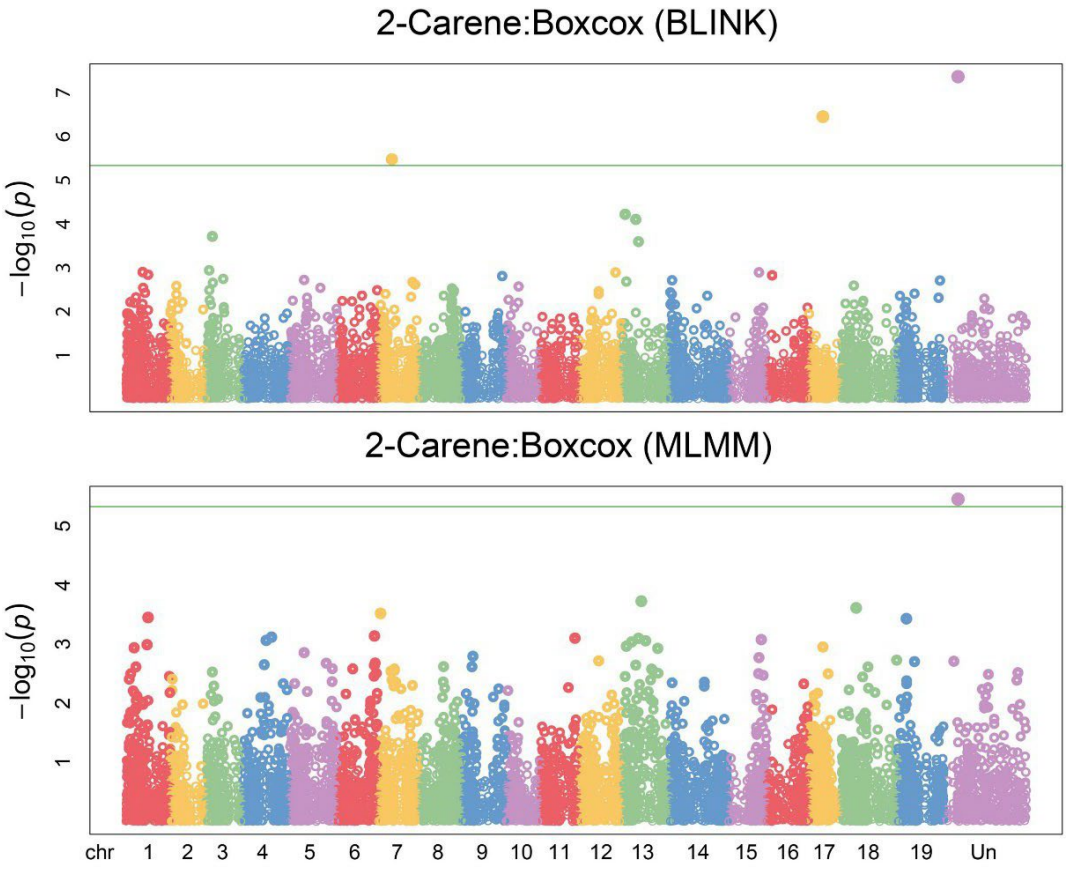


Figure 37 GWAS of 2-carene in the germplasm collection using the BLINK and MLMM models. The horizontal line represents the significance threshold and markers grouped by chromosomes with unmapped markers indicated by 'Un'.

Unfortunately, the most significant SNP was in a region of the genome that has not been mapped to a linkage group in the PN40024 genome build. The SNP fell into the intron of a gene that encodes a transcription factor from the *Indeterminate Domain* subfamily of the Cys2His2 zinc-finger domain family. These transcription factors have roles in various aspects of plant growth and development (Kumar, et al. 2019). This is a credible candidate gene that may be involved in the change of gene expression related to the production of monoterpenes. BLINK also predicted the presence of a QTL on chromosome 17. This was located in the intron of a formin-like gene, which plays a role in actin polymerisation in cells, which is unlikely to be involved in the regulation of monoterpene biosynthesis. Other genes located close to this gene in the genome are also unlikely to be good candidates for controlling the trait. The marker on chromosome 7 falls into an intergenic

region some distance away from a gene and may be indicating that a surrounding gene is important for 2-carene production. There were no obvious candidate genes in this region of chromosome 7.

β-Citronellol

This compound is one of the monoterpenes found in high concentrations in 'Muscat' varieties and is described as having a 'sweet, rose-like' aroma (Elsharif and Buettner 2017). Similar to 2-carene, the mapping was more consistent across different models when non-transformed data were used. Only one QTL was detected using the MLMM model and this was on chromosome 15 (Figure 38). One of the three QTLs detected by the BLINK method was also on chromosome 15, but surprisingly the significant marker was not the same as the one in the MLMM model, but they were only ~2 kb apart in the genome. There are no obvious candidate genes in this region, but there are some transcription factors and genes that encode oxidases that have possible functions in monoterpene metabolism. Two other QTLs were detected using BLINK, one on chromosome 13 and the other on chromosome 17 (Figure 38). The SNP in chromosome is located in the promoter of a gene involved in the demethylation of histones and is part of the photoperiod pathway in other plant species.

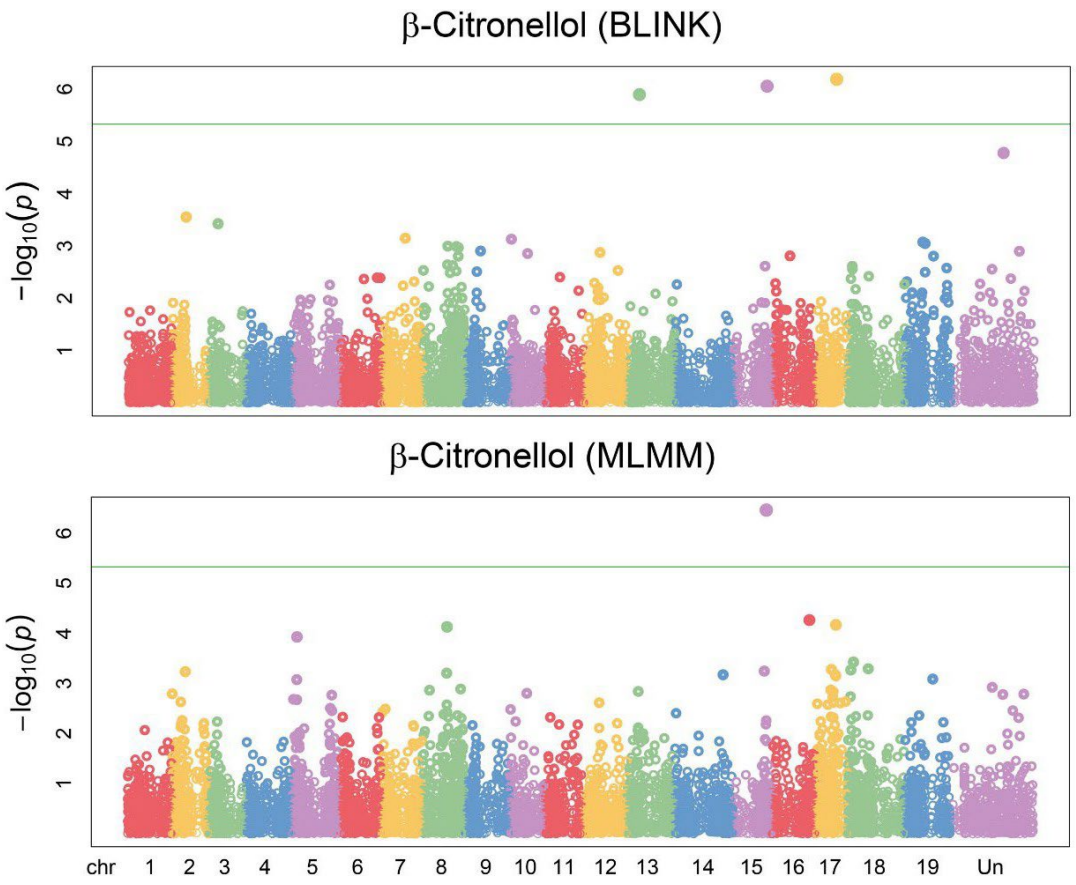


Figure 38 GWAS of β-citronellol in the germplasm collection using the BLINK and MLMM models. The horizontal line represents the significance threshold and markers grouped by chromosomes with unmapped markers indicated by 'Un'.

Other genes around this genomic region are mainly involved in gene regulation and transporters, which could conceivably be involved in the phenotype, but which would need much characterisation to prove function. Likewise, the SNP on chromosome 17 is located in a gene that is involved in the regulation of transcription, in this case in response to dehydration stress, which may relate to the

ripening process in grapevine. Other genes in this region include transporters and pentatricopeptide repeat proteins that alter RNA stability in organelles (Barkan and Small 2014) and could be involved in metabolism that occurs in the chloroplast, the location of monoterpene biosynthesis in grape berries (Luan and Wüst 2002).

p-Cymene

The monoterpene p-cymene has been described as having 'minty', 'solvent' sensory characters (Perez-Cacho and Rouseff 2008) and in this study were present in higher concentrations in the 'Muscat' varieties. The data were not normally distributed and so a Box-Cox transformation was conducted before GWAS, as non-normalised data gave inconsistent results among the models. There also appeared to be a vintage effect, with concentrations being much greater in the 2020 vintage compared to 2019, but GWAS results after rank-based inverse normalisation of the raw data gave similar results to the Box-Cox transformation (Figure 39).

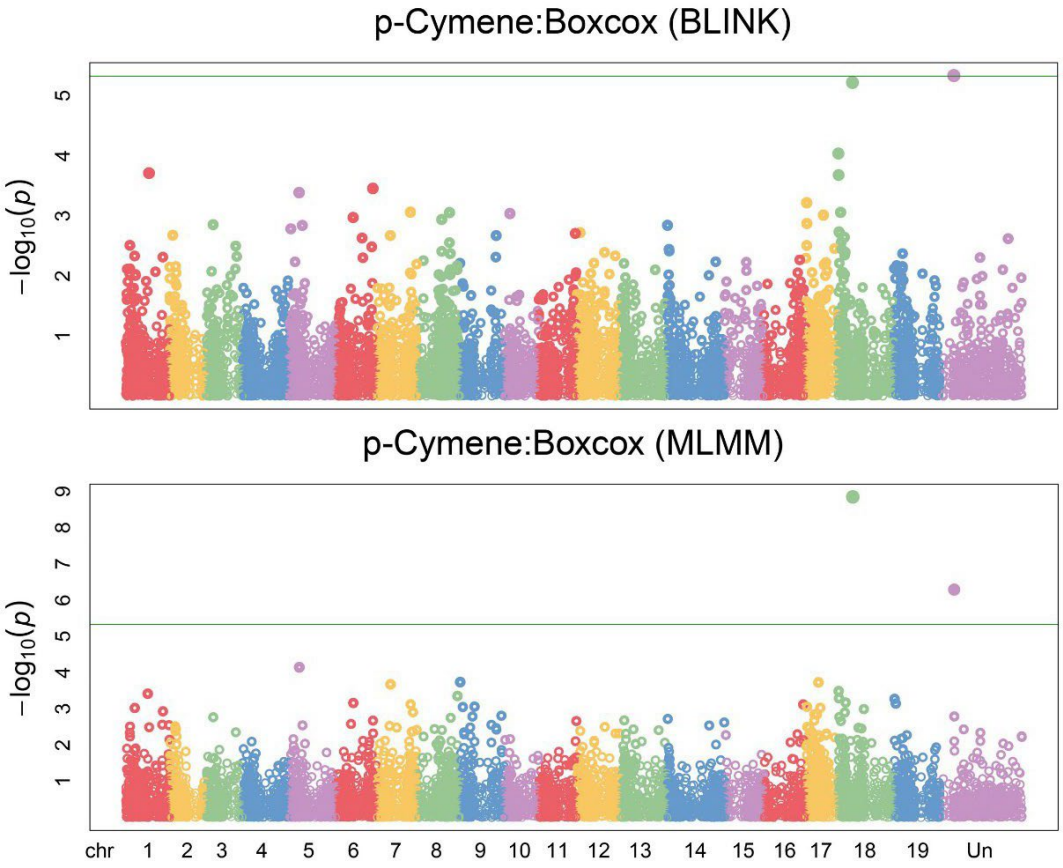


Figure 39 GWAS of p-cymene in the germplasm collection using the BLINK and MLMM models. The horizontal line represents the significance threshold and markers grouped by chromosomes with unmapped markers indicated by 'Un'.

Interestingly, the SNP that was predicted to indicate a QTL in the unmapped markers was the same as the one predicted from the 2-carene GWAS (Figure 37, Figure 39). This was in the gene encoding an *Indeterminate Domain* subfamily protein from the Cys2His2 zinc-finger domain family of transcription factors. This further confirms this gene as a potential candidate gene that controls aspects of monoterpene synthesis, and could be a good marker for the production of these types of

monoterpenes in berries. The SNP on chromosome 18 falls in the intron of a gene whose function is uncertain but contains motifs for binding RNA and other proteins and is potentially involved in regulation of transcription. Other genes in this QTL are mainly related to gene regulation and so there are multiple candidates that may regulate the trait.

β-Damascenone

β-Damascenone was one of our target compounds in the biparental cross mapping (Section 6.1.3) and we were interested to see if those results were reflected in the GWAS. This compound is thought to be very important for increasing the perception of fruity wine characters (Pineau, et al. 2007). After the initial GWAS with non-normalised data, there was some concern that we may have a strong vintage effect, as there was evidence of a strong QTL on the bottom of chromosome 2 (Figure 40). However, the marker was located ~2.3 Mbp from the colour locus and, given the rapid decay of linkage disequilibrium in grapevine, it is potentially not related to the colour/vintage effect.

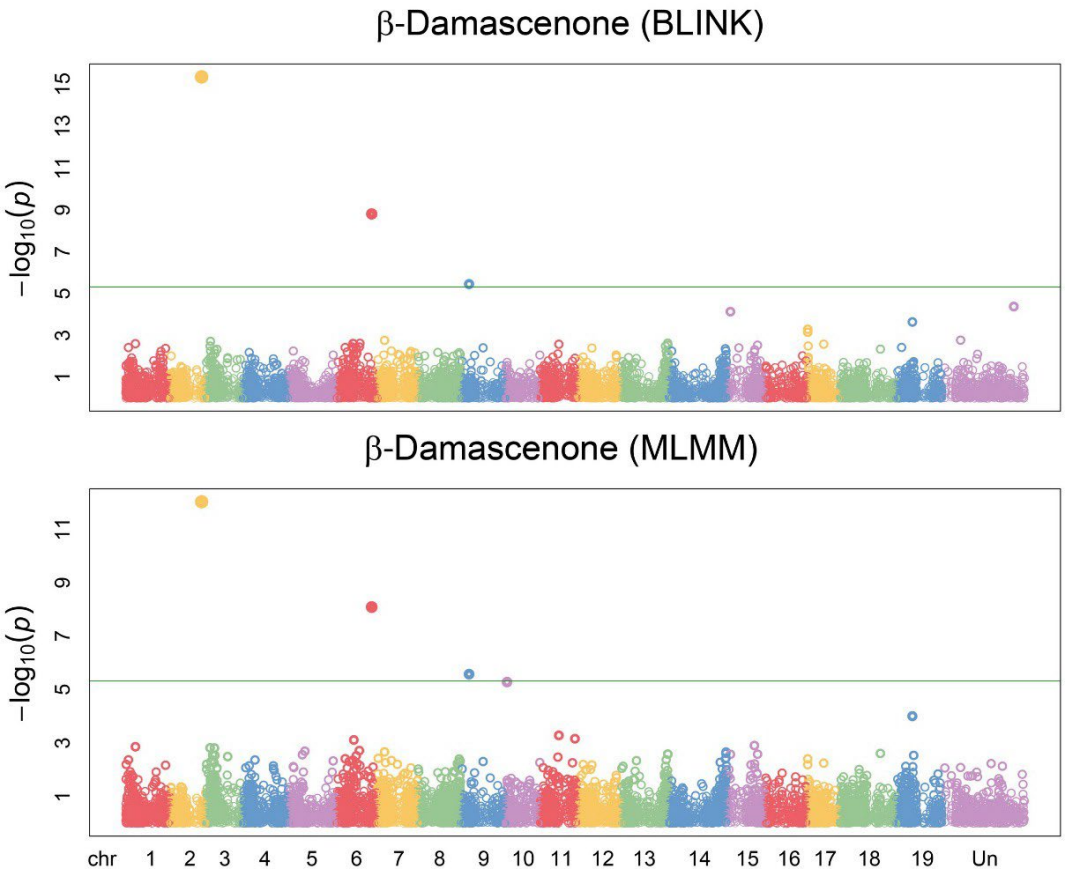


Figure 40 GWAS of β-damascenone in the germplasm collection using the BLINK and MLMM models. The horizontal line represents the significance threshold and markers grouped by chromosomes with unmapped markers indicated by 'Un'.

To confirm the lack of a vintage effect, the raw data from both years of sampling were plotted (Figure 41) and this revealed that while there was some skewing of the values towards greater concentrations in 2020, this was not as evident as that observed for 1-hexanol (Figure 35).

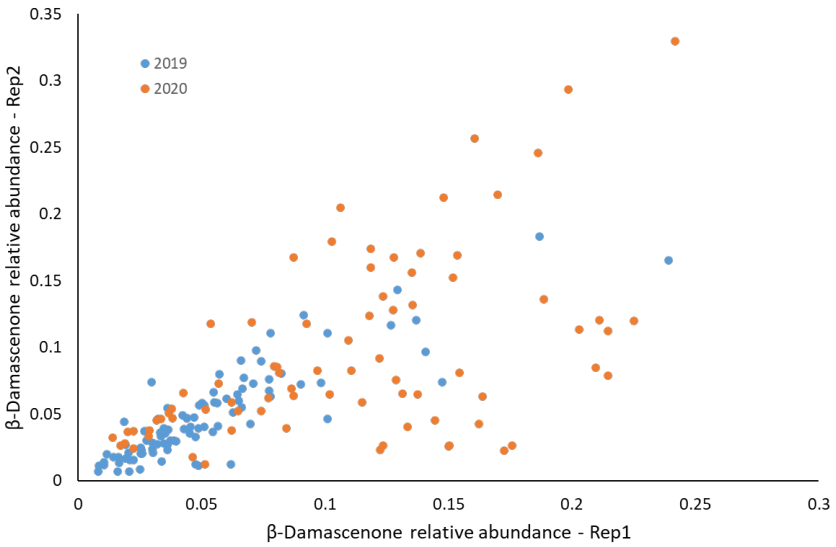


Figure 41 Plot of β-damascenone abundances in replicate wines produced from various varieties from the germplasm collection with blue data points from the 2019 vintage and orange from the 2020 vintage.

The SNP on chromosome 2 that was highly significant in the GWAS falls in a gene that encodes a DNA-binding transcriptional activator involved in nodulation in legume plant species (Yano, et al. 2008) and perhaps mycorrhizal infections in other plant species. While no carotenoid biosynthesis genes were in the vicinity of this SNP, some pentatricopeptide repeat-containing protein genes are nearby in the genome and other potential candidate genes such as a dehydration-responsive protein. Identifying the candidate gene does not preclude this SNP from being a useful marker for this phenotype and so we plotted genotype against the concentration of β-damascenone to gauge the relationship (Figure 42). The SNP would appear to be a good marker for β-damascenone concentration, with the 'TT' genotype being undesirable as it associated with low concentrations.

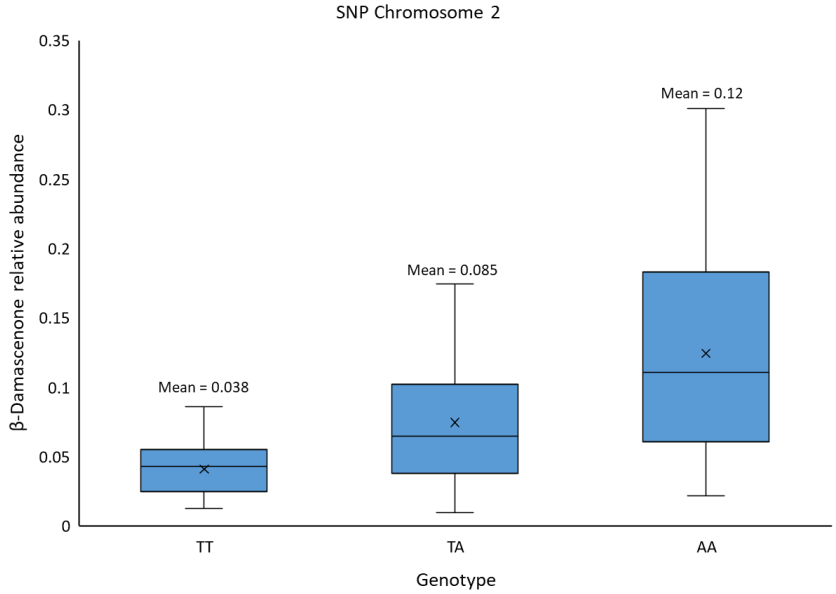


Figure 42 Box plots for the significant β-damascenone SNP marker on chromosome 2 showing the relationship between phenotype and genotype in wine replicate 1 of the variety collection.

However, we do need to be mindful of the potential for this marker to be linked to the colour locus as there is evidence that this part of the genome has reduced haplotype diversity due to selection for colour (Migicovsky, et al. 2017). Therefore, this relationship may be due to a sampling effect or because there is genuinely a QTL for this compound linked to the colour locus. The other two SNPs that were above the significance threshold of the GWAS were located on chromosomes 6 and 9. The one on chromosome 6 was in an intergenic region between a PHD finger transcription factor and a 'harpin-induced' protein. This area of the chromosome is rich in genes that encode regulatory proteins. The SNP on chromosome 9 falls in the intron of a heat shock protein 70 gene. These proteins are important in the response of plants to various stresses and there is evidence that they associate with carotenoid biosynthesis genes (Bonk, et al. 1996). Silencing of certain members of this gene family can result in reduced carotenoid production (Anaraki, et al. 2018). Other potential candidate genes in this region include a pentatricopeptide-repeat-containing protein, transcription factors and a glycosyl transferase.

Diethyl pentanedioate

Diethyl pentanedioate, otherwise known as diethyl glutarate, is described as having an 'estery, floral' sensory character, however details about its contribution to wine sensory characters are scant. Even though this compound is synthesised by the yeast during fermentation, the correlation between winemaking replicates (Table 11) was such that it suggests this synthesis is regulated by a grape-derived component. GWAS was conducted with non-normalised data and was inconsistent across the models used. One SNP was significant in the GWAS with both BLINK and MLMM and this was located on chromosome 18 (Figure 43).

This SNP was located in a gene encoding a pseudouridine synthase which is involved in nucleic acid metabolism but can feed into other pathways from which glutarate could be formed. The QTL on chromosome 14 was highly significant using the BLINK model and was located in a transcription factor known to be involved in the response to ABA in plants (Bi, et al. 2017). However, more biochemistry to understand the link between berry composition and the production of this compound by yeast is required before a full list of candidate genes in these regions can be compiled, and it is doubtful if this compound is of high priority at this time.

Diethyl succinate

This compound in ubiquitous in wine and is formed by the yeast during fermentation. It is described in sensory terms as having a 'wine-like' character (Cortés-Diéguez, et al. 2015), which suggests it is important for the base flavour of wine upon which other odorants add nuance depending on variables such as variety and region. The only GWAS model that detected QTLs above the significance threshold was BLINK (Figure 44). Three SNPs, one each on chromosomes 5, 6 and 14 were associated with the diethyl succinate concentration in the wines. The one on chromosome 6 falls in a gene that encodes a proton-dependent oligopeptide transport (POT) family protein involved in transporting peptides, amino acids and other compounds in plants, animals and bacteria. On chromosome 5, the significant SNP is in a kinesin phragmoplast orienting gene which produces a protein involved in microtubule organisation, while the one on chromosome 14 is in a gene annotated as 'unknown' but could be a regulator of potassium channels. As with diethyl

pentanedioate, more biochemical knowledge about the link between grape composition and diethyl succinate in wine is required before better predictions of potential candidate genes can be made in the mapped areas.

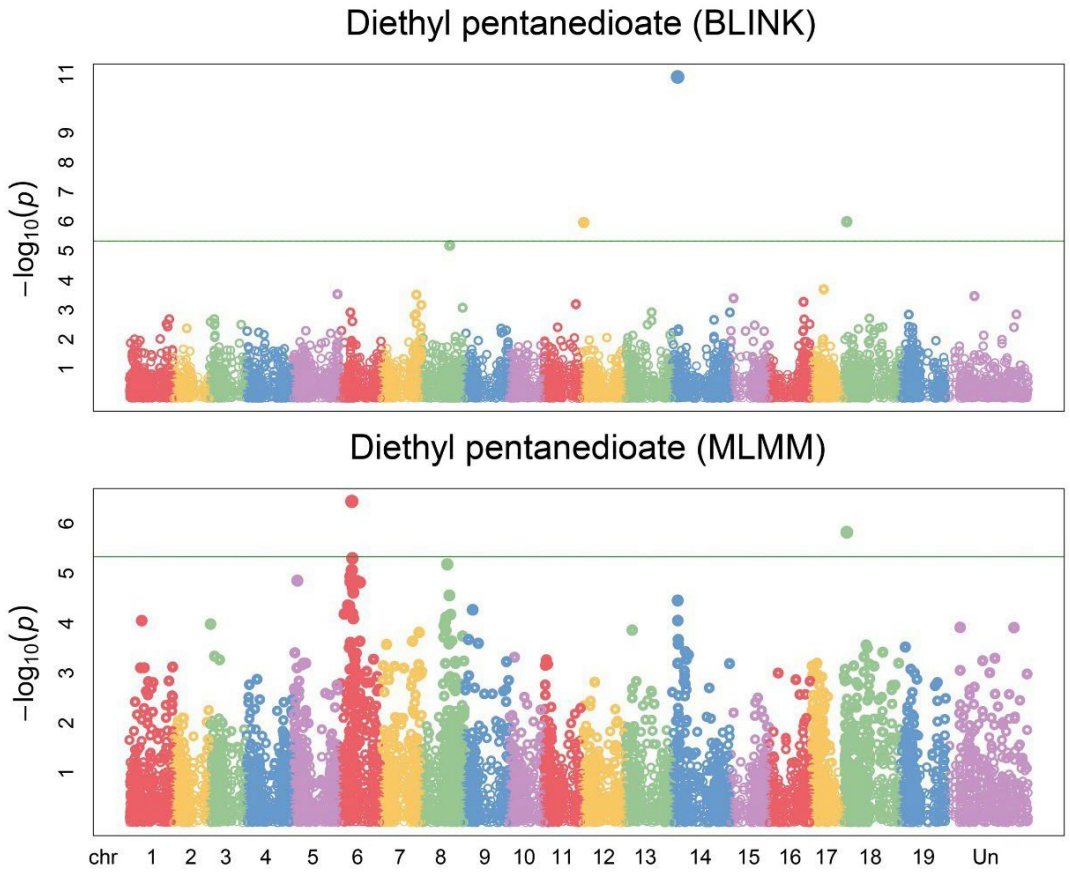


Figure 43 GWAS of diethyl pentanedioate in the germplasm collection using the BLINK and MLMM models. The horizontal line represents the significance threshold and markers grouped by chromosomes with unmapped markers indicated by 'Un'.

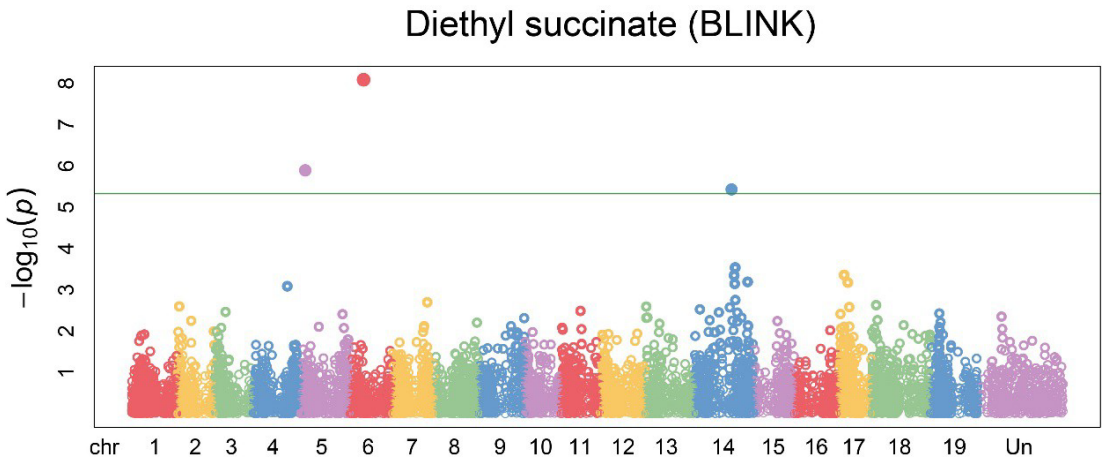


Figure 44 GWAS of diethyl succinate in the germplasm collection using the BLINK and MLMM models. The horizontal line represents the significance threshold and markers grouped by chromosomes with unmapped markers indicated by 'Un'.

3-Ethoxy-1-propanol

This compound has often been identified in wine volatile compound metabolite studies and significantly reduces in abundance during fruit development (Boss, et al. 2018). This compound is listed as having a 'mild fruit aroma', but it has been correlated with the 'blackcurrant' descriptor in some Cabernet Sauvignon wines (Tao and Zhang 2010). Unfortunately, none of the GWAS models predicted any significant QTLs for this compound. It is formed during fermentation, and it could be that this variable outweighs any variation in fruit composition, although the correlation between winemaking replicates would suggest that this is not the case (Table 11). Perhaps other variables such as the ripeness of the fruit are masking any effect due to genetic differences in the varieties.

Ethyl 3-hydroxybutanoate

This compound has an interesting combination of taste descriptors with its odour being described as 'fresh', 'fruity' and 'grape', whereas the taste descriptors include 'green' in addition to the listed odour descriptors. Ethyl 3-hydroxybutanoate has been shown to enhance the perception of wine 'fruity' aromas even if present in sub-threshold concentrations (Lytra, et al. 2015). There were no significant QTLs detected using any of the GWAS models with raw or normalised data. This suggests that non-genetic variables have a greater influence on the phenotype than genotype, or that three are no markers linked to this phenotype in the SNP dataset.

Ethyl 3-methylbutyl succinate

Otherwise known as ethyl isopentyl succinate, this compound has been described as having a 'pungent' sensory character (Yu, et al. 2019). Like the two previous compounds, there were no significant QTLs detected by GWAS. This may reflect the lack of a SNP in the dataset close to a causal gene or the preponderance of environmental variables masking any difference due to genetic differences between the varieties.

Ethyl 4-hydroxybutanoate

There is not a lot of information in the literature about this compound and it has been described as having a 'caramel-like' sensory character. The GWAS revealed the presence of a single significant QTL using the BLINK model with raw (Figure 45) and rank-based inverse normalisation but this was not replicated with other models, although in some cases the SNP did approach significance without being higher than the threshold. The genotypic difference was in an intergenic region on chromosome 13 flanked by two genes of unknown function. A potential precursor in the fruit is the corresponding acid 4-hydroxybutanoate which is placed in pathways associated with butanoate metabolism which links to many other pathways such as the citrate cycle and glutamate, arginine and proline metabolism. Some genes around this SNP encode genes with functions in amino acid transport and metabolism and are potential candidate genes for further study if this compound is considered a priority for future work.

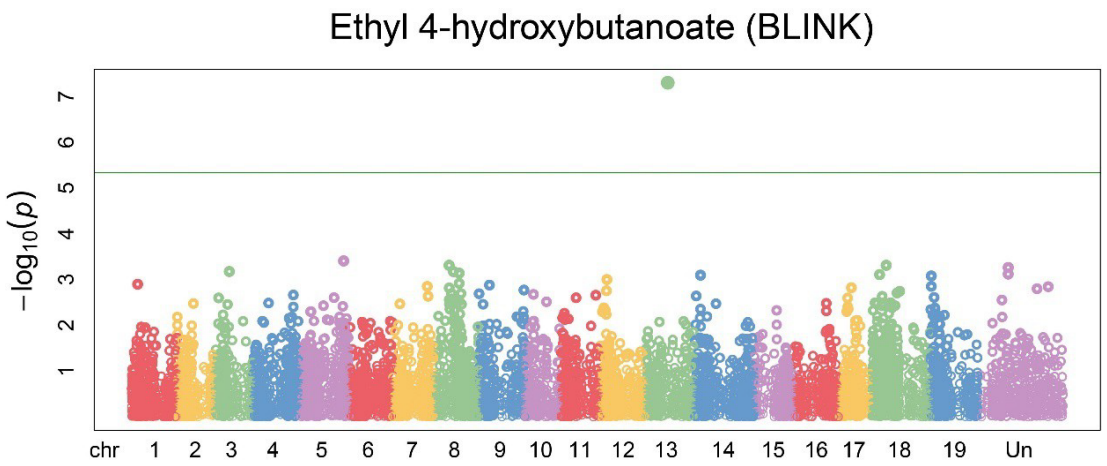


Figure 45 GWAS of ethyl 4-hydroxybutanoate in the germplasm collection using the BLINK model. The horizontal line represents the significance threshold and markers grouped by chromosomes with unmapped markers indicated by 'Un'.

Ethyl 9-decenoate

Ethyl 9-decenoate is another compound that is ubiquitous in wine volatile compound analyses and often shows variation in vineyard trials. It is thought to have 'fruity' and/or 'fatty' sensory influence on wines. No QTLs were detected using any of the GWAS models and this could be due to the reasons already listed for other compounds above where no significant SNPs were identified.

Ethyl isovalerate

This compound, also called ethyl 3-methylbutanoate, has 'fruity' sensory attributes, but has also been described as having a 'lemon' and 'anise' aroma (Gómez-Míguez, et al. 2007). Unfortunately, none of the GWAS models detected any significant QTLs for this compound.

Ethyl octanoate

Ethyl octanoate is one of the most abundant esters in wines and has a pleasant 'cooked fruit' sensory character (Fang and Qian 2005). GWAS using the raw data was unsuccessful and a comparison of the two vintages suggested that the abundances in the 2019 wines were generally lower than those in 2020, although the colour marker was not detected as a significant QTL. Nevertheless, we repeated the analysis with the rank-based inverse normal transformation and some significant QTLs were detected (Figure 46). With the BLINK model, three significant QTLs were detected corresponding to SNPs present on chromosomes 4, 8 and 18. Only the SNP on chromosome 18 was significant using the less powerful MLMM method. This SNP was located in an intergenic region, upstream of a gene of unknown function. There are several genes in the vicinity that are involved in amino acid metabolism, which could play a role in determining the amount of nitrogen present in the fermentation or influence the abundance of CoA precursors in the must. The SNP detected by BLINK on chromosome 4 falls in a transmembrane protein of unknown function and no obvious candidate genes are in the vicinity of this gene without knowing more about the grape biochemistry that influences the production of the ethyl esters. The location of the significant SNP on

chromosome 8 coincides with an 1,4-alpha-glucan branching enzyme gene that is involved in polysaccharide biosynthesis. A gene of interest close to this location is a lipase gene, as it is known that triacyl glycerides can influence ethyl ester production in fermentations (Blackford, et al. 2018).

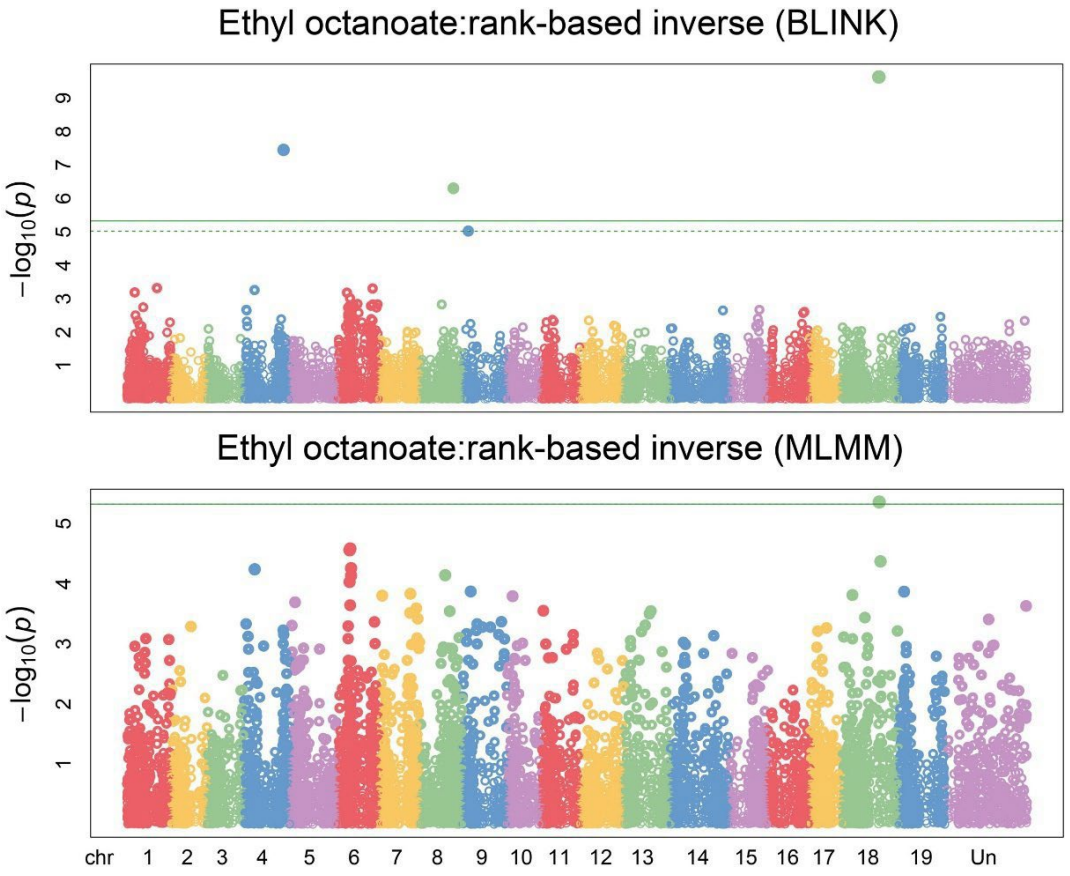


Figure 46 GWAS of ethyl octanoate in the germplasm collection using the BLINK and MLMM models. The horizontal lines represent the significance thresholds (solid line = p<0.01, dotted line = p<0.05) and markers grouped by chromosomes with unmapped markers indicated by 'Un'.

1-Hexanol

1-Hexanol is derived from the degradation of linoleic acid via the lipoxygenase pathway (Dunlevy, et al. 2009), which is activated when plant tissue is wounded. The compound is best known for contributing to the smell of cut grass and so contributes to the 'green' or 'stemmy' sensory character in wine. As discussed above, the production of 1-hexanol showed a strong vintage effect and as such, multiple normalisation transformations were conducted, but they either failed to detect QTLs or always had SNPs in the colour locus indicating this region as a significant QTL. Comparing the various models under different data transformations, the most consistent results occurred when the Box-Cox transformation was used. While the colour locus on chromosome 2 was significant, there was also another QTL detected on chromosome 5 in both BLINK and a linear model method (in this case the compressed mixed linear model (CMLM)) (Figure 47). The SNP on chromosome 5 was present in the intron of a gene encoding a zinc-finger DNA-binding protein. There are some interesting nearby genes that could play a role in the production of 1-hexanol, including a number of alcohol dehydrogenase genes and fatty acid desaturases. The function of these genes should be investigated

as a possible means of reducing 1-hexanol concentrations in wine.

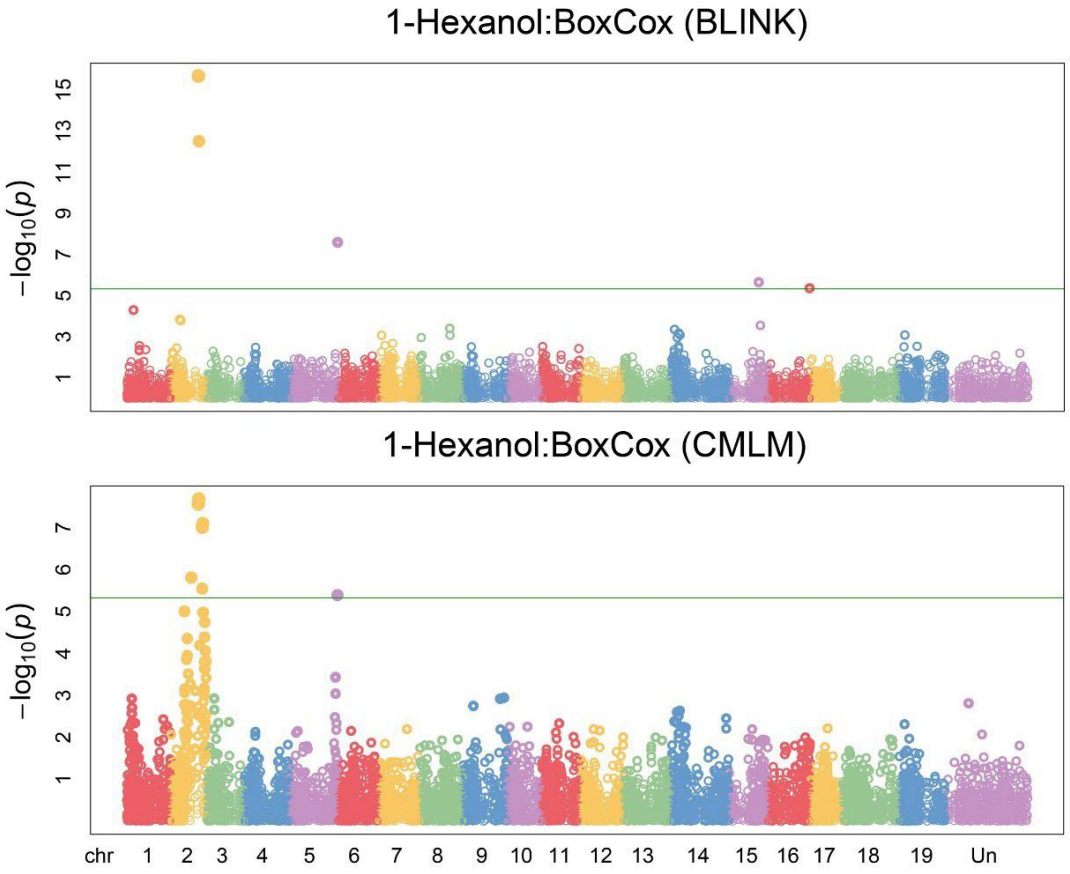


Figure 47 GWAS of 1-hexanol in the germplasm collection using the BLINK and CLMM models. The horizontal line represents the significance threshold (p<0.01) and markers grouped by chromosomes with unmapped markers indicated by 'Un'.

(*E*)-3-Hexenol

The biosynthesis of this compound is still unclear, although it has been predicted that it is an intermediate in the spontaneous isomerisation of (*Z*)-3-hexenal to (*E*)-2-hexenal (Hatanaka 1993). It has similar sensory characteristics to 1-hexanol, contributing to 'green' and 'grassy' sensory characters. The results of the GWAS mapping were very similar to those obtained with 1-hexanol (Figure 48). For this compound only one of the two SNPs flanking the colour locus was significant, and the same SNP on chromosome 5 was significant in many of the models. It was possible that this marker was associated with a differential flux down either the linoleic or linolenic part of the lipoxygenase pathway (Figure 5) which result in 1-hexanol or unsaturated C_6 compound production respectively. To test this assumption, the genotype:phenotype relationship was plotted using box plots for both of these compounds at this SNP (Figure 49). It was evident that for both compounds the 'AA' genotype corresponded to those varieties with lower concentrations of this compound in wines. So, any candidate gene in this should influence the flux down both the saturated and unsaturated branches of the pathway and would be good targets for manipulating the 'grassy' character in wines. The desaturase genes could be involved in this regulation if they catalyse the

desaturation at the 9 or 12 position of the 18-carbon fatty acid as these are present in both substrates required for the production of both of these C_6 compounds. This should be tested in future work.

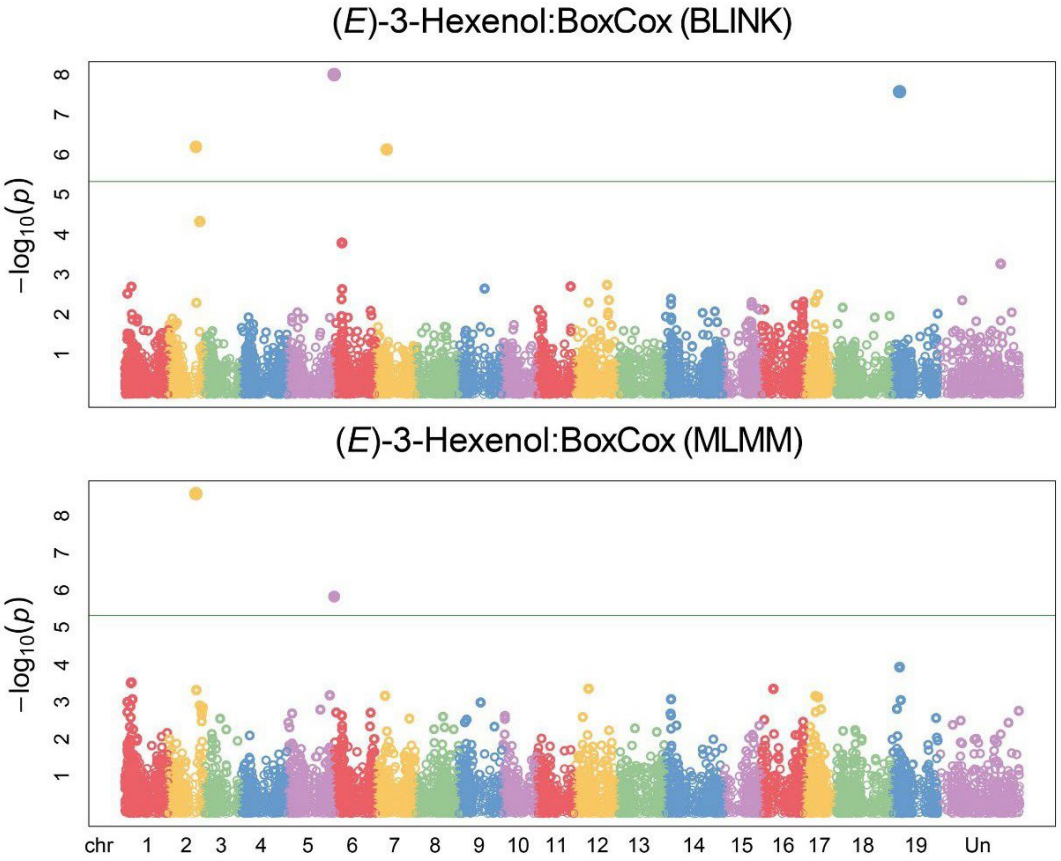


Figure 48 GWAS of (*E*)-3-hexenol in the germplasm collection using the BLINK and MLMM models. The horizontal line represents the significance threshold (p<0.01) and markers grouped by chromosomes with unmapped markers indicated by 'Un'.

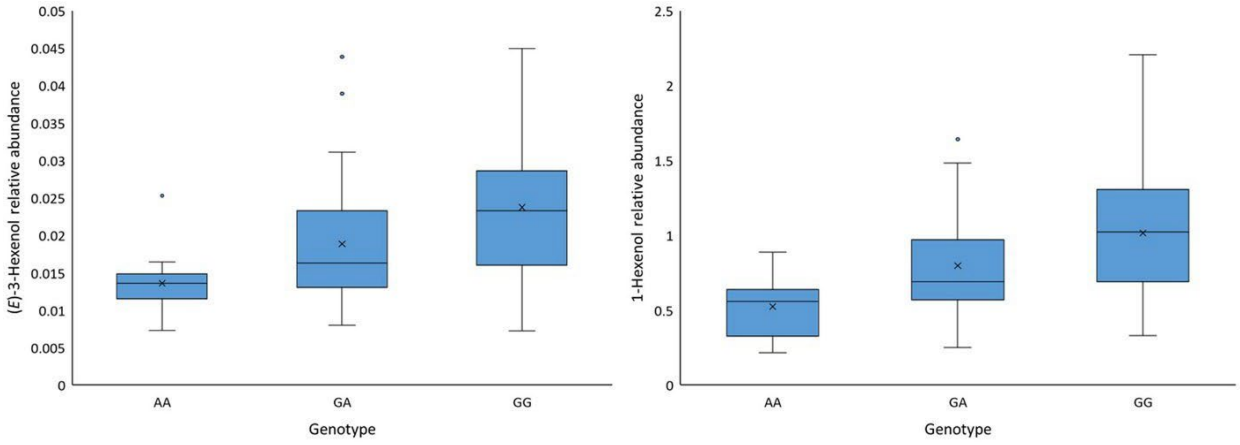


Figure 49 Box plots for the significant SNP marker for (*E*)-3-hexenol and 1-hexanol concentrations on chromosome 5 showing the relationship between phenotype and genotype in wine replicate 1 of the variety collection.

(*Z*)-3-Hexenol

This compound was a primary target of this project because of its link with 'green' and unripe sensory characters in wine as outlined in Section 6.1.3. We were also interested to see if the GWAS analysis confirmed the findings of the biparental crosses, which suggested that there was a region on chromosome 12 that was important for the concentration of (Z)-3-hexenol found in the wine. The most consistent GWAS modelling was observed when the data had been normalised using the Box-Cox transformation. The analysis suggests that the region on chromosome 12 is indeed important for this trait with two markers in the vicinity of the region identified in the biparental crosses also being significant in the BLINK GWAS analysis and one when the MLMM model was used (Figure 50). The SNP significant in both is located at ~17.1 Mbp from the top of chromosome 12 which is just upstream of the QTL region identified in the Mv crosses. The other SNP on chromosome 12 in the BLINK GWAS is at ~18.6 Mbp, which is within the QTL previously identified.

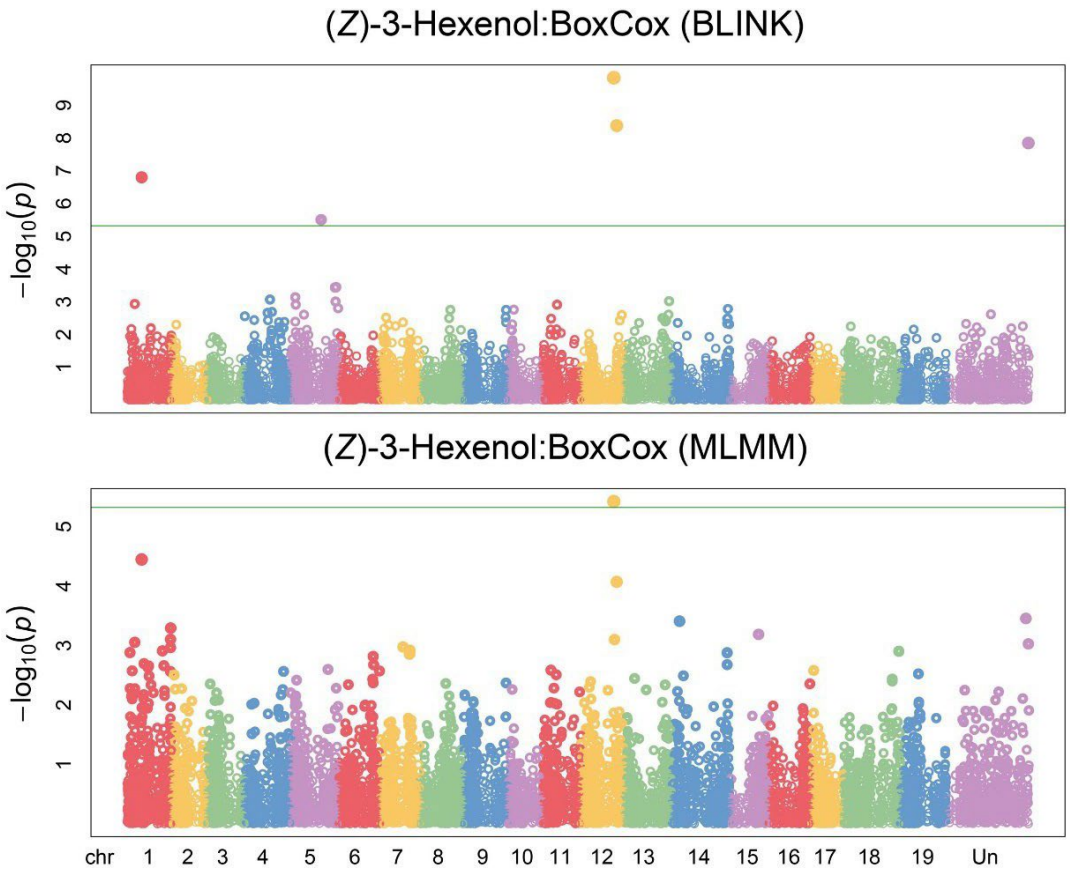


Figure 50 GWAS of (*Z*)-3-hexenol in the germplasm collection using the BLINK and MLMM models. The horizontal line represents the significance threshold (p<0.01) and markers grouped by chromosomes with unmapped markers indicated by 'Un'.

We decided to investigate the nature of these markers on chromosome 12 further to try and understand the phenotype:genotype relationship better. Box plots of the (Z)-3-hexenol concentrations measured in the wines from varieties with different genotypes were generated for either SNP on their own and in combination (Figure 51). With regards to the marker at 17.1 Mbp, the presence of a 'C' allele was enough to predict that the wine would have relatively less (*Z*)-3-hexenol than those with at 'GG' genotype. Similarly, a 'C' at 18.6 Mbp also resulted in less of this

compound in wines compared to the 'TT' genotype in this case. When both markers are considered, it is those with the combination of 'TT GG' that are likely to have greater (Z)-3-hexenol concentrations in wine. This suggests that the trait is recessive because the presence of a 'C' allele at either locus reduced the (Z)-3-hexenol concentration. A combination of these two markers would work well as a means of predicting the amount of (Z)-3-hexenol we would expect to get in wine made from a new variety.

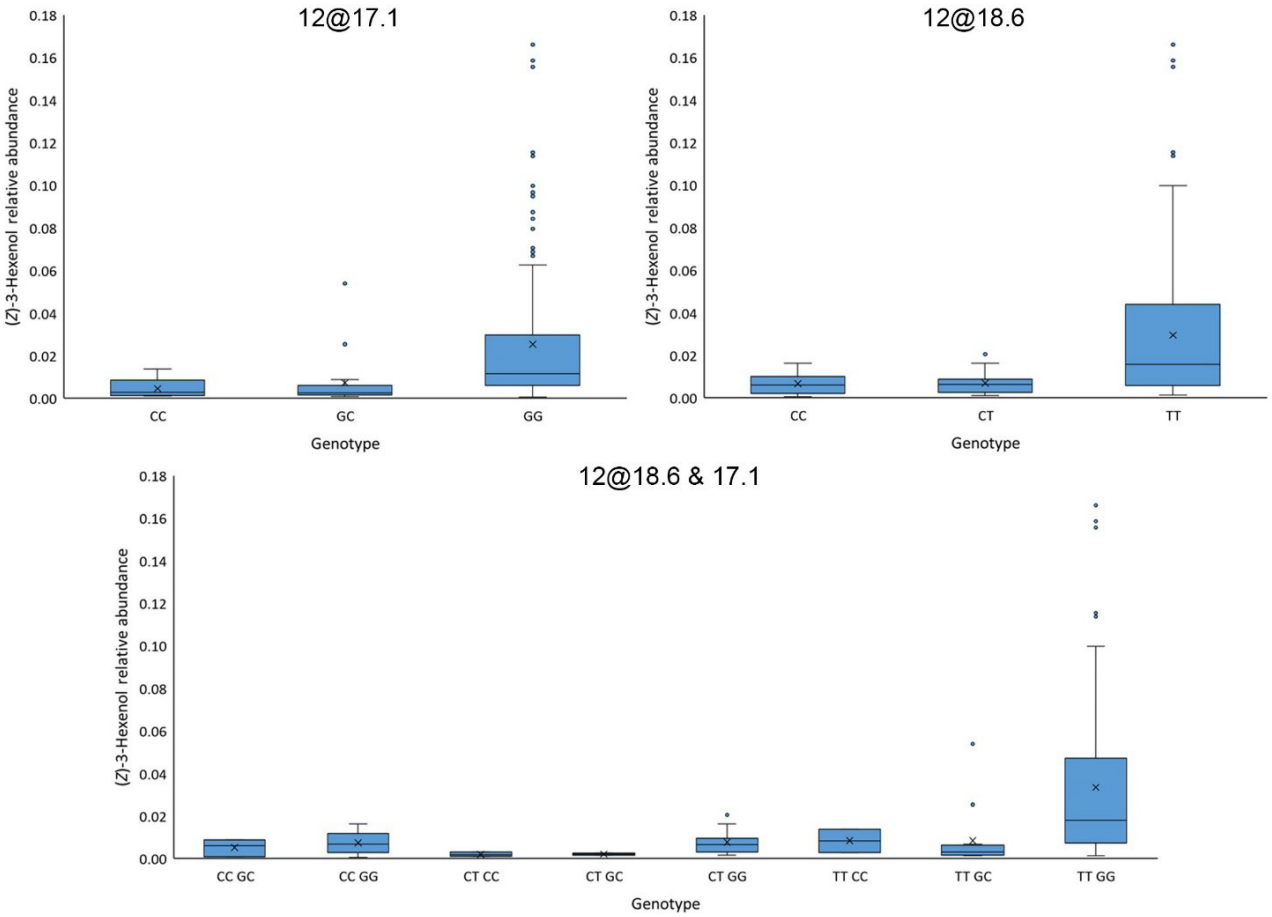


Figure 51 Box plots for the significant SNP markers for (*Z*)-3-hexenol concentrations on chromosome 12 showing the relationship between phenotype and genotype in wine replicate 1 of the variety collection for the individual markers and when used in combination.

Hotrienol

There is surprisingly little information on the sensory properties of hotrienol even though it is often reported to be present in GC-MS analyses of wine. In studies on other food products, it is described as having a floral character and is predominant in honey made from heather (Rodríguez-Flores, et al. 2021). The GWAS detected one QTL consistently above the significance threshold in more than one model (Figure 52) and this was located on chromosome 8. The SNP falls in a gene encoding a protein of unknown function, which shows little homology to genes in other plant species. There are several genes in the surrounding region of the genome with potential roles in biosynthesis of hotrienol including transcription factors, pentatricopeptide repeat-containing proteins and a hydrolase that would be a primary target for future work if this compound was a high priority.

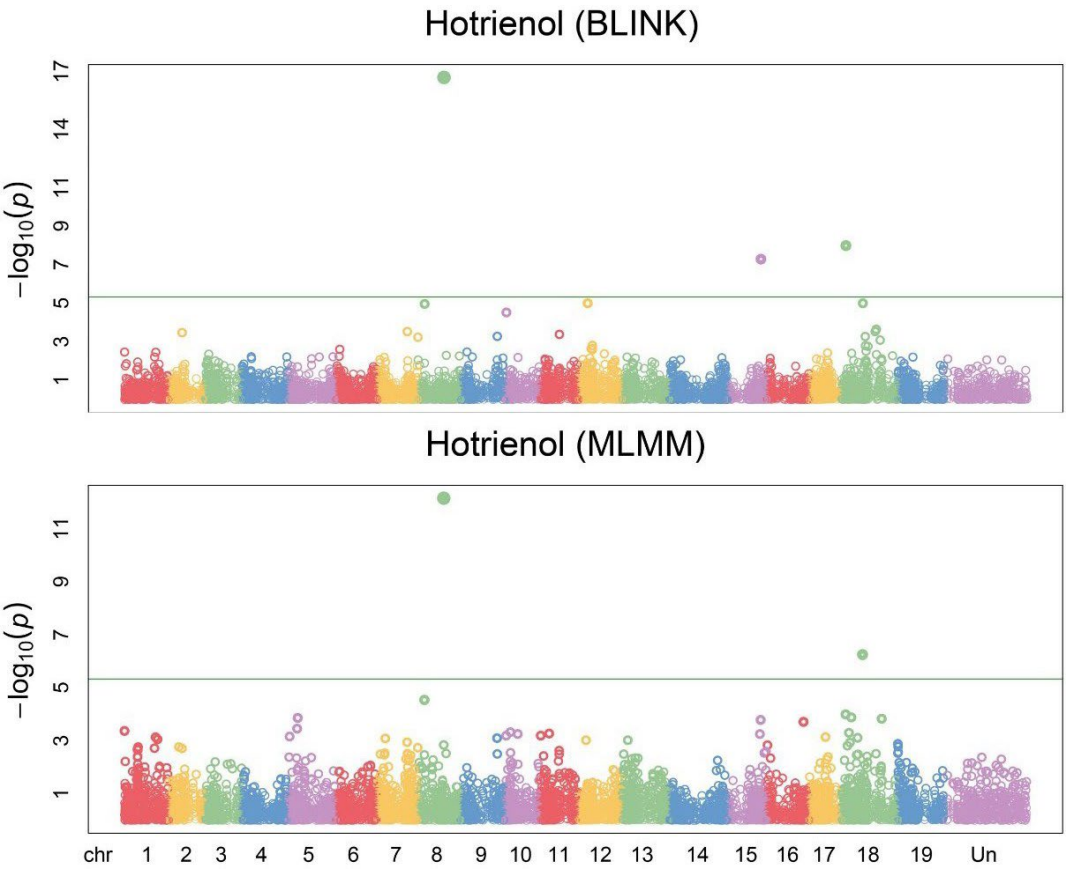


Figure 52 GWAS of hotrienol in the germplasm collection using the BLINK and MLMM models. The horizontal line represents the significance threshold (p<0.01) and markers grouped by chromosomes with unmapped markers indicated by 'Un'.

Isoamyl alcohol

Isoamyl alcohol is one of the 'higher' or 'fusel' alcohols that are produced by yeast during alcoholic fermentation. It is a product of leucine metabolism and is the major fusel alcohol found in wines. Isoamyl alcohol in combination with isobutyl alcohol has been shown to increase the 'spirit-like/solvent' character in wine while reducing 'red fruit' and 'coconut/vanilla' sensory attributes (De-La-Fuente-Blanco, et al. 2016). The production of this compound showed a profound vintage effect, especially in one of the winemaking replicates and so the GWAS was conducted using a rank- based inverse transformation of the dataset. There were no significant QTLs identified using any of the GWAS models.

Linalool

Linalool is a monoterpene found in high concentrations in 'Muscat' varieties. There is a marker already available for the high production of this compound and therefore distinguishing the 'Muscat' varieties from other aromatic and non-aromatic varieties (Emanuelli, et al. 2014). The biochemistry and sensory attributes are already discussed in section 6.1.3, but we were interested to see if GWAS on a wide selection of varieties would reveal any other QTLs that may be associated with moderate linalool production and perhaps confirm our prediction of an important locus on chromosome 1 (section 6.1.3). The data showed a vintage effect and as the GWAS was conducted after a rank-based inverse transformation, the GWAS results were not consistent across the different models used, but

we have shown here the results using the BLINK model (Figure 53), as this has often been the most informative with the other compounds studied.

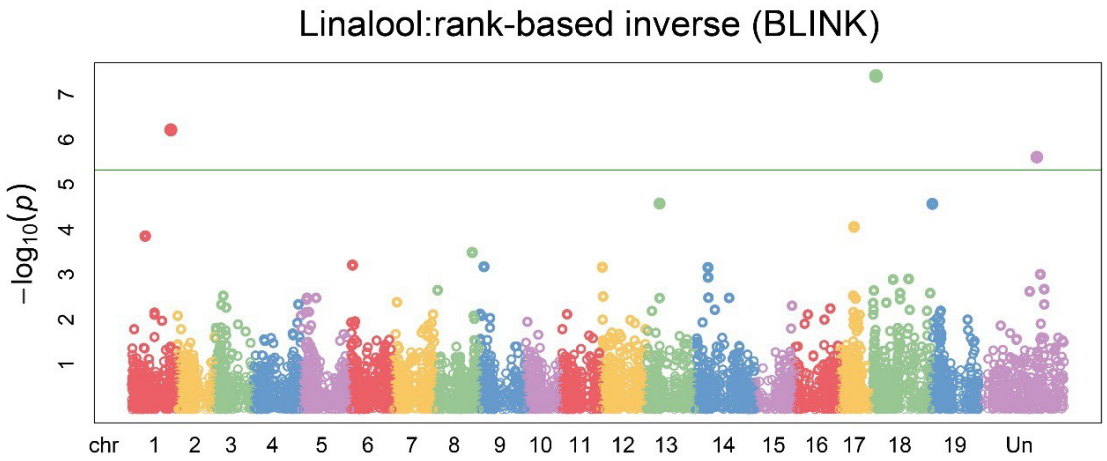


Figure 53 GWAS of linalool in the germplasm collection using the BLINK model. The horizontal line represents the significance threshold (p<0.01) and markers grouped by chromosomes with unmapped markers indicated by 'Un'.

Three SNPs were above the significance threshold (Figure 53). The one on chromosome 1 was located towards the bottom of the chromosome and so did not overlap with the putative QTL found in the biparental mapping (section 6.1.3). The SNP falls in a gene encoding a protein involved in lipid anabolism in the chloroplast, although we are not aware of any links between lipid and monoterpene metabolism. This genome region is rich in transporter and calcium-binding proteins and no obvious monoterpene-related genes were evident. On chromosome 18, the significant SNP fell in an intergenic region. Again, no obvious monoterpene-linked genes were found in this region of the genome, but a regulator of cytokinin biosynthesis was nearby and may be a link between late stages of ripening (Böttcher, et al. 2015) and monoterpene accumulation (Boss, et al. 2014). The significant unmapped marker was located in a region with no clear candidate genes for monoterpene production.

4-Methyl-1-pentanol

This compound is reported to have a 'nutty' aroma and is produced during fermentation. There is little information about the metabolic pathway(s) leading to 4-methyl-1-pentanol production. GWAS failed to find consistent QTLs using the various models, and different data transformations gave inconsistent results as well. As such, we did not consider that any of the predicted significant QTLs were reliable.

3-(Methylthio)-1-proponal

Otherwise known as methional, this compound contributes 'raw potato' attributes to food products. It is derived from the amino acid methionine and is one of the 'higher' or 'fusel' alcohols produced

during fermentation. 3-(Methylthio)-1-propanol is not thought to play a major role on wine sensory attributes directly (De-La-Fuente-Blanco, et al. 2016), but when converted to its corresponding aldehyde methional during wine aging or oxidation, this can then impart a 'cooked vegetable' sensory character on the wine (Escudero, et al. 2000). GWAS identified two genome regions with potential QTLs (Figure 54). The SNP on chromosome 14 was only significant in the GWAS using the BLINK model and fell in a gene encoding an aminopeptidase, which cleaves amino acids from the N-terminus of polypeptides. Given that the N-terminal residue on a protein is most commonly a methionine, this is a potential candidate gene for the accumulation of the methionine metabolite methional. The SNP on chromosome 7 was significant at the 0.01 level in the MLMM analysis and at the 0.05 level using the BLINK model (Figure 54). This SNP is located in a gene that encodes a protein of unknown function, which is bordered by a pair of ammonium transporters that could influence total nitrogen and amino acid concentrations in the plant.

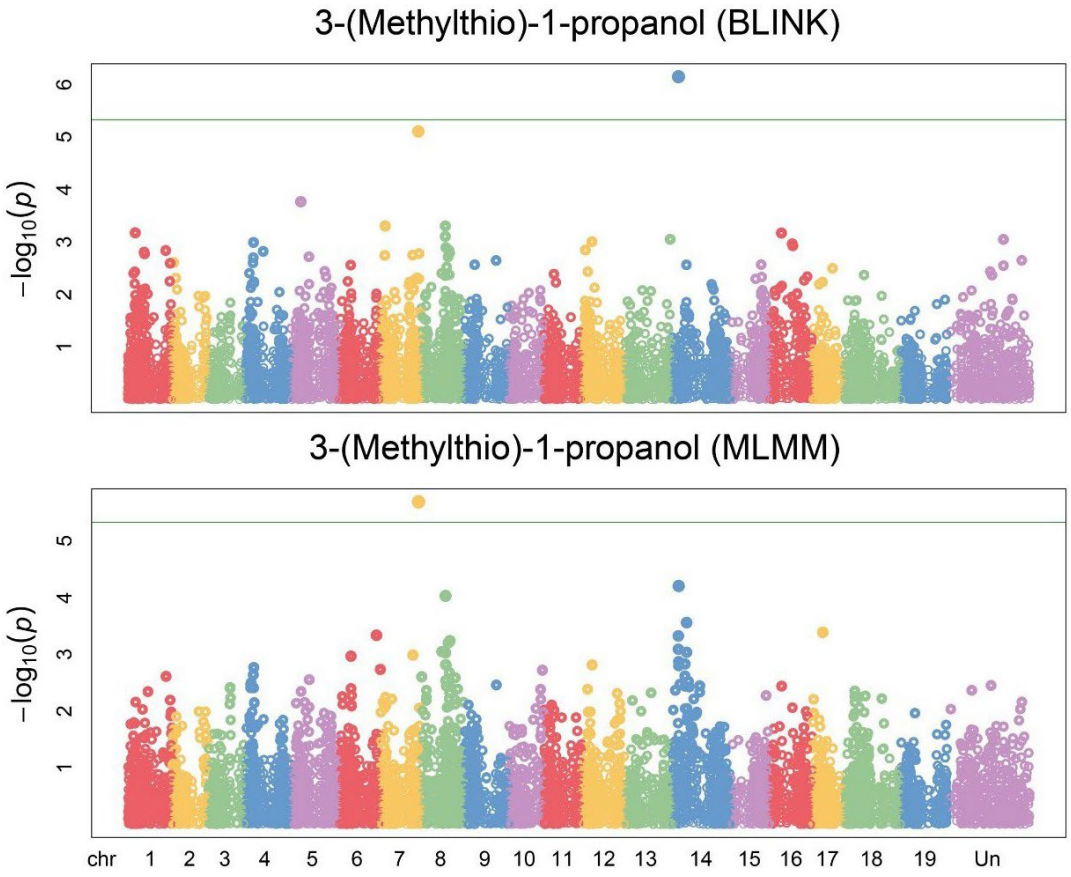


Figure 54 GWAS of 3-(methylthio)-1-propanol in the germplasm collection using the BLINK and MLMM models. The horizontal line represents the significance threshold (p<0.01) and markers grouped by chromosomes with unmapped markers indicated by 'Un'.

2-Nonanol

The sensory descriptors for 2-nonanol are quite broad, ranging from 'waxy' and 'green' to 'citrus orange' and 'cheesy with fruit undertones'. Its biosynthesis is not clear, although the most obvious

pathway involves the reduction of the methyl ketone 2-nonanone. The GWAS suggested that the production of 2-nonanol is multigenic with several SNPS from various chromosomes above the 0.01 significance threshold (Figure 55). However, there was a limited number of these SNPs that were found in more than one GWAS model. One was located on chromosome 17, in a gene encoding a G-protein signalling receptor and no obvious candidate genes were identified in the vicinity of this SNP. Another marker on chromosome 11 is in an intergenic region and several protein families are found around this part of the genome that could be involved in the phenotype including alcohol dehydrogenases, glycosyl transferases and enzymes involved in cofactor synthesis. Finally, there was also a significant SNP with both models on chromosome 7 and it falls in a zinc-finger DNA-binding transcription factor. No clear candidate genes were identified in this area of the genome. This compound is probably a lower priority target and given the multigenic nature of the trait it would be difficult to design markers to predict its accumulation in new varieties.

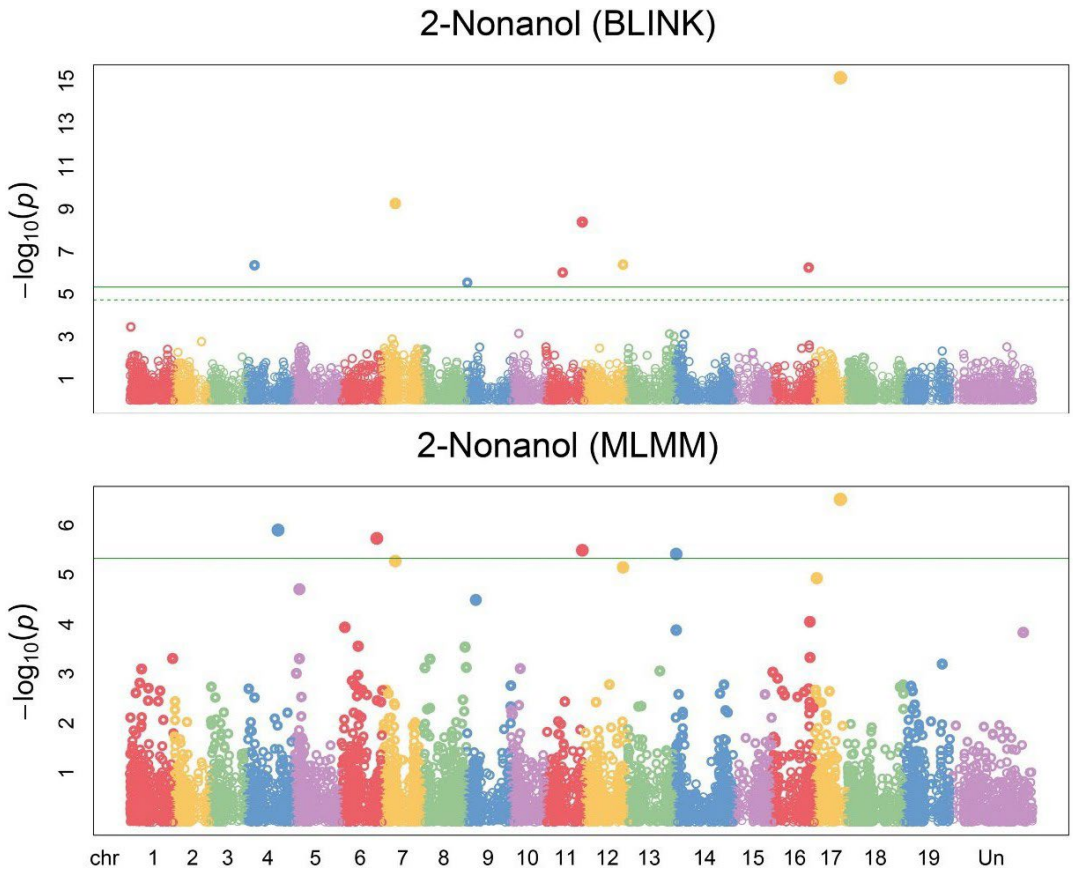


Figure 55 GWAS of 2-nonanol in the germplasm collection using the BLINK and MLMM models. The horizontal line represents the significance threshold (p<0.01) and markers grouped by chromosomes with unmapped markers indicated by 'Un'.

2-Nonanone

The methyl ketone 2-nonanone could arise through lipid beta-oxidation or biosynthesis as 3-keto fatty acids, when released from CoA or acyl carrier protein, can spontaneously decarboxylate to methyl ketones. The sensory descriptors for 2-nonanone are, like 2-nonanol, relatively confusing with terms ranging from 'green fruity' and 'weedy' to 'blue cheese' and 'mushroom'. The compound

appears to change in character markedly depending on the concentration in the food product and the matrix. There was a distinct vintage effect in the concentrations of 2-nonanone with abundances in the wines produced in 2020 higher than those in 2019. As a consequence, the rank-based inverse transformation was applied to the data before GWAS. Unfortunately, there were no significant SNPs detected with any of the models for this compound.

1-Octanol

1-Octanol is described as having 'waxy', 'green' and 'citrus' sensory attributes. The origin of this compound in grapes is not clear but could originate from lipid degradation and/or the reduction of octanoic acid. GWAS analysis of the 1-octanol concentrations from the varietal wines did not detect any significant QTLs.

Phenylethyl acetate

The sensory attributes and biosynthetic origins of phenylethyl acetate have already been listed in section 6.1.3, but briefly it has a 'fruity' or 'rose' sensory percept and is generated by yeast during fermentation. The biparental crosses found a significant QTL at the top of chromosome 18 that was associated with the production of a number of acetate esters. The GWAS of phenylethyl acetate concentrations in the variety collection was first conducted with raw data as this gave the most consistent results among the different models (Figure 56). The marker of chromosome 2 that was significant in both models was the same one that was significant in the analysis of β-damascenone which raises suspicions with regards to the sampling effect of analysing white wine varieties one year and red wine varieties in the second year. The other significant marker was located on chromosome 7 and was significant at the 0.01 level using the BLINK model and at the 0.05 level using the MLMM model. The SNP is located in a gene that encodes a pectin esterase gene that is involved in modifying cell walls and could conceivably change berry cellular structure and alter extraction during fermentation. Another gene in this region is involved in the urea cycle that could influence yeast assimilable nitrogen concentrations, which have been shown to alter ester production during fermentation (Hernández-Orte, et al. 2006). The box plot of the trait versus the genotype at this SNP on chromosome 7 shows a weak relationship (Figure 57), although the 'CC' genotype would be one that should be selected for in a breeding program.

Given there was evidence that the vintage effect might have influenced the analysis, the GWAS was repeated after a rank-based inverse transformation of the data. No significant QTLs were detected from this analysis.

It was disappointing that the GWAS did not confirm the quite convincing mapping of this trait to chromosome 18 in the biparental crosses. However, there are some reasons why this might be so. Firstly, as explained in the introduction, linkage disequilibrium decays rapidly in grapevine. This generally means that a SNP needs to be very close to the causal gene to be detected and presumably this has not occurred in our analysis. Secondly, the polymorphism causing the phenotype in the biparental crosses could be due to the segregation of uncommon alleles in those specific varieties, which is not going to be evident in the broader variety collection. Certainly, the range of phenotypes

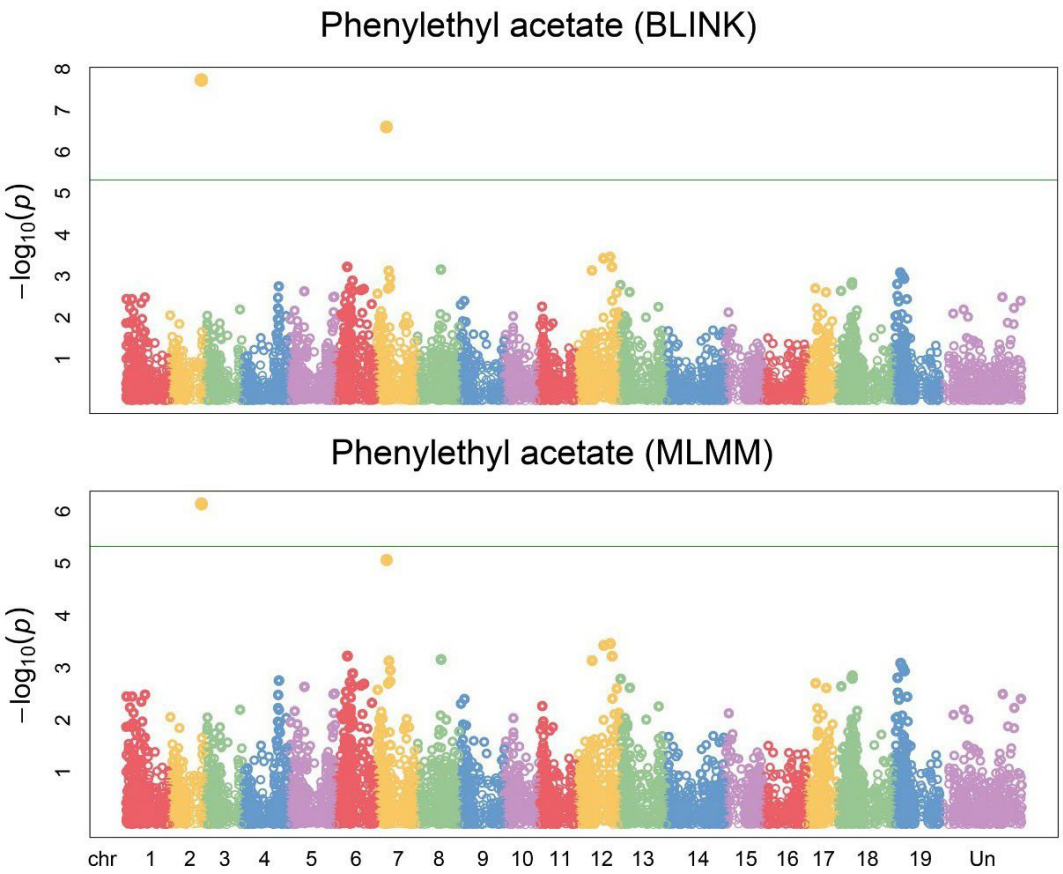


Figure 56 GWAS of phenylethyl acetate in the germplasm collection using the BLINK and MLMM models. The horizontal line represents the significance threshold (p<0.01) and markers grouped by chromosomes with unmapped markers indicated by 'Un'.

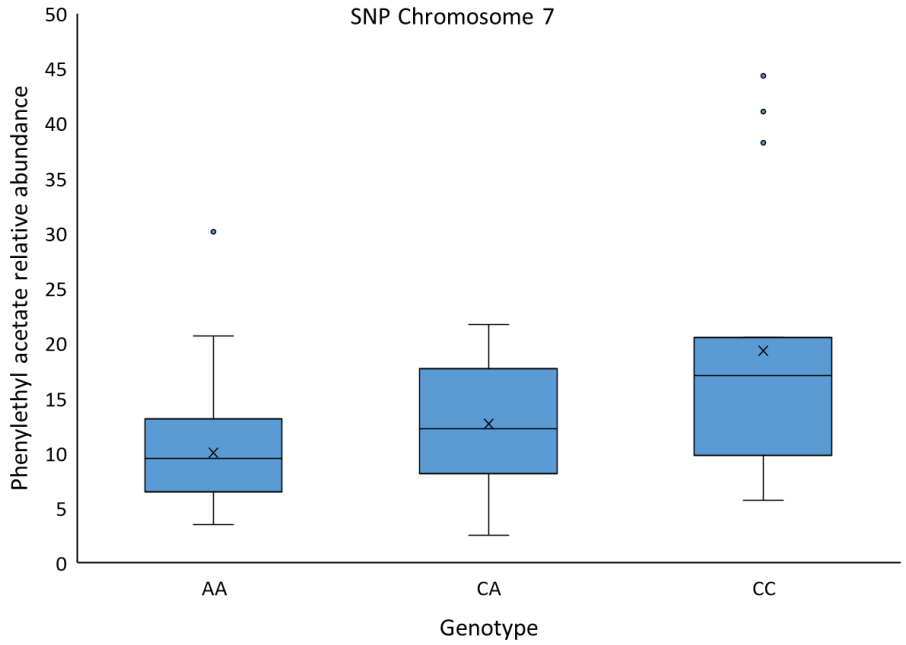


Figure 57 Box plot of the significant SNP markers for phenylethyl acetate concentrations on chromosome 7 showing the relationship between phenotype and genotype in wine replicate 1 of the variety collection.

in the biparental crosses changed ~46 fold from minimum to maximum, whereas the varietal collection only changed ~17 fold and this was with the vintage effect as well suggesting a greater effect was evident for the genetic variation in the biparental crosses. Perhaps the trait is multigenic in the varietal collection and the polymorphisms are epistatic to the one on chromosome 18, thus masking its effect. It is a good demonstration of the difference in the power of both analyses.

Phenylethyl alcohol

Being a fusel alcohol, phenylethyl alcohol is present in most, if not all, wines. It is derived from phenylalanine and is described as having a 'rose-like' aroma and as such is not seen to be as negative on wine sensory attributes as the other higher alcohols. The concentrations of phenylethyl alcohol did show a vintage effect with higher concentrations found in the second year of sampling when the white varieties were analysed. This is reflected in the GWAS with the SNPs at the location of the *VvMYBA1* gene being significant (Figure 58). Another SNP on chromosome 10 was above the significance threshold (p<0.01) when using the BLINK and GLM models and one on chromosome 19 only in the BLINK model (Figure 58).

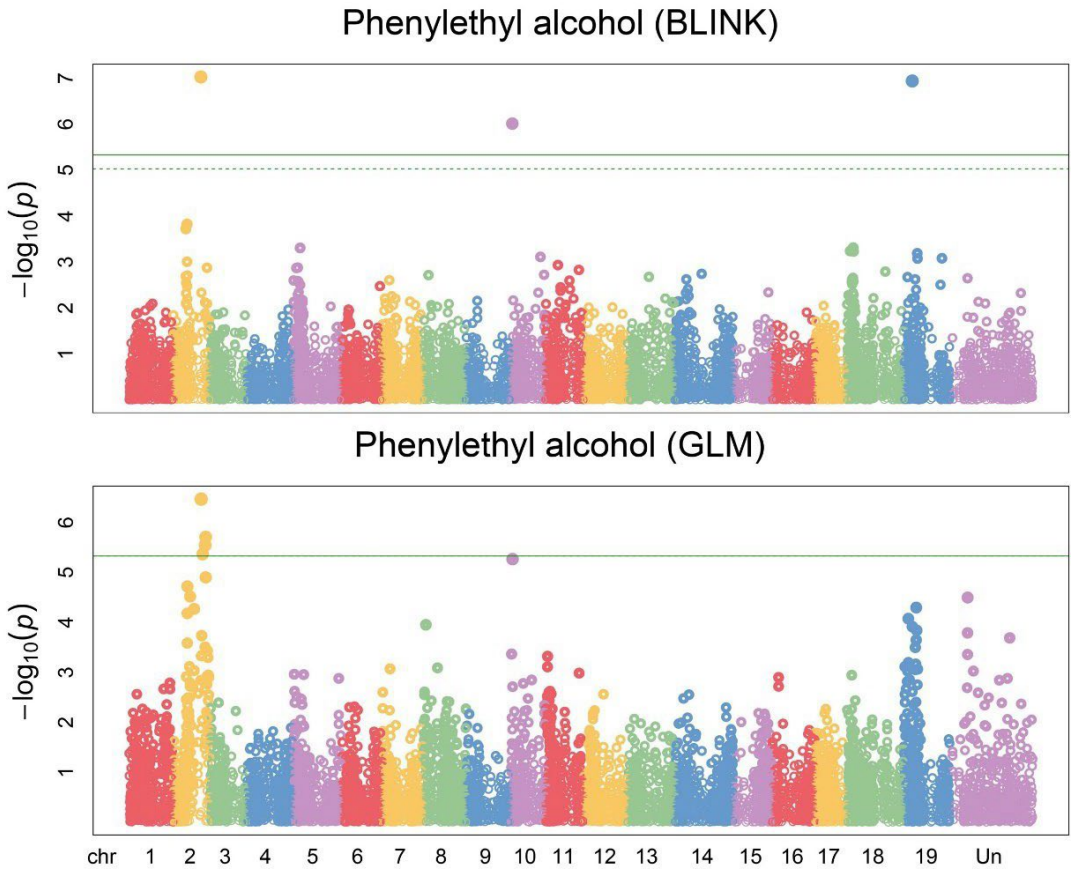


Figure 58 GWAS of phenylethyl alcohol in the germplasm collection using the BLINK and GLM models. The horizontal line represents the significance threshold (p<0.01) and markers grouped by chromosomes with unmapped markers indicated by 'Un'.

The one on chromosome 10 falls in a starch synthase gene that is unlikely to influence phenylethyl alcohol concentrations. However, near to this SNP there are two genes with an involvement in aromatic amino acid production and these are credible targets.

Summary

The GWAS proved to be a good complimentary analysis to the biparental crosses, confirming the findings with regards to the (Z)-3-hexenol concentration and giving us more QTLs to investigate further in the case of many other target compounds. It should be remembered that the power of GWAS relies on the amount of genome covered by the SNPs and with LD decaying quickly in grapevine, this is even more crucial. So, the lack of significant SNP markers on chromosome 18 for acetate ester production in the GWAS could be simply due to this issue (as discussed above) and may be overcome if specific SNP markers in candidate genes in this region are analysed with the data collected in the varietal collection. This could be done in future work if important targets are identified. However, GWAS has also revealed the presence of a number of previously unidentified QTLs for a range of compounds which can now be converted into SNP markers to predict breeding outcomes or investigated further to identify the source of the phenotypic difference. This knowledge could then be applied to other strategies to alter berry composition in the field or via gene-editing.

6.3 Candidate genes underlying QTLs of interest

6.3.1 Introduction

The QTL analyses of wine volatile compound concentrations from the biparental crosses and GWAS resulted in the identification of genome regions that were associated with these specific phenotypes. In the case of the GWAS analyses, the SNPs identified can be visualised using box plots to examine if these are likely to help in the prediction and selection of progeny in a breeding program that have a desirable phenotype or, in some cases, lack a negative phenotype. Examples of this are given in section 6.2 above. With regard to the biparental crosses, the markers found to predict phenotypes in the QTL mapping population are excellent for this purpose in the cross being studied, but further work is required to determine if they have a broader application in other varieties and crosses. One means to do this is to identify SNPs in genes under the QTLs identified in the biparental cross and test them in the varietal collection. There are some assumptions being made here, the main one being that the mutation causing the different phenotypes in the biparental cross is also represented in the broader varietal collection and not unique to the cross. Secondly, if it is present, we must also assume that other mutations are not epistatic to the origin mutation thus masking its effectiveness. In reality, the best (or perfect) marker is one based on the actual mutation that is causing the phenotype, such as the 'Muscat' marker (Emanuelli, et al. 2014) or at least present in the causal gene and tightly linked to the phenotype (Dunlevy, et al. 2013). Given this, we set out to select genes underlying the QTLs identified in the biparental crosses that could be responsible for the phenotype and to characterise their activities, if possible, to test their ability to alter the wine volatile phenotype.

6.3.2 Candidate genes underlying the wine (*Z*)-3-hexenol concentration QTL

Putative (Z)-3:(E)-2 Hexenal Isomerase (HI) genes lie in the QTL region on chromosome 12

There has already been data presented above about the role of (Z)-3-hexenol in wine and how we have mapped the concentration of this compound to a region of chromosome 12 using both biparental crosses and GWAS. The conclusion from this was that between approximately 17.4 to 19.0 Mbp from the top of chromosome 12 there was a strong QTL that influenced the concentration of (Z)-3-hexenol measured in the micro-scale wines made from the progeny of Mv × R and Mv × CS crosses and the varietal collection.

This region of the genome contained 114 annotated genes. Between 17.4 and 17.9 Mbp, the region is dominated by many 'R-gene-like' disease resistance loci, and another cluster of these genes is present downstream between 18.7 and 19 Mbp from the top of chromosome 12. These genes are commonly found to be clustered in plant genomes. However, they are unlikely to influence (*Z*)-3-hexenol concentrations in wine. Between these clusters there is an array of genes with putative functions in aspects of metabolism, signalling and RNA stability, but with no clear role in C_6 alcohol production. However, in the region between 18.1 and 18.2 Mbp from the top of the chromosome there exists a group of genes tentatively identified as cupin seed-storage proteins. A search of the NCBI protein database found that these genes encoded 'germin-like' proteins that are a sub-family of the cupin protein family and have been shown to have a range of functions in plants (Dunwell, et al. 2008). Interestingly, recently a subclass of the germin-like proteins has been shown to possess (*Z*)-3:(*E*)-2-hexenal isomerase (HI) activity and these have been characterised in paprika,

alfalfa, cucumber and rice (Kunishima, et al. 2016, Spyropoulou, et al. 2017). A comparison of the grapevine genes in this region and those characterised as having HI activity found that they grouped with the HI-like proteins (Figure 59).

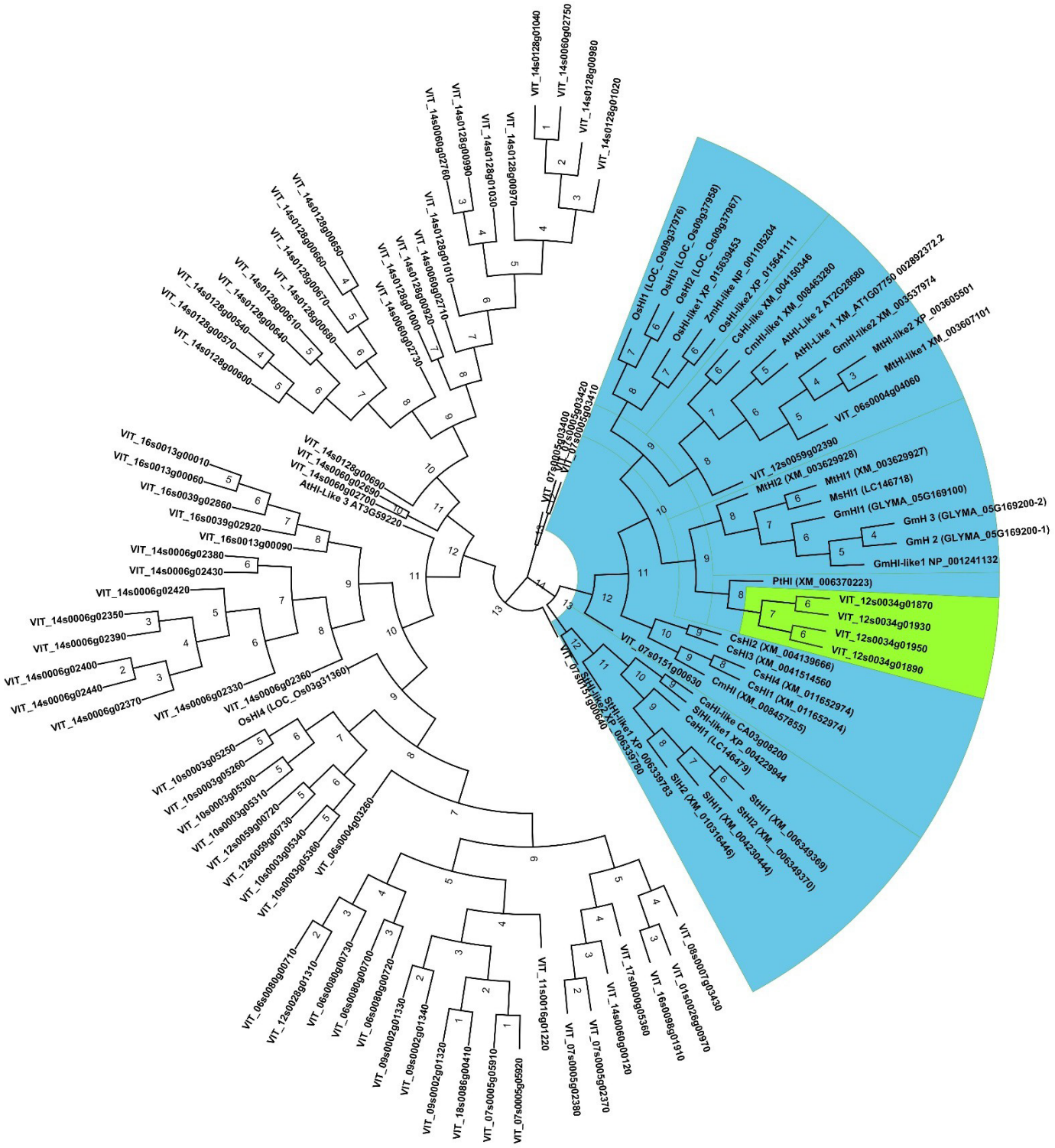


Figure 59 Phylogenetic tree of grapevine 'germin-like' proteins compared with those from other species identified as belonging to the HI subfamily (blue shading) or from other subfamilies. The four grapevine genes in the QTL region on chromosome 12 are highlighted with green shading.

The closest match to the four grapevine genes on the phylogenetic tree is a protein encoded by a gene from poplar (*PtHI*) that has not been functionally characterised, however, the MsHI1 protein in the closest branch of the tree to the grapevine genes is an HI from alfalfa that has been shown to have HI activity. Interestingly, the authors of the paper that first described the HI function in these germin-like proteins (Kunishima, et al. 2016) also showed that some members of this clade did not have HI function, and so it is important to confirm this *in vitro*.

The HI genes in this region are highly homologous

When the list of genes in this region was examined, it was found that there were several sequences that had been annotated as being 'cupin family', 'globulin-like' or 'legumin' in this region. It was apparent that the annotation software may have had difficulty identifying the coding regions for all these genes, but a comparison of different genome builds led us to suggest that there are 4 HI genes in this region with at least three partial genes also located in this part of the genome (Table 13).

Table 13 Gene symbol on NCBI, location on version 2 of the grape genome build and other annotated names given to the 7 putative HI genes in the QTL identified on chromosome 12. Those in italics are listed here as pseudogenes as they do not encode full HI proteins.

GENE SYMBOL (NCBI)	GENE NAME	LOCATION (v2)	OTHER ANNOTATED NAMES
LOC100249816	*VviHI1*	18100764..18103212	VIT_12s0034g01870 (v2), Vitvi12g01876_t002 (v4)
LOC100254961	*VviHI2*	18112381..18114553	VIT_12s0034g01890 (v2), Vitvi12g01876_t001 (v4)
LOC100854631 (Pseudo)		18130831..18139662	VIT_12s0034g01910 (v2), Vitvi12g04481 (v4)
LOC100260054	*VviHI3*	18149101..18151309	VIT_12s0034g01930 (v2), Vitvi12g02659 (v4)
LOC100265260	*VviHI4*	18177564..18179736	VIT_12s0034g01950 (v2), Vitvi12g02659 (v4)
LOC100263489 (Pseudo)		18180999..18182146	VIT_12s0034g01970 (v2), Vitvi12g01880 (v4)
LOC100246341 (Pseudo)		18190709..18197889	VIT_12s0034g01980 & VIT_12s0034g02000 (v2), Vitvi12g01881 & Vitvi12g01882 (v4)

The HI genes in this genome region show a high amount of similarity, with the amino acid sequences encoded by these four genes having between 92-97% identity (Figure 60). This suggests there has been a relatively recent duplication in this region of the genome with our hypothesis being that one pair of genes has replicated further along the chromosome. The close amino acid and DNA sequence homology between all these genes created issues for further molecular studies as it was difficult to design primers to amplify each gene individually and to design markers that were specific to a single SNP without amplifying other sequences such as the other full-length genes or the pseudogene fragments also in this region.

Expression of the HI-like genes in berries

Despite the close homology in nucleotide sequences, primers were eventually designed so that qRT-PCR could be conducted to determine the expression pattern of these genes during berry development. Gene expression was studied in three varieties all grown in the same vineyard in Eden Valley to minimise regional effects on results, although there would have been intra-vineyard

```
Consensus       MELNLAPKFAQKIFEGEGGTYY.WSSAEYELLKEAKVGGGRLVLQPRGFALPHYADSN.I   60
VviHI1          ......................S.....................G. ............K.   60
VviHI2          ......................R.....................................R.   60
VviHI3          ......................R.....................................R.   60
VviHI4          ..................E...S.....................................K.   60

Consensus       GYVLQGSCGVVG.VSPKASQEVVLRLKKGDIIPVPSGAVSWWYNDGDSELIIVFLGETSK  120
VviHI1          ............M.F.E..E....K..E.....................V.........  120
VviHI2          ............I.P.............................................  120
VviHI3          ............M...N.............................................  120
VviHI4          ............I.............................................  120

Consensus       AYVPGEFTYFLLTGTQGILGGFSTEFNSRAYDINNEEAKKLAKSQSGVL.IKLPEGHKMP  180
VviHI1          ...............................N.S....E......T...L......Q...  180
VviHI2          ........N.....................................I..........  180
VviHI3          ..............................................L.L..........  180
VviHI4          ........................................R......I..........  180

Consensus       HPCKNSTDKLVFNIDAALPDIHVQNAGLLTALTAKKFPFLGEVGLSATLVKLDANAMSSP  240
VviHI1          ............Y.............................................  240
VviHI2          ..............................................................  240
VviHI3          ......P.....................................................  240
VviHI4          ..............................................................  240

Consensus       .YAADSSVQVIYVAKGSGRIQVVGINGERALDTKVKAGHLLVVPRFFVASAIADGEG.EY  300
VviHI1          V.....................................Y.........T. .....M..  300
VviHI2          M..G.............................A........L. ..............L..  300
VviHI3          V.....................................F........A. ......M..  300
VviHI4          M ..............................................L..  300

Consensus       FSLITAT.PVFGEFTGKTSVWGALSP.VLQASLNVAPEFEQLFRAKIKKSTILVPPQN    358
VviHI1          .......Q...................Q.............................    358
VviHI2          .......E.I.................H.............................    358
VviHI3          ..M..T.Q...................Q........G.....................    358
VviHI4          .......E...................H.............................    358
```

Figure 60 Alignment of the predicted amino acid sequences encoded by the 4 putative full-length HI genes in the PN40024 genome located on chromosome 12 in the QTL identified by biparental crosses and GWAS. The consensus sequence is shown at the top of the alignment, and derivations from the consensus are indicated below each residue, those being identical to the consensus are shown as a '.' dot.

variation given that the blocks were some distance apart. Pinot gris was used as the 'Pinot' variety in the berry development gene expression analysis as a proxy for the microvine.

To put the expression of the HI genes in context, we needed to know when the concentrations of the unsaturated C_6 compounds changed during berry development. HS-SPME GC-MS was used to quantify the relevant C_6 compounds in the berries that could be detected in the same samples used for gene expression analysis. The four compounds focussed on were (*E*)-2-hexenal, (*E*)-2-hexenol, (*Z*)-3-hexenal and (*Z*)-3-hexenol, with the aldehydes being present in higher concentrations than the alcohols. In Cabernet Sauvignon there is a trend for the aldehydes to reduce after veraison with (Z)-3-hexenal showing a more drastic reduction than (E)-2-hexenal (Figure 61). In contrast, the concentration of (E)-2-hexenol was much higher in all samples taken after veraison than before, while (Z)-3-hexenol increased in the preveraison period and decreased after the onset of ripening. So, there is an increase in the proportion of '(E)-2' compounds compared to '(Z)-3' after veraison, especially the alcohols.

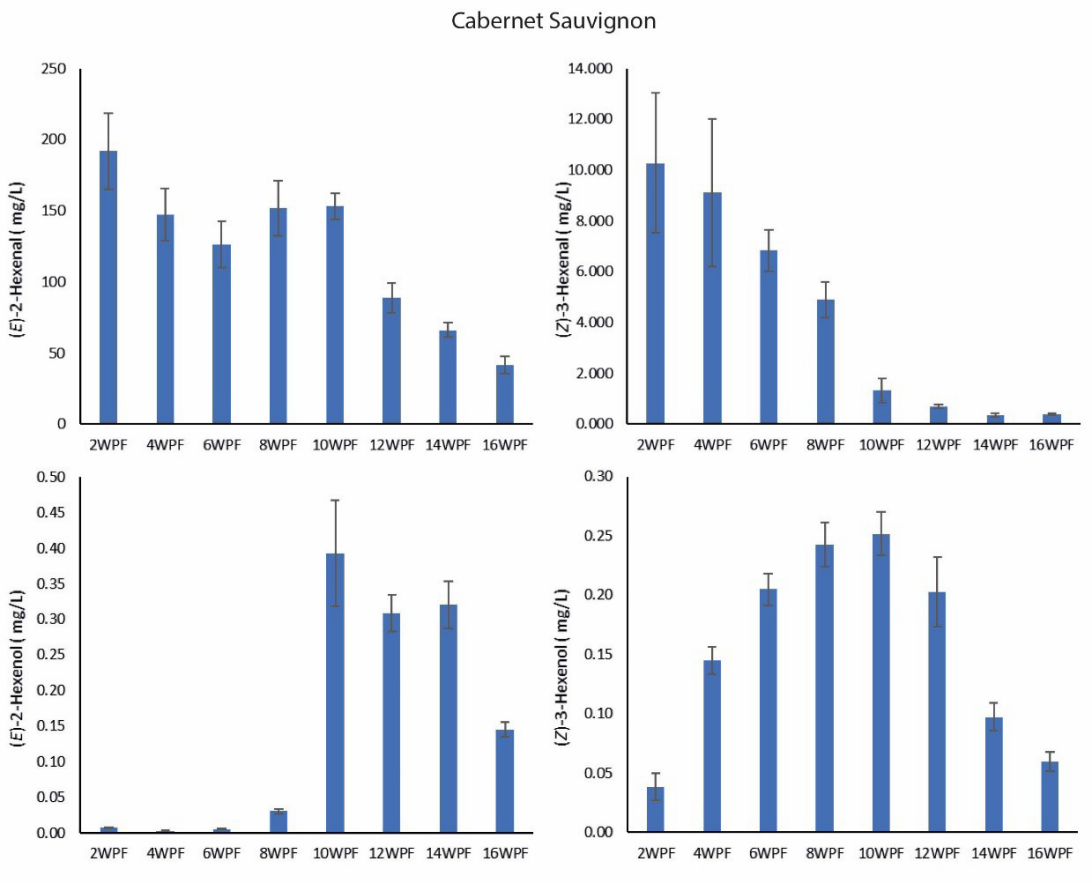


Figure 61 Concentrations of (*E*)-2-hexenal, (*E*)-2-hexenol, (*Z*)-3-hexenal and (*Z*)-3-hexenol in the Cabernet Sauvignon developmental samples taken from Eden Valley in the 2018/19 vintage.

The trends in the Pinot gris samples were somewhat similar to those seen in the Cabernet Sauvignon samples. In short, concentrations of the aldehydes decreased after veraison, but more so for (Z)-3-hexenal and the proportion of (E)-2-hexenol to (Z)-3-hexenol increased, in this case due mainly to a significant increase in (E)-2-hexenol after veraison (Figure 62).

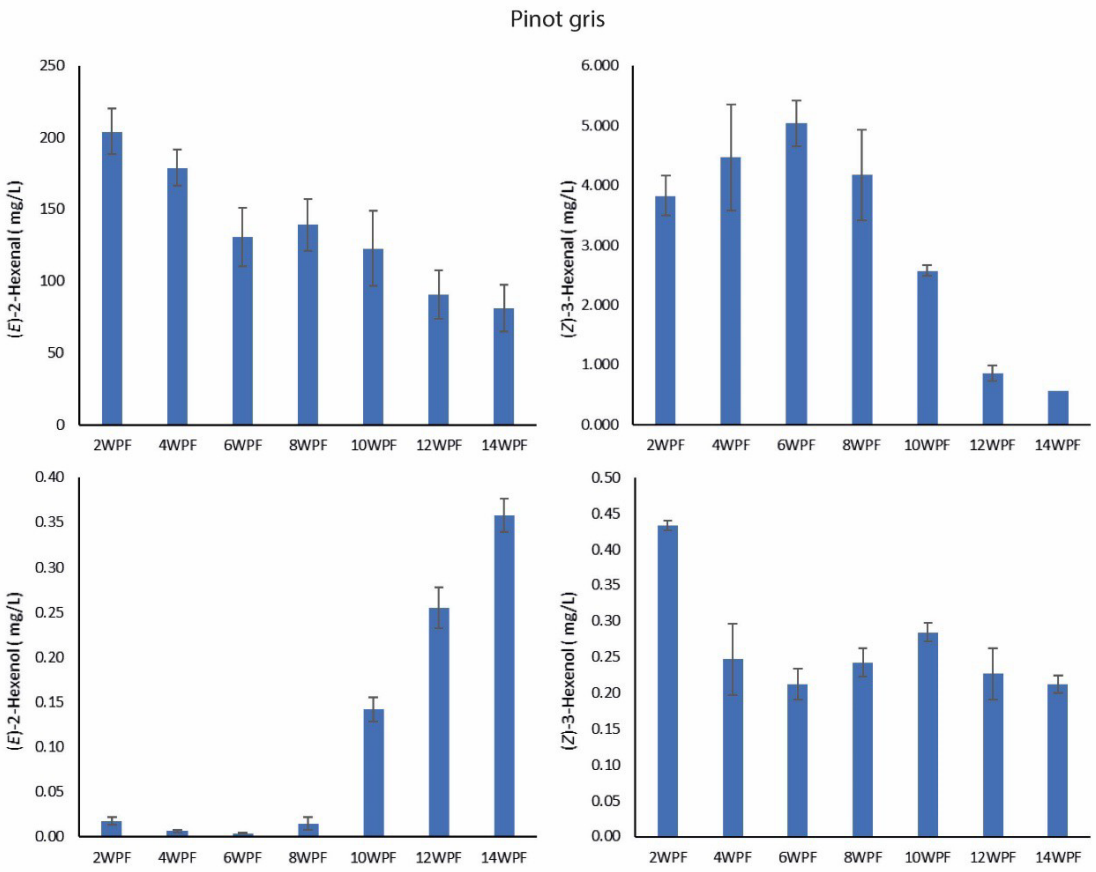


Figure 62 Concentrations of (*E*)-2-hexenal, (*E*)-2-hexenol, (*Z*)-3-hexenal and (*Z*)-3-hexenol in the Pinot gris developmental samples taken from Eden Valley in the 2018/19 vintage.

The Riesling developmental samples also followed the same trend as the other two varieties. There was a general decrease in aldehydes during development, but a greater proportional reduction of (Z)-3-hexenal after veraison compared to (E)-2-hexenal (Figure 63). There was an unusual spike in the alcohols at 11 wpf, but in general, there was more (E)-2-hexenol in the samples after veraison compared to before veraison, whereas the (Z)-3-hexenol concentrations trended to decrease. This suggests that the berries have greater capacity to convert (Z)-3-hexenal to (E)-2-hexenal after veraison, especially in the region of the berry where the aldehydes are reduced to alcohols, presumably by alcohol dehydrogenase.

Expression of the four putative HI genes was quantified in the development series samples from the three varieties. The *VviHI4* gene was not expressed in any of the berry samples tested. The *VviHI1* gene was only expressed in the flowers of the three varieties and the highest copy number accumulated of this gene was in the Riesling flower sample (Figure 64). This might explain the slightly lower concentrations of (Z)-3-hexenol in the earliest Pinot gris and Riesling berry samples compared to the other pre-verasion samples (Figure 62, Figure 63). The most highly expressed gene was *VviHI2* with estimated copy numbers in ripe Cabernet Sauvignon berries exceeding 30 million (Figure 65). Expression of this gene was also highest after veraison in Riesling and Pinot gris, but not to the same extent as Cabernet Sauvignon (Figure 65). The expression overlaps with the increase in (E)-2-hexenol and decrease in (Z)-3-hexenol concentrations in these samples. The *VviHI3* gene was also up-regulated in the later stages of ripening (Figure 66), but expression of this gene was much lower than *VviHI2*. What this gene analysis suggests is that *VviHI2* and *3* contribute to the conversion of

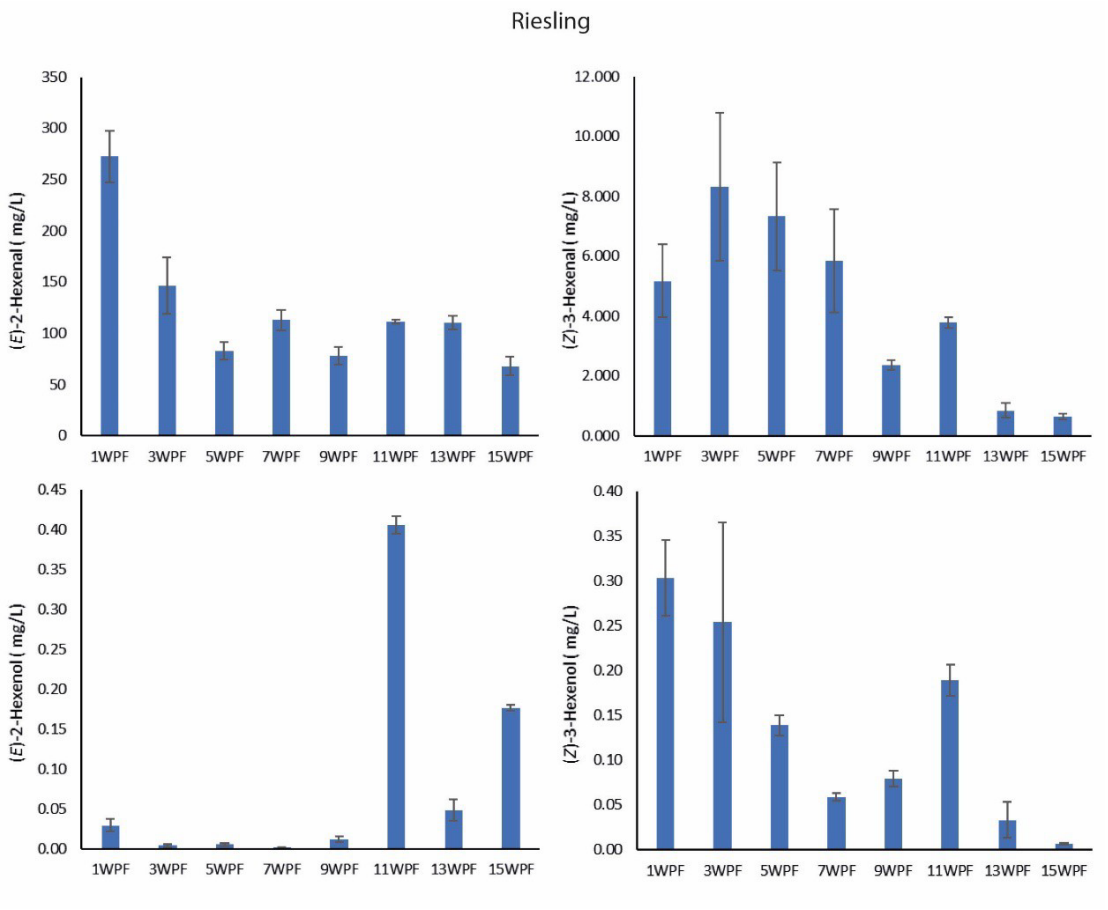


Figure 63 Concentrations of (*E*)-2-hexenal, (*E*)-2-hexenol, (*Z*)-3-hexenal and (*Z*)-3-hexenol in the Riesling developmental samples taken from Eden Valley in the 2018/19 vintage.

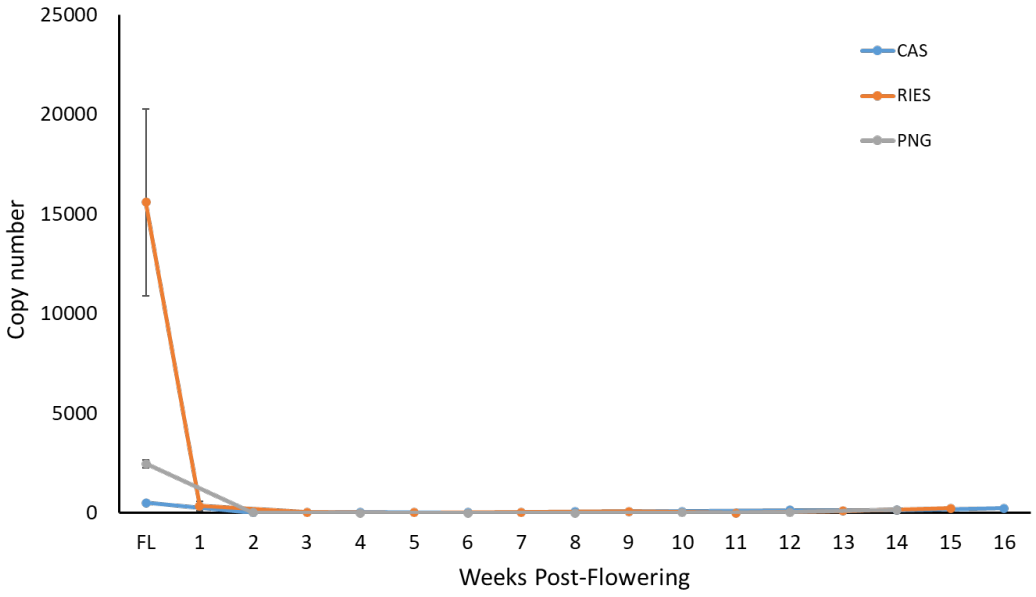


Figure 64 Expression of the *VviHI1* gene in the Cabernet Sauvignon (CAS), Riesling (RIES) and Pinot gris (PNG) berry samples from throughout development in the 2018/19 growing season.

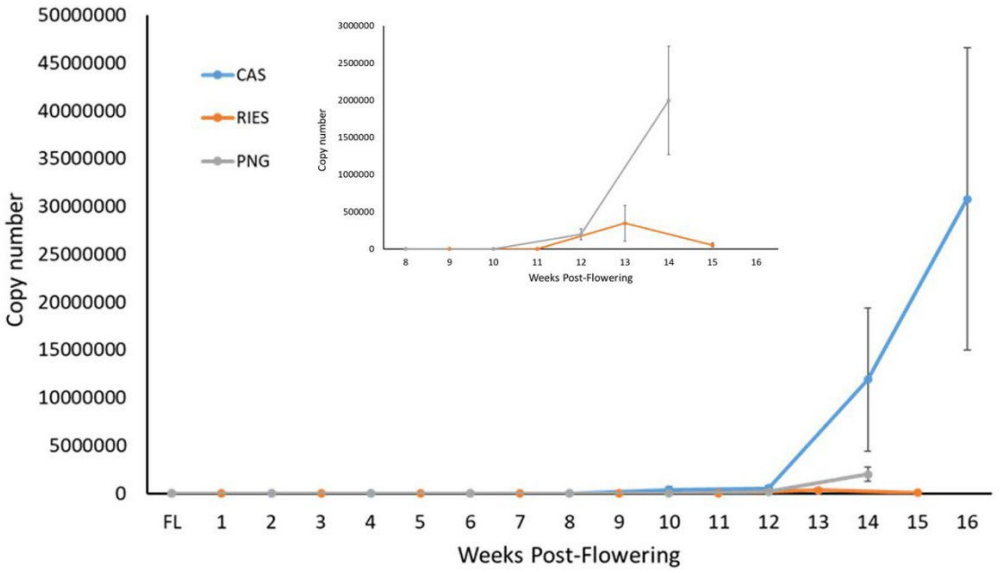


Figure 65 Expression of the *VviHI2* gene in the Cabernet Sauvignon (CAS), Riesling (RIES) and Pinot gris (PNG) berry samples from throughout development in the 2018/19 growing season. The inset shows the Riesling and Pinot gris samples from 8-15 wpf to demonstrate the increase in expression in these samples.

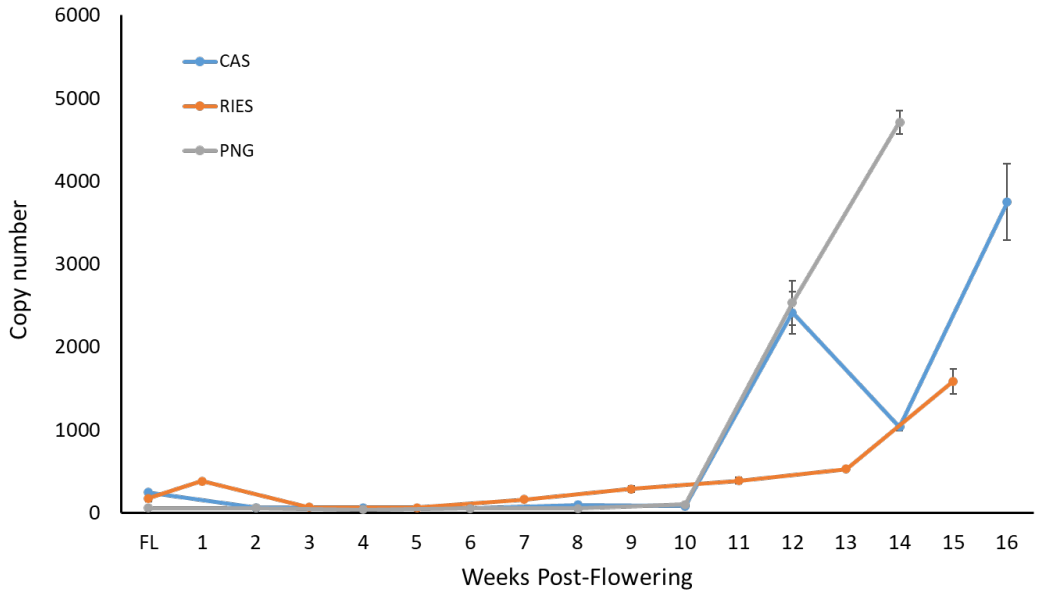


Figure 66 Expression of the *VviHI3* gene in the Cabernet Sauvignon (CAS), Riesling (RIES) and Pinot gris (PNG) berry samples from throughout development in the 2018/19 growing season.

(Z)-3-hexenal to (E)-2-hexanal and their respective alcohols in the later stages of ripening in all three varieties.

VviHI1, *2* & *3* show HI activity *in vitro*

Previous papers have shown that these cupin-like protein family members can be expressed as recombinant protein in *E. coli* and their function tested *in vitro* (Kunishima, et al. 2016, Spyropoulou, et al. 2017). We conducted similar experiments with the four putative HI genes found under the QTL on chromosome 12 that was associated with (*Z*)-3-hexenol concentrations in wine from progeny of the two biparental crosses.

Alleles from all four HI genes from Riesling were cloned into a recombinant enzyme expression vector and the protein subsequently purified using a poly-histidine binding column. The proteins were then tested for activity using conditions described in materials and methods (section 5.5.2). The (Z)-3-hexenal substrate is stored in triacetin in order to stabilise the compound, but this seems to have introduced some impurities in the assays, especially the compound (E)-2-hexenal dimethyl acetal, which elutes at the same time as (E)-2-hexenal. This can be seen in the 'no-enzyme' assays (Figure 67), but it becomes a minor contaminant in the assays with enzyme added.

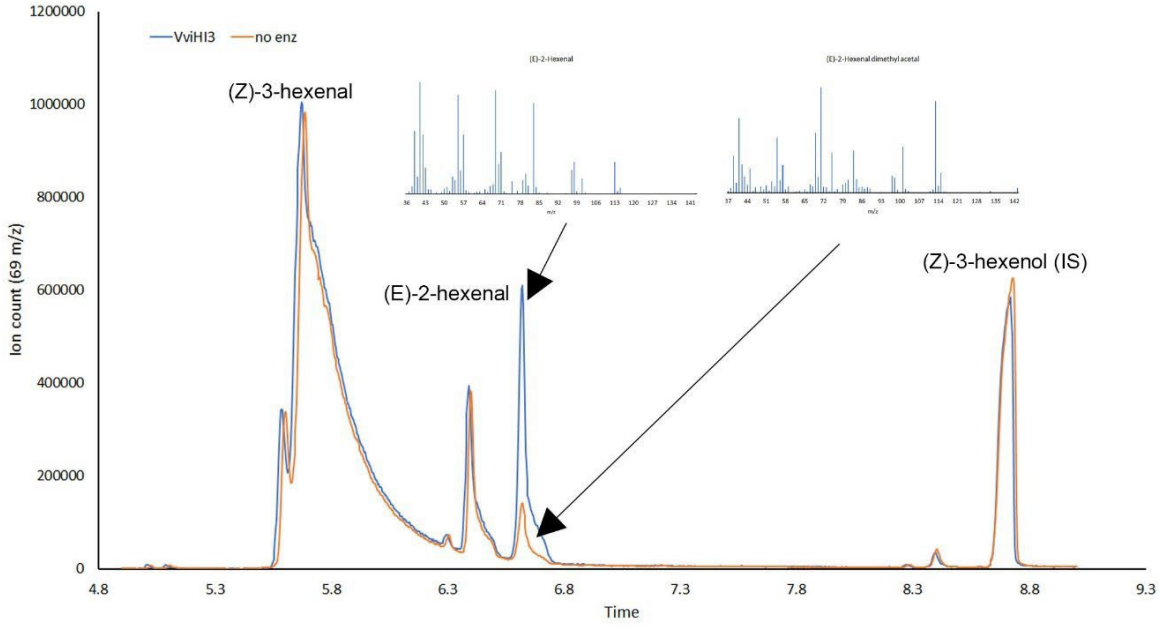


Figure 67 Ion count for 69 m/z versus elution time in minutes for hexane extracts of assays conducted using (*Z*)-3-hexenal as the substrate incubated with either recombinant VviHI3 protein (blue) or with no enzyme added to the assay mix (orange). Mass spectra for the peaks eluting at about 6.7 minutes are shown near the top of the graph. The peak corresponding to the (*Z*)-3-hexenol internal standard is also indicated on the chromatogram.

Three of the four genes tested showed hexenal isomerase activity. Interestingly, the three that were shown to be expressed in berries were the three that showed HI activity when expressed as recombinant proteins in *E. coli* (Figure 68), namely VviHI1, 2 and 3. Similar amounts of recombinant VviHI were used in each of the assays and resulted in a similar amount of (E)-2-hexenal being produced. This showed that the proteins encoded by these genes are likely to be involved in the isomerisation of (Z)-3-hexenal to (E)-2-hexenal in berries. The assay was also conducted with (Z)-3-hexenol, but no corresponding (E)-2-hexenol was produced (data not shown), indicating that this conversion occurs at the aldehyde point in the pathway (Figure 5). This means that if there is high

alcohol dehydrogenase activity in the berry causing a rapid reduction of (Z)-3-hexenal to (Z)-3-hexenol, there is less opportunity to isomerase the aldehyde to (E)-2-hexenal. So, this could be an important control point in the pathway.

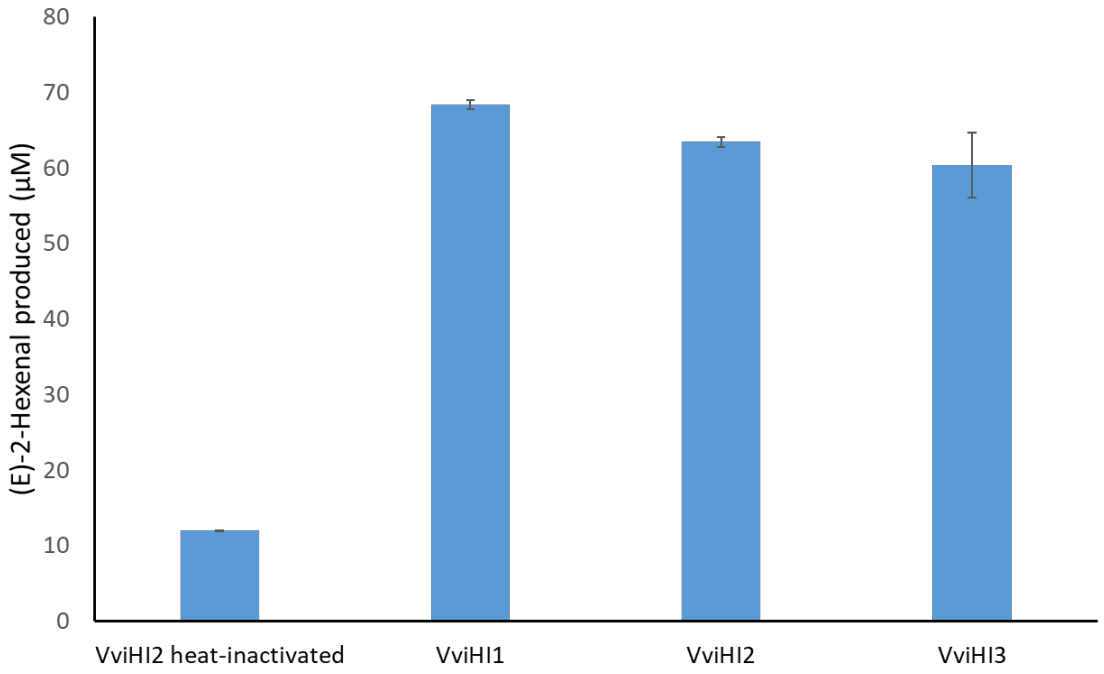


Figure 68 Activity assays of recombinant VviHI proteins assayed with 200 µM (Z)-3-hexenal and compared to the assay performed with heat-inactivated VvHI2. Similar amounts of recombinant protein were used for each assay.

Marker design for predicting wine (Z)-3-hexenol concentration

Because of the high DNA sequence identity between the *VviHI* genes and alleles, the causal mutations behind the different phenotypes could not be identified. This similarity also made marker design problematic, as designing primers unique to each gene was extremely difficult when certain regions of the gene had to be targeted. It was decided to design markers to the two SNPs shown in Figure 51 as they had already been shown in the DArTseq analysis to have good predictive power.

The SNP at ~17.1 Mbp from the top of chromosome 12 is the result of a G>C change in the transcribed region of a grape patatin-like phospholipase domain-containing protein. The SNP introduces a *Ppu*MI restriction enzyme site when the 'C' nucleotide is present as it recognises the sequence 5'..RGGWCCY..3'. Primers were designed to generate a cleaved amplified polymorphic sequences (CAPS) marker and the details are listed in Table 14.

The SNP at ~18.6 Mbp is caused by a C>T change in an UDP-glucose:glycoprotein glucosyltransferase gene. The presence of a 'T' eliminates an *Hpa*II restriction enzyme site which recognises the sequence 5'..CCGG..3'. Although this would be considered a common cutter given the four base pair recognition site, there are no other *Hpa*II sites in the close vicinity of this gene and so a marker could be designed using this enzyme. The marker details are listed in Table 14.

Table 14 Details of CAPS markers for predicting wine (*Z*)-3-hexenol concentrations in new varieties.

MARKER	5'PRIMER	3'PRIMER	ENZYME	ALLELE 1	ALLELE 2
Vv12@17.1	AACCCTTTCCTGAACTTTTCCACGGA	GGCTTCCTCCCTCTCGTCCCT	*Ppu*MI	'C' 431+114bp	'G' 545bp
Vv12@18.6	TGGTCAAACAATGTTAGGGGGTGGG	CTGCAGGTTATATTTGTCGATGCCGA	*Hpa*II	'C' 447+312bp	'T' 759bp

6.3.3 Candidate genes linked to *VvGAI1* potentially involve in monoterpene biosynthesis

A gene next to the crossover in plant #131 is predicted to hydrolyse polyphosphates

As outlined in section 6.1.3, a Mv × Riesling cross gave a strange segregation result for the production of moderate amounts of monoterpenes, which suggested that there was a gene linked to the dwarfing locus that inhibited monoterpene production (Figure 27). This was supported by QTL mapping (Figure 28). From the mapping it was predicted that the causal gene fell somewhere between ~3.4 and 4.6Mbp from the top of chromosome 1 (Figure 27). This area contained approximately 130 annotated genes with a broad range of functions. Interestingly, the first gene located upstream from the crossover in plant #131 (which produced Riesling-like concentrations of linalool) comes from a family of genes that encode enzymes known as Nudix hydrolases (NUDX). These enzymes are part of a superfamily of proteins with both enzymatic and structural roles in organisms from the three domains of cellular life (Srouji, et al. 2017), however, the subfamily of hydrolases is of interest in the case of the gene on chromosome 1.

The NUDXs are defined by a conserved motif, which is the catalytic site of the enzyme and imparts a pyrophosphohydrolase activity on the enzyme (Yoshimura and Shigeoka 2015). These enzymes cleave phosphate groups off a range of substrates including nucleotide sugars, NAD(P)H, FAD, coenzyme A and diadenosine polyphosphates. Of interest to monoterpene production, there are two recent reports of NUDX genes being involved in controlling the biosynthesis of these compounds, but in conflicting ways. In roses, a NUDX named RhNUDX1 was shown to have geranyl diphosphate (GPP) diphosphohydrolase activity, converting GPP to geranyl monophosphate (Magnard, et al. 2015). This was proposed to be part of a pathway in the flowers that involved another phosphatase to convert GPP into geraniol and the transient expression of this gene in tobacco leaves resulted in an increase in the production of geraniol glycosides. The other relevant study looked at two NUDXs from *Arabidopsis* and showed that they could convert the five carbon diphosphate precursors of GPP, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), to their monophosphate forms isopentenyl phosphate (IP) and dimethylallyl phosphate (DMAP). Interestingly, mutations in these two *Arabidopsis* NUDX genes increased linalool concentrations in the plants (Henry, et al. 2018). We began characterisation of this gene as it was considered to be likely involved in the phenotype.

LOC100247306 is most similar to the *Arabidopsis* gene *AtNUDX25*

The grapevine NUDX gene next to the plant #131 chromosome 1 crossover point was compared to NUDX genes from *Arabidopsis,* as the family of genes has been best studied in this plant species. A phylogenetic tree comparing the NUDX family of genes from both grapevine and *Arabidopsis* was generated (Figure 69). The closest *Arabidopsis* protein sequence to the one by the crossover point was annotated as At1G30110, which is otherwise known as *AtNUDX25* and so we named the grapevine NUDX gene of interest *VvNUDX25*. As a control to help develop functional assays we decided to also conduct gene characterisation using the closest grapevine gene to the one shown to be involved in rose monoterpene production. This was most similar to the *Arabidopsis* gene annotated as AT1G68760 and otherwise known as *AtNUDX1*. To follow convention, we called this gene *VvNUDX1*.Interestingly, this gene is also located on chromosome 1, but only ~2.5 Mbp from the top of the chromosome and many of the Mv × R F1 progeny have crossovers covering this region but do not produce monoterpenes to the same extent as Riesling or plant #131.

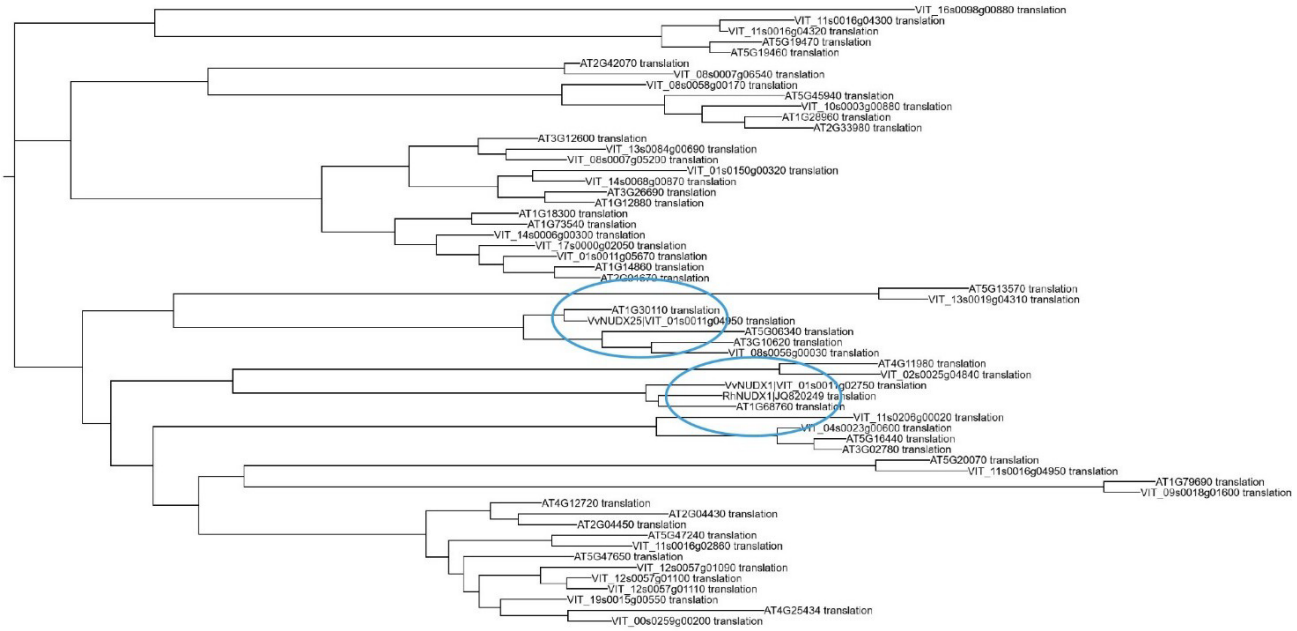


Figure 69 Phylogenetic tree showing a comparison of the protein sequences of putative NUDX genes from *Arabidopsis* and grapevine. The cycled groups correspond to the gene found closest to the crossover point in the #131 plant at the top and the lower circled group corresponds to another NUDX gene found further upstream on chromosome 1 which was used as a control.

Both VvNUDX25 and VvNUDX1 function as pyrophosphatases

Microvine berry cDNA was used to amplify and clone VvNUDX1 and 25 into appropriate vectors for transient expression in tobacco leaves and for the production of recombinant protein in *E. coli*. The first assay conducted using these grapevine NUDXs was transient expression in tobacco leaves as this had been shown for the rose *RhNUDX1* gene to increase the concentration of glycosylated geraniol (Magnard, et al. 2015). The leaves were infiltrated with expression constructs containing either NUDX gene and an empty vector control and sampled after 7 days. The leaves were then ground and the glycosidic monoterpenes extracted and analysed using an LC-MS QTOF instrument.

There is a certain amount of endogenous geraniol glycoside present in the tobacco leaves, but the infiltration of *Agrobacterium* carrying a plasmid to express either *VvNUDX1* or *VvNUDX25* resulted in a substantial increase in geraniol glycoside production (Figure 70). The introduction of *VvNUDX1* increased the production of these compounds 14-fold and *VvNUDX25* almost five-fold. The relative increase cannot be taken as an indication of the enzyme affinity or activity towards GPP as the infiltration procedure brings into play aspects of transcription efficiency and mRNA stability that could change the accumulation of the enzyme and hence the compound measured. This was an interesting finding with regards *VvNUDX25*, as it has been shown that it has activity towards diadenosine tetraphosphate, diadenosine pentaphosphate, NADH, CoA (Yoshimura, et al. 2007) and guanosine-3',5'-tetraphosphate (Ito, et al. 2012), but does not seem to have been tested against other phosphorylated substrates.

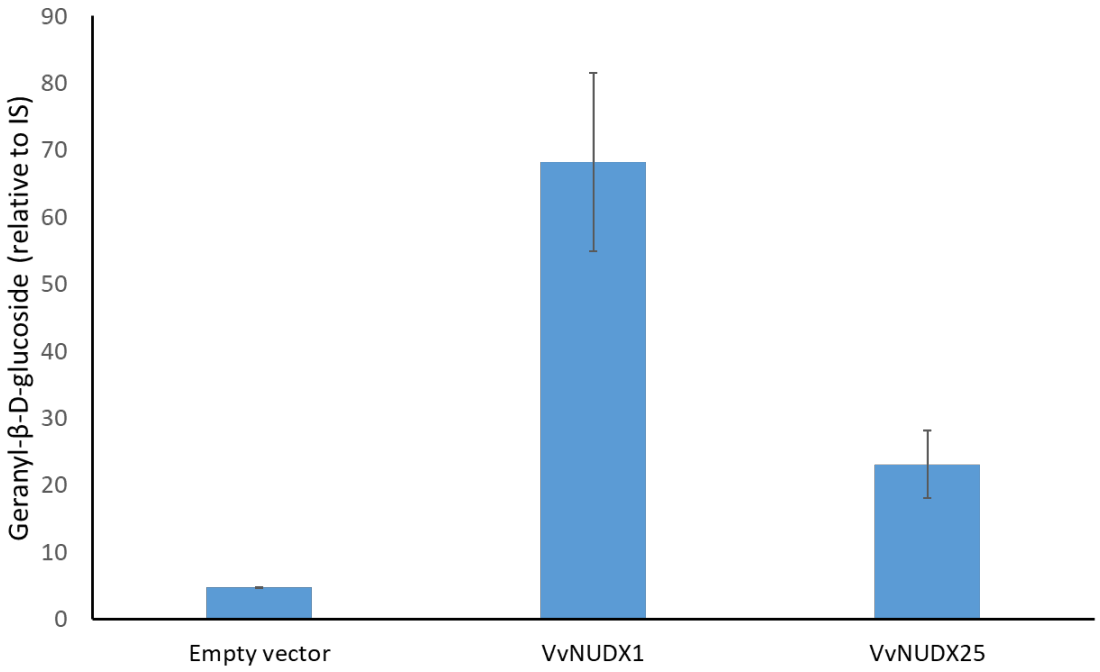


Figure 70 Relative geranyl-β-D-glucoside concentrations produced by VvNUDX1 and 25 proteins assayed by transient inoculation into tobacco leaves and compared to an empty vector control.

The results of the tobacco leaf inoculation suggested that the cDNA clone for *VvNUDX25* encoded an active enzyme and so we decided to use it to produce recombinant enzyme in *E. coli* to test activity against other relevant substrates. VvNUDX1 protein was also expressed and purified as a control. The first substrate tested was one that was known to be a compound that AtNUDX25 could hydrolyse - diadenosine tetraphosphate (Ap4A). This was confirmed in the grapevine homologue as we could detect ATP and AMP when VvNUDX25 enzyme was incubated with the Ap4A substrate (Figure 71). Neither the VvNUDX1 enzyme nor heat-denatured VvNUDX25 showed any activity against Ap4A.

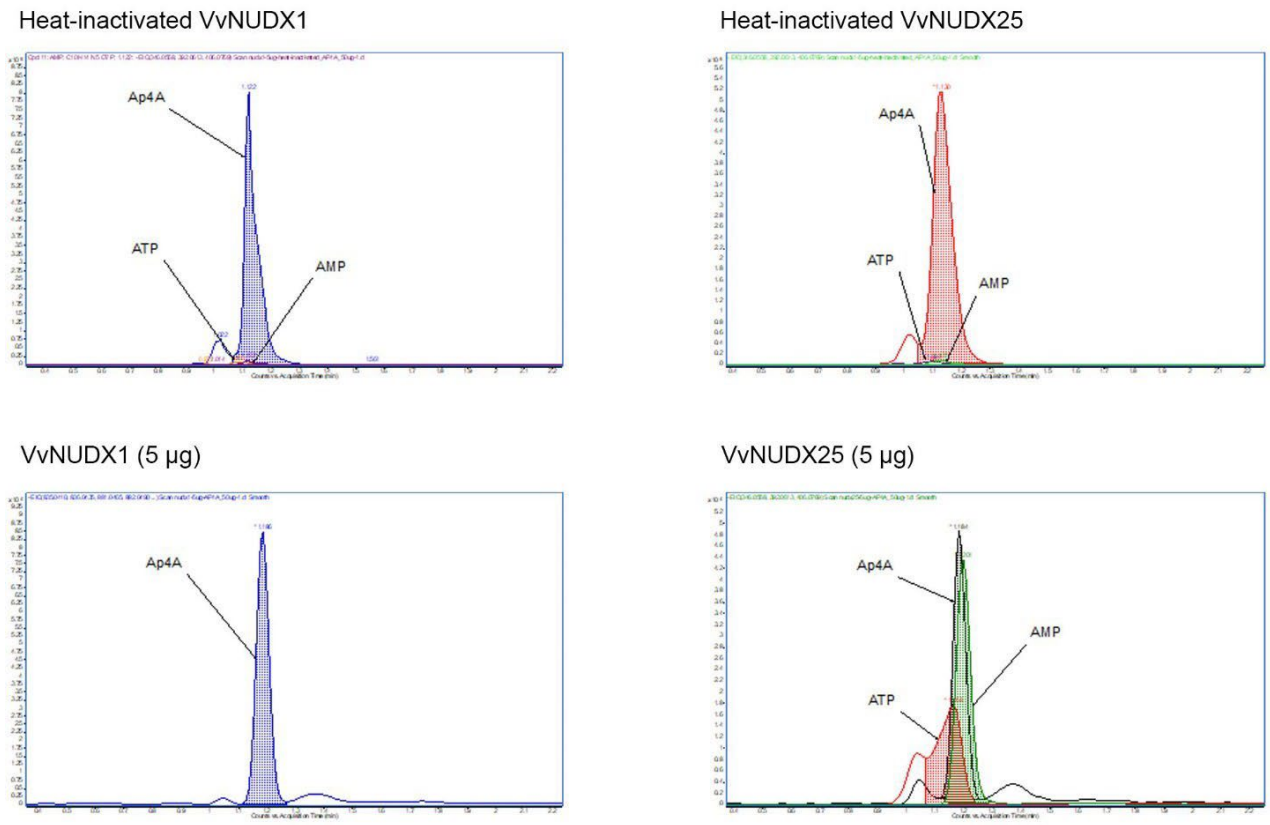


Figure 71 Extracted ion chromatograms showing the Ap4A substrate and the identified reaction products (ATP and AMP) after dephosphorylation.

Following this, we tested the activity of both VvNUDX1 and VvNUDX25 against precursors for both monoterpene and sesquiterpene production in plants, namely GPP and farnesyl diphosphate (FPP) respectively. In both heat-denatured samples, there was a trace of geranyl phosphate (GP), but this greatly increased when both VvNUDX1 and 25 were incubated with the substrate (Figure 72).

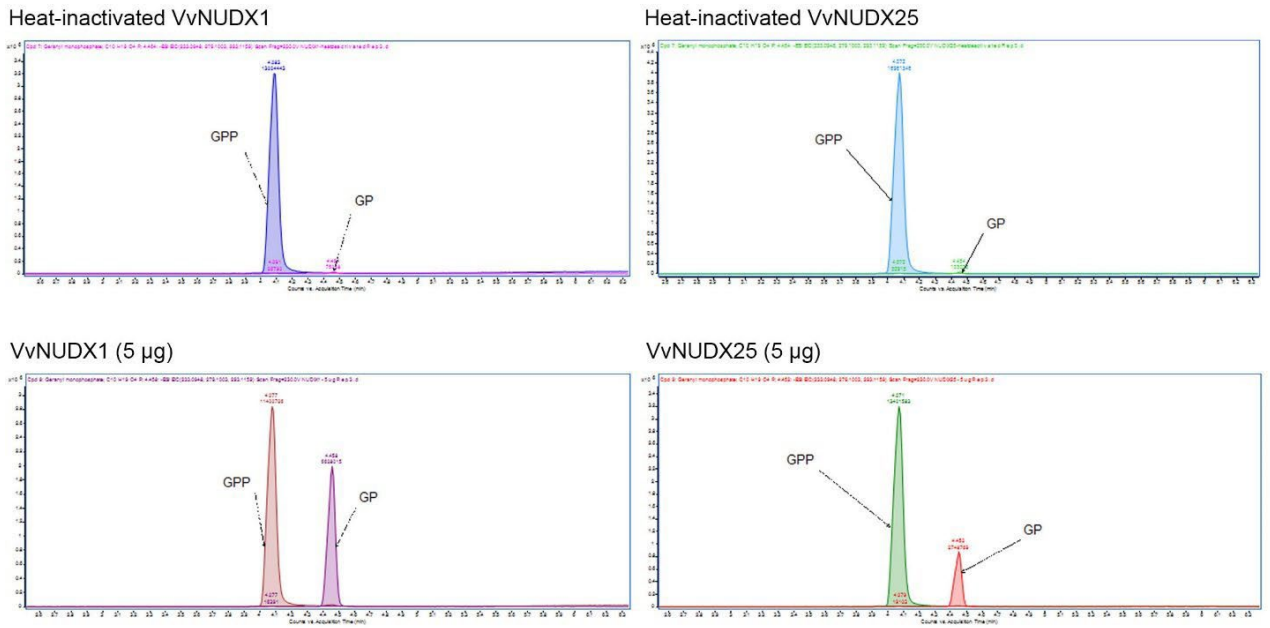


Figure 72 Extracted ion chromatograms showing the GPP substrate and the identified reaction product (GP) after dephosphorylation.

Likewise, with FPP there were traces of farnesyl phosphate (FP) present in the heat-denatured enzyme samples. When the FPP substrate was incubated with active VvNUDX1 or 25 enzyme, all of the substrate was dephosphorylated to FP (Figure 73).

The two main conclusions from these results are that: VvNUDX25 has a broad substrate specificity that includes at least Ap4A, GPP and FPP; both VvNUDX1 and VvNUDX25 have the ability to convert the diphosphate precursors for monoterpene (GPP) and sesquiterpene (FPP) biosynthesis into monophosphates that may not be utilised by relevant terpene synthases. So, it is possible that both could affect the accumulation of these compounds in berries, but more work is required to test this hypothesis.

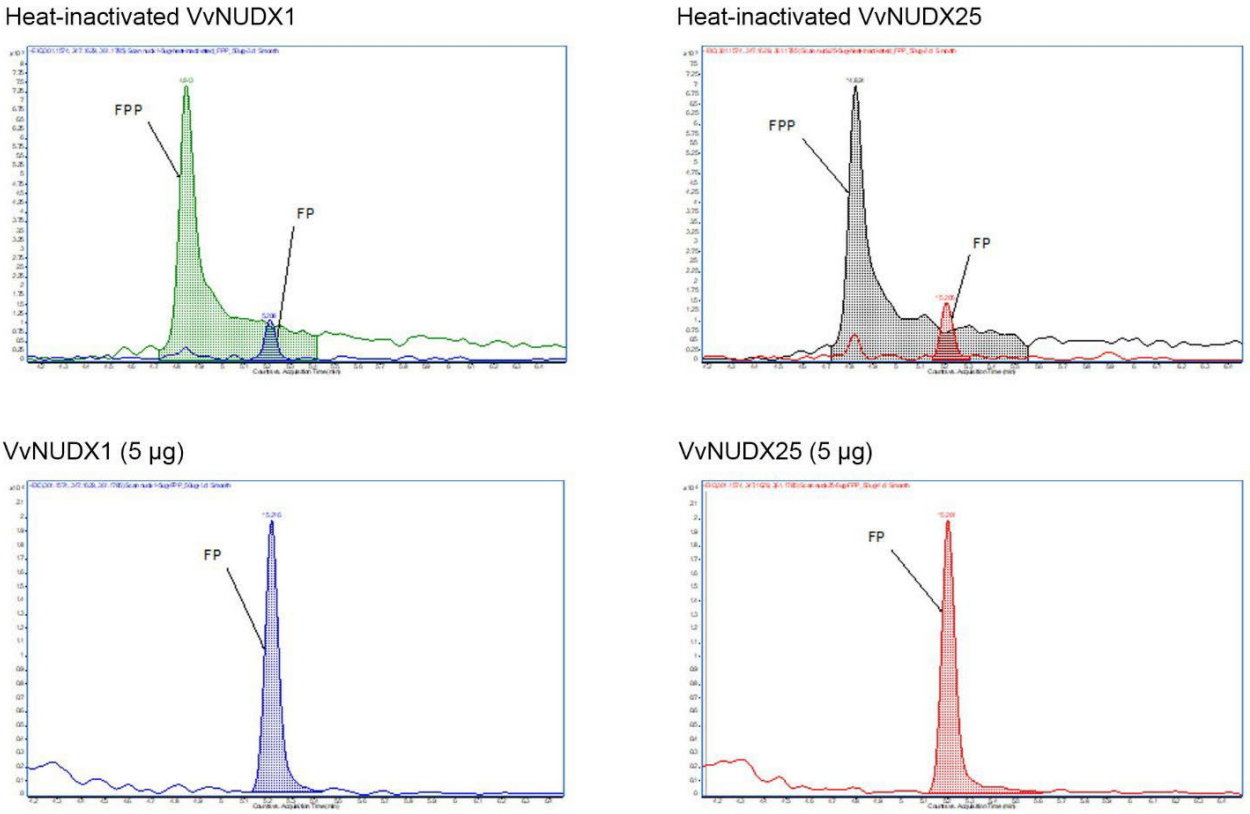


Figure 73 Extracted ion chromatograms showing the FPP substrate and the identified reaction product (FP) after dephosphorylation.

VvNUDX25 is expressed throughout berry development

To provide more evidence about the role of VvNUDX25 in the regulation of monoterpene biosynthesis in grape berries, the expression of this gene during berry development was examined. The Riesling and Pinot gris development series collected from Eden Valley were used and to put the gene expression in context, the abundances of linalool in these samples was determined using GC-MS. The observed changes in free linalool concentrations in the Riesling samples match those observed for micro-scale wines made from a Riesling developmental series (Boss, et al. 2018) where linalool concentrations are high early in development, decrease until veraison, then increase later in the ripening process (Figure 74). The Pinot gris samples are shown on the same scale (Figure 74) and it is obvious that concentrations are much lower and do not increase after veraison compared to the Riesling samples.

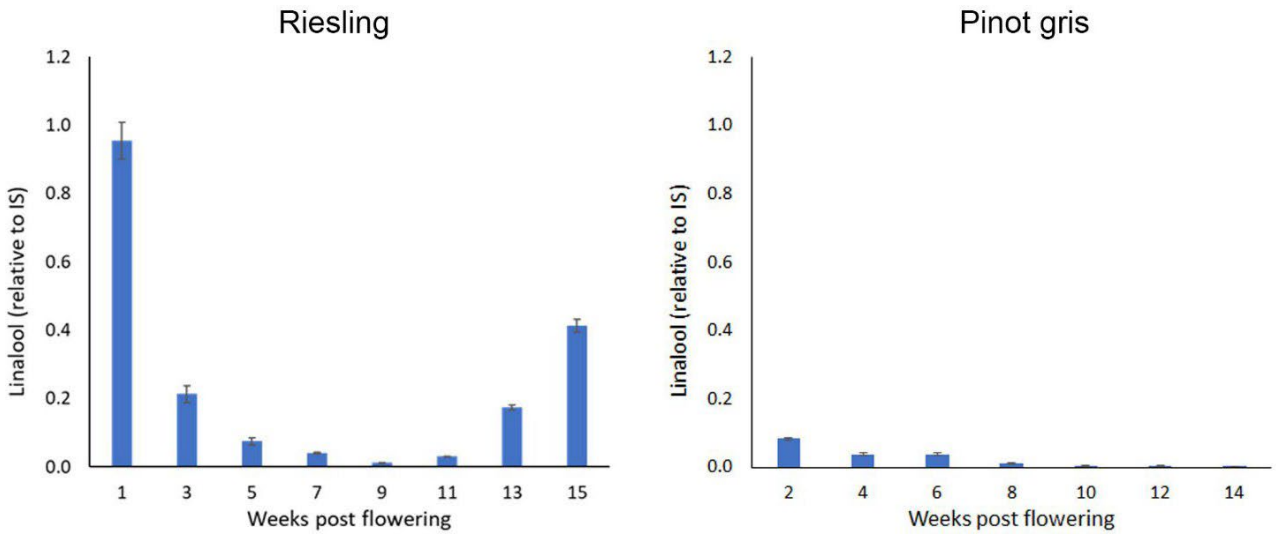


Figure 74 Concentrations of free linalool in the Riesling and Pinot gris berry developmental samples taken from Eden Valley in the 2018/19 vintage.

Expression of *VvNUDX1* and *VvNUDX25* were tested in the same samples used for the GC-MS analysis. No expression of *VvNUDX1* was detected at any stage of berry development (data not shown). In contrast, *VvNUDX25* was highly expressed in both Riesling and Pinot noir (Figure 75). The patterns of expression were different between the cultivars with Pinot gris having a decrease in expression pre-veraison that is not so marked in Riesling. In general, the expression suggests that there is neither a positive nor negative role for this gene in linalool production in Riesling.

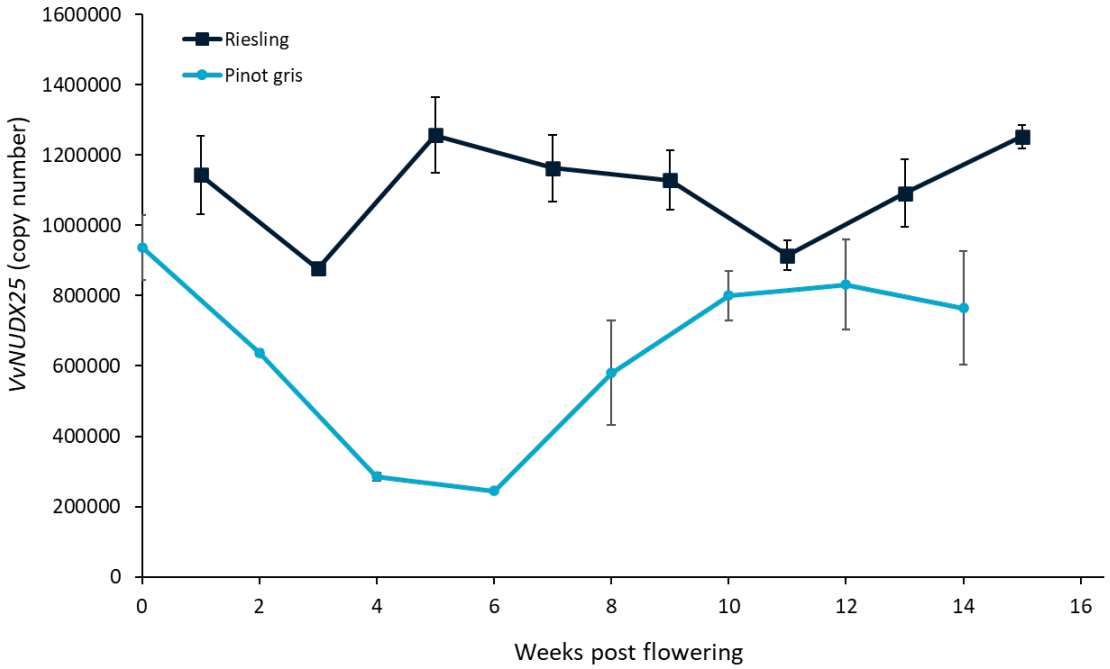


Figure 75 Expression of the *VvNUDX25* gene in the Riesling and Pinot gris berry samples from throughout development in the 2018/19 growing season.

Sequence analysis of the Riesling and Pinot noir *VvNUDX25* genes

The *VvNUDX25* gene is just over 5.2 kb in size. However, the coding region of the mature mRNA is only 522 bp, made up from six exons in the gene and encoding a protein of 173 amino acids. These exons are separated by some large introns; for example, the first intron between exon 1 and 2 is just under 1.8 kb in length. Full genome sequences for Pinot noir and Riesling are now available on the website of Dario Cantu's group at UC Davis (www.grapegenomics.com) which allowed us to compare the *VvNUDX25* alleles in both cultivars. The coding regions of the Pinot noir alleles are identical, and the Riesling alleles are also identical to those from Pinot noir except for a single base difference in one allele, which does not change the amino acid residue encoded by that codon (data not shown). Therefore, differences in enzyme activity due to amino acid differences between Riesling and the microvine are not likely. The genes were remarkably conserved outside of the coding sequence, with ~1 kb upstream and downstream of the translation start site being almost identical between the two Pinot alleles and one of the Riesling alleles (99.9% identical across the 7.2 kb). One Riesling allele was somewhat different to the other three but being ~98.8% identical to the others means the gene is highly conserved with suggests an important functional role in the plant.

Summary

The conclusions from this part of the project are that the function of the *VvNUDX25* gene and its position relative to the predicted genome region of interest for the phenotype suggest it could be involved in regulating monoterpene production in berries. However, the sequence data and expression analysis, especially when comparing Riesling and Pinot varieties did not present any causal differences that might explain a difference in phenotype between these varieties. It should be remembered that we are using the Pinot noir genome sequence and Pinot gris developmental series as proxies for the microvine, which is derived from Pinot Meunier. It is possible that differences exist between these 'sports' of Pinot noir and the microvine. Furthermore, it is possible that the crossover itself in plant #131, which appears to have occurred extremely close to the *VvNUDX25* gene, could have disrupted the expression and/or activity of this gene. This requires more study. However, such differences still would not explain the differences between the ability of Riesling and the microvine to produce monoterpenes. It was decided that plant #131 might be a good tool to use to identify QTLs responsible for moderate production of monoterpenes as observed in Riesling by producing a F2 population by self-pollinating this plant and this is outlined in section 6.4.1 below.

6.3.4 Candidate genes in the QTL associated with wine acetate ester concentrations

A gene list from this region identifies several potential candidate genes

The QTL analysis of the MV × R and Mv × CS biparental crosses identified a region near the top of chromosome 18 associated with the concentration of acetate esters in the small-scale wines made from the berries harvested from the progeny. A total of 188 annotated genes were present in this region (Table 15).

Table 15 List of genes underlying the QTL for wine acetate ester production identified in the Mv × CS and Mv × R biparental crosses.

NUMBER	LOCUS	DESCRIPTION	NUMBER	LOCUS	DESCRIPTION
1	LOC100264280	N-Glycosylase/DNA lyase OGG1	95	LOC100264291	Mitogen-activated protein kinase kinase kinase
2	LOC100262510	Beta-galactosidase 16	96	LOC100262519	Mediator of RNA polymerase II transcription subunit 1
3	LOC100250625	Proteasome subunit beta type-like	97	LOC100242038	Transmembrane protein 50 homolog
4	LOC100854126	E3 ubiquitin ligase BIG BROTHER-related	98	LOC100245365	Unknown
5	LOC100255807	NAC domain-containing protein 2	99	LOC100247177	Histone acetyltransferase (MYST)
6	LOC100260874	Cellular synthase interactive	100	LOC100853421	Unknown
7	LOC100266082	Replication factor C subunit 3	101	LOC100252293	Vacuolar protein sorting-associated protein 9A
8	LOC100248995	Kinesin-like protein KIN-7C	102	LOC100257428	Beta-1,3-galactosyltransferase
9	LOC100243763	Aspartic proteinase Asp1	103	LOC100262554	Eukaryotic translation initiation factor 5
10	LOC100248915	Aldehyde dehydrogenase	104	LOC100255673	Auxin response factor 17
11	LOC100254009	L-Ascorbate peroxidase 6	105	LOC100260744	protein GrpE
12	LOC100254101	Nitrate regulatory gene2	106	LOC104882474	Unknown
13	LOC100259138	Protein disulfide-isomerase	107	LOC100265954	Pentatricopeptide repeat-containing protein
14	LOC100259216	Two-component response regulator ORR9	108	LOC100249007	Unknown
15	LOC100264314	Delta-aminolevulinic acid dehydratase	109	LOC104882475	Calcium-binding and coiled-coil domain-containing protein
16	LOC100264396	IAA-amino acid hydrolase ILR1-like 6	110	LOC100267713	Serine hydroxymethyltransferase
17	LOC100242108	Caffeic acid 3-O-methyltransferase	111	LOC100245399	BPS1
18	LOC100247171	F-box protein 7	112	TRNAR-UCU	transfer RNA arginine
19	LOC100242031	Acetyltransferase	113	LOC100259231	RHOMBOID-like protein 8
20	LOC100252288	V-type proton ATPase subunit c1	114	TRNAW-CCA	transfer RNA tryptophan
21	LOC100257421	BTB/POZ domain-containing protein	115	LOC100264408	Phosphopantothenate--cysteine ligase 2
22	LOC100262547	Zinc ribbon 15 superfamily	116	LOC100250529	Chaperone protein DnaJ
23	LOC100247245	Chlorophyllide a oxygenase	117	LOC100255706	Transcription factor TGA4
24	LOC100267707	TLC domain-containing protein 2	118	LOC100242121	ABSCISIC ACID-INSENSITIVE 5
25	TRNAQ-UUG	Transfer RNA glutamine	119	LOC100247254	Retrotransposon-like
26	TRNAM-CAU	Transfer RNA methionine	120	LOC100260778	DNA-directed RNA polymerase II
27	LOC100252368	Filament-like plant protein	121	LOC100265983	Unknown

NUMBER	LOCUS	DESCRIPTION	NUMBER	LOCUS	DESCRIPTION
28	LOC100245393	Photosystem II 22 kDa protein	122	LOC100252377	Pentatricopeptide repeat-containing protein
29	LOC100250521	Unknown	123	LOC100257505	Glutaredoxin
30	LOC100262628	Esterase	124	LOC100262638	Alkylated DNA repair protein
31	LOC104882465	Esterase	125	LOC100267796	Aminotransferase ALD1
32	LOC109121704	Esterase	126	LOC100255792	Centromere-associated protein E
33	LOC100260772	Unknown	127	LOC100243670	Membrane-anchored ubiquitin-fold protein 3
34	LOC100265978	Long chain acyl-CoA synthetase	128	LOC100248808	F-box/kelch-repeat protein
35	LOC100250597	NEDD1 – WD40 repeat	129	LOC100260859	Unknown
36	LOC100255781	sister chromatid cohesion protein PDS5 homolog A	130	LOC104882477	Unknown
37	LOC100260849	Unknown	131	LOC100253901	60S Ribosomal protein L30-1
38	LOC100243662	UDP-galactose transporter	132	LOC100248897	Apoptotic chromatin condensation inducer
39	LOC100248801	Universal stress protein family	133	LOC100253991	Aspartic proteinase
40	LOC104882466	AAA protein family	134	LOC100259038	Heavy metal-associated isoprenylated plant protein
41	LOC100253893	lysM Domain-containing GPI-anchored protein	135	LOC100852940	Unknown
42	LOC104882467	Aspartic proteinase nepenthesin-1	136	LOC109121591	MADS-box protein
43	LOC100253978	RNA polymerase II transcription factor B subunit 4	137	LOC104882592	MADS-box protein
44	LOC100259029	RRP15-like	138	LOC104882593	Unknown
45	LOC100259114	Glutamic acid-rich protein	139	LOC100242012	Unknown
46	LOC100264430	Gamma-glutamylcyclotransferase	140	LOC104882479	Bifunctional aspartate aminotransferase and glutamate/aspartate-prephenate aminotransferase
47	LOC100242142	WAT1-like transporter	141	LOC100247155	Zinc finger CCCH domain
48	LOC100264284	WAT1-like transporter	142	LOC100252269	Low affinity sulfate transporter 3
49	LOC100242002	WAT1-like transporter	143	LOC100233023	Sulfate transporter
50	LOC100247142	DUF836; Glutaredoxin domain	144	LOC109121687	Sulfate transporter
51	LOC100247281	Transmembrane emp24 domain-containing protein	145	LOC109121688	Sulfate transporter
52	LOC100252402	Eukaryotic translation initiation factor 1A	146	LOC109121573	Unknown
53	LOC100257526	Aminoacylase	147	LOC100242149	tRNA modification GTPase MnmE
54	LOC100853357	Ethylene-responsive transcription factor ERF014-like	148	LOC100247288	Acetyl-CoA carboxylase 1
55	LOC109121696	Non-coding RNA (ncRNA)	149	LOC100252409	Unknown
56	LOC100252259	Prolyl-tRNA synthetase associated domain-containing protein	150	LOC100257534	Glucan 1,3-beta-glucosidase A
57	LOC100262515	E3 ubiquitin-protein ligase RING1	151	LOC100854472	Pollen-specific protein C13

NUMBER	LOCUS	DESCRIPTION	NUMBER	LOCUS	DESCRIPTION
58	LOC100245359	Phosphatidylinositol 4-phosphate 5-kinase 5	152	LOC100245371	2,3-Bisphosphoglycerate-dependent phosphoglycerate mutase
59	LOC104882468	Aspartate aminotransferase	153	LOC100250501	Random slug protein 5
60	LOC100250491	Nodulation-signaling pathway 2 protein (GRAS) **miRNA171g in reverse orientation**	154	LOC100255680	ATP-dependent helicase
61	LOC100262661	Unknown	155	LOC100260752	Metallophosphoesterase
62	LOC100855011	Unknown	156	LOC100265962	Beta-D-xylosidase 7
63	LOC100260740	DIS3-like exonuclease	157	LOC100262668	Beta-D-xylosidase 7
64	LOC100265950	G-type lectin S-receptor serine/threonine-protein kinase	158	LOC100267825	WD repeat-containing protein
65	Not on NCBI	Penicillin-binding protein	159	LOC100245515	Ethylene-responsive transcription factor
66	Not on NCBI	Photosystem II CP43 protein	160	LOC100250637	Endoplasmic reticulum-Golgi intermediate compartment protein 3
67	LOC100267819	Flavonol synthase	161	LOC100243648	Trehalose-phosphate phosphatase
68	LOC100267262	Reticulon-like protein B2	162	LOC104882481	NDR1/HIN1-like protein
69	LOC100250629	Selenoprotein K	163	LOC100254118	F-box protein
70	FLS1	Flavonol synthase	164	LOC104882482	Transcription termination factor
71	LOC100232939	Flavonol synthase	165	LOC100264413	RNA-directed DNA methylation
72	LOC100243852	Flavonol synthase	166	LOC100242128	Thioredoxin TTL1
73	LOC100249002	Flavonol synthase	167	LOC104882483	VQ motif-containing protein
74	LOC100254107	CrcB-like protein	168	LOC100255818	Spinster homolog 3
75	LOC100266087	Glycine-rich protein 2b	169	LOC100252381	Methyltransferase
76	LOC100243769	Auxin transporter-like protein	170	TRNAK-CUU	Transfer RNA lysine
77	LOC100259225	Thaumatin-like protein	171	LOC100260886	Unknown
78	LOC100264401	Ubiquitin-fold modifier	172	LOC100257510	Pumilio-family RNA binding
79	LOC100242116	DUF246, O-fucosyltransferase	173	LOC100266093	Annexin D1-like
80	LOC100247249	Pentatricopeptide repeat-containing protein	174	LOC109121642	Unknown
81	LOC100257501	Zinc finger	175	LOC100243776	Serine/arginine-rich splicing factor
82	LOC100262634	Dual specificity protein kinase	176	LOC100232961	Hexose transporter HT2
83	LOC100267791	Histone-lysine N-methyltransferase	177	LOC100262645	Unknown
84	LOC100248922	Biotin--protein ligase	178	LOC100267801	Carbon catabolite repressor protein 4
85	LOC100245477	Oxysterol-binding protein-related protein 1C	179	LOC100250613	Synaptonemal complex protein 2
86	LOC100250604	Phosphatidylinositol 4-phosphate 5-kinase	180	LOC100255796	Mitochondrial adenine nucleotide transporter
87	MIR394C	microRNA MIR394c	181	LOC100260863	N-terminal cysteine-rich region

NUMBER	LOCUS	DESCRIPTION	NUMBER	LOCUS	DESCRIPTION
88	LOC100855348	Terminal ear1 homolog	182	LOC100266070	Signal peptidase complex-like protein
89	LOC100254016	30S Ribosomal protein S13	183	LOC100243754	MYB39
90	LOC104882590	Keratin, type I cytoskeletal	184	LOC100254022	Cryptochrome-1
91	LOC100259145	Calcium-binding protein CML29	185	LOC100259151	Protein phosphatase 2C
92	LOC100264320	Nitrate reductase [NADH]-like	186	LOC109121672	Unknown
93	LOC104882472	Unknown	187	LOC100242043	Monoacylglycerol lipase ABHD6
94	LOC100259117	Unknown	188	LOC100247183	14-3-3 protein

It was difficult to predict what type of gene may be involved in changing the composition of the grape must to then influence the acetate ester concentration in the wines. This in turn makes candidate gene selection difficult as well. Because all three of the main acetate esters mapped to the same region, we predicted that the causal gene did not have a role in the production of the alcohol precursors, but perhaps involved something the yeast requires for esterification. Three initial candidate genes from the region were chosen (highlighted blue in Table 15) based on the proposed function based on the genome annotation and their role in nitrogen metabolism or coenzyme-A (CoA) synthesis. The first chosen was a WAT1-like transporter (LOC100264284), because this family of genes encode a number of transporters termed 'usually multiple acids move in and out transporters' (UMAMIT), many of which transport amino acids (Yao, et al. 2020). The second was a phosphopantothenate-cysteine ligase (LOC100264408) because of its role in the production of CoA (Leonardi and Jackowski 2007), which is essential for ester production by yeast. The third targeted gene encodes a bifunctional aspartate aminotransferase and glutamate/aspartate-prephenate aminotransferase (LOC104882479). These enzymes are thought to regulate carbon and nitrogen metabolism in all organisms (de la Torre, et al. 2014) and as such may play a key role in the nitrogen composition of the berry, which in turn will influence yeast metabolism.

LOC100264284

Primers were designed to amplify the whole genomic region covering the coding region of this gene in Mv, CS and R, which was just under 2.3 kb in length. The amplified bands were cloned and sequenced across the entire length to identify SNPs and especially those that might result in functional differences due to a change in amino acid sequence. The sequences were very similar with one allele being identical in all three varieties and the other allele being 97.9-99.9 % identical. Nevertheless, a SNP was detected in the fourth exon which removed an *Afl*III restriction enzyme site in one of the CS and R alleles and primers were designed to amplify a region containing this SNP to use as a molecular marker. The marker was first tested in the Mv × R cross to confirm the marker trait association in this region of the genome in this biparental cross. As the marker was homozygous in Mv and heterozygous in R, there were only two genotypes observed and the concentration of phenylethyl acetate in the plants with the 'kk' genotype was generally greater than that seen in those with the 'hk' genotype (Figure 76). In contrast, in the variety collection, the difference in phenylethyl acetate concentration among the genotypes was not so evident. Although the median for the 'kk' genotype is somewhat higher than the others, the means are very similar (Figure 76).

This suggests that this SNP is not causing the phenotypic difference and is not closely linked to any potential casual gene in the varieties. This is not surprising given the rapid decay of linkage disequilibrium in grapevine.

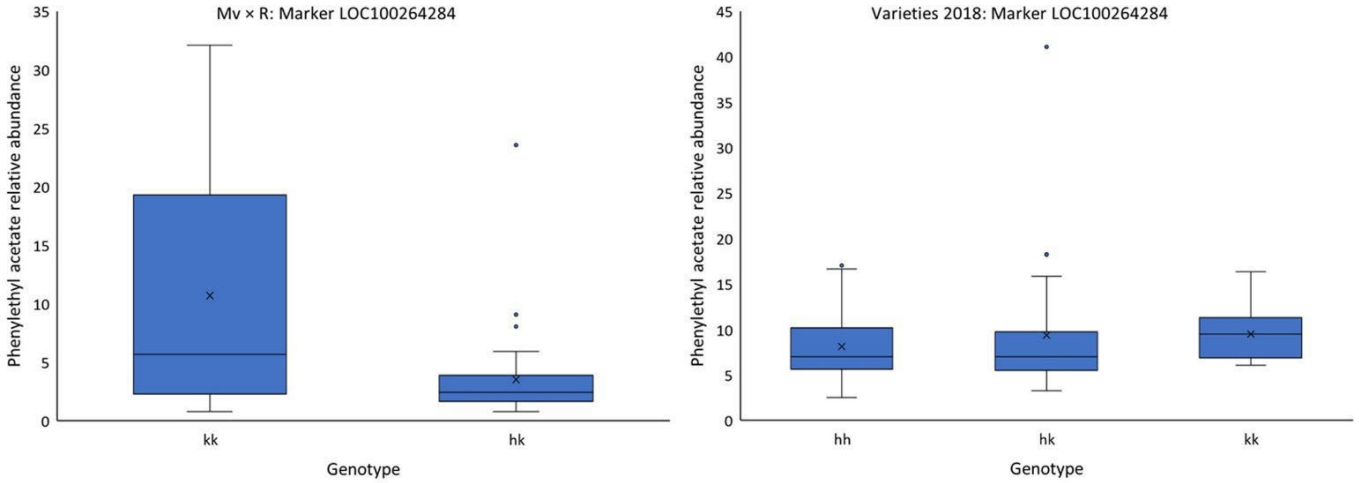


Figure 76 Box plots of the marker designed to a SNP in LOC100264284, a candidate gene for acetate ester production, plotting phenylethyl acetate concentration in wine versus plant genotype.

LOC100264408

This gene was our primary target given its role in the production of CoA and its precursors in the berry, which could be a source of these compounds for the yeast. The gene is quite complex comprising 10 exons (one non-coding at the start) and covering 5.5 kb of the genome. Primers were designed to amplify six overlapping regions of the gene and sequences were compared to identify SNPs that caused amino acid changes in the protein sequence. Two in the first exon were used to design markers and these were tested across the Mv × R progeny and the 2018 samples from the variety collection. The first of these was an A to T polymorphism which was present in one allele of Mv and R, and which causes an isoleucine to phenylalanine amino acid change in the protein. The SNP also removed an *Eco*RV restriction enzyme site, which allowed the easy design of a marker. The marker was first tested in the Mv × R cross and clearly showed a marker trait association with the plants containing the 'hh' genotype having much higher concentrations of phenylethyl acetate than the 'hk' and 'kk' genotypes (Figure 77). In contrast, this marker did not appear to be predictive of wine phenylethyl acetate esters in the varieties tested in the 2018 vintage. In fact, varieties with the 'hh' allele appeared to have slightly lower phenylethyl acetate concentrations than those with other genotypes (Figure 77). A second polymorphism in this gene was detected in both Mv and R varieties. The SNP converted a 'C' to a 'T' in the DNA sequence, which converted a serine to a phenylalanine in the amino acid sequence and introduced a *Tfi*I restriction enzyme site into the sequence. The genotypes from the Mv × R progeny were identical to the other marker in this gene, but 12 of the 89 varieties sampled in 2018 for which we had genotypic data from had different genotypes at the two SNPS identified in this gene. However, these differences did not significantly change the association between the SNP in this gene and the concentration of phenylethyl acetate in the wines made form the different varieties (Figure 77). This gene remains a good candidate but will require a more in-depth investigation of the genetic variants in *vinifera* cultivars to determine if it is predictive

beyond the Mv crosses analysed in this project.

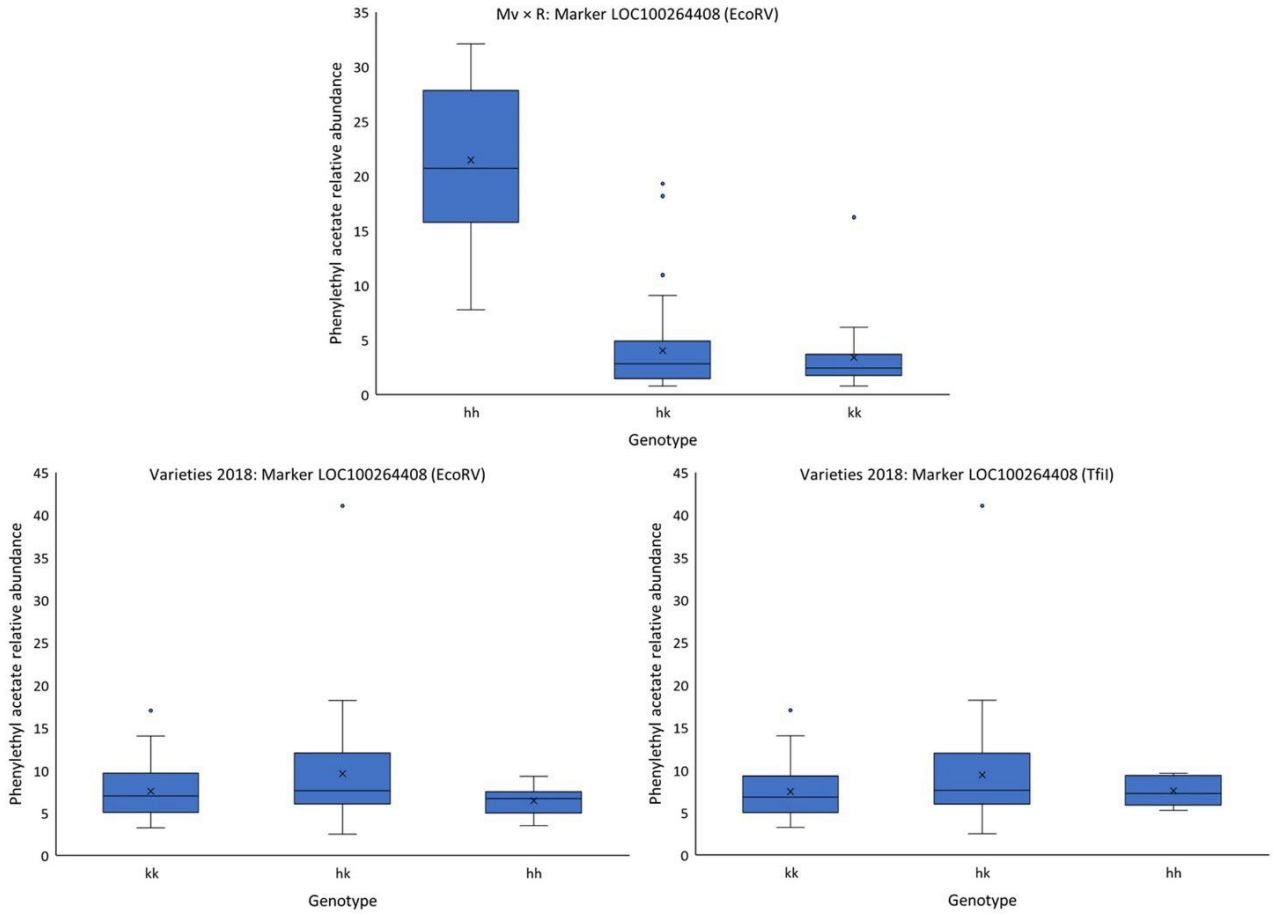


Figure 77 Box plots of two markers designed to a SNP in LOC100264408, a candidate gene for acetate ester production, plotting phenylethyl acetate concentration in wine versus plant genotype.

LOC104882479

This bifunctional aspartate aminotransferase and glutamate/aspartate-prephenate aminotransferase gene was examined for the presence of polymorphisms that may cause differences in nitrogen sources in the berries and alter acetate ester production by yeast. Sequence analysis of the genomic region covering this gene amplified from Mv, CS and R showed the presence of a SNP in one allele in Mv and R. The 'A' to 'T' nucleotide change results in a tyrosine residue in the protein being converted into a cysteine. The SNP also introduces a *Bsm*FI restriction enzyme site into the sequence that can be used to develop a marker for this polymorphism. The marker was tested in the Mv × R F1 population to see if it showed an association with the acetate ester phenotype. There was clearly an association, with the 'kk' genotype correlating with higher acetate ester production in the wines compared to the 'hk' and 'hh' genotypes (Figure 78). However, this association was not apparent when the genotype/phenotype association was tested in the varieties sampled in 2018 (Figure 78).

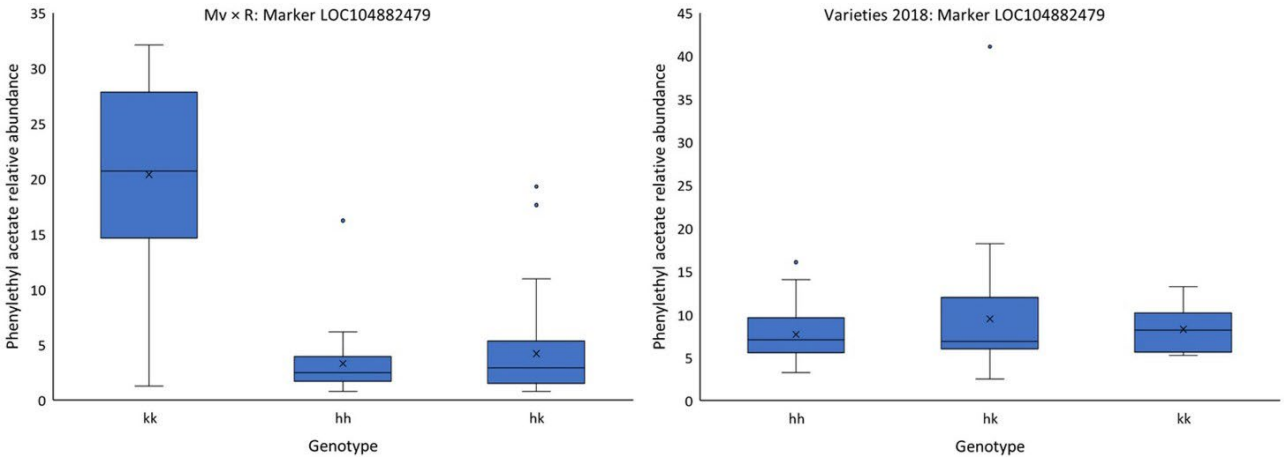


Figure 78 Box plots of the marker designed to a SNP in LOC104882479, a candidate gene for acetate ester production, plotting phenylethyl acetate concentration in wine versus plant genotype.

Summary

There was an obvious marker:trait relationship in the Mv × R F1 population with regard to the markers designed to the three candidate genes identified in the genomic region associated with acetate ester concentrations in wine. The strength of this relationship in both the Mv × R and Mv × CS F1 populations was an exciting discovery that gives us an opportunity to understand how grape composition can influence the production of these important yeast-derived aroma compounds. However, our subsequent attempts to determine the causal gene for this trait were not successful. The two paths followed were: a) the GWAS analysis of the variety collections described in section 6.2.2; b) the identification of candidate genes in the QTL and the testing of SNPs for marker:trait association in the 2018 variety collection samples described in this section. The GWAS did not detect any QTLs overlapping with the region identified in the biparental crosses, but rapid linkage disequilibrium may explain this given the number of markers was limited. The candidate gene approach also failed to detect a significant association between acetate ester concentration in the varietal wines and SNPs in the chosen genes. It is possible that this trait mapped in the biparental cross because the presence of a rare allele with a major effect on the phenotype in Mv (and perhaps R and CS as well), which is not evident in the other varieties. This would mean these approaches using the variety collection cannot be used to determine the causal gene. The production of the F2 population from the self-pollinated F1 Mv × R plant #131 will allow this trait to be further studied in these genetic backgrounds and may narrow down the QTL region given the density of the map is greater for this population (see section 6.4.1).

6.3.5 Candidate genes in the QTL associated with wine β-damascenone concentrations

List of genes in the QTL region mapped in the Mv × CS F1 population

Biparental mapping revealed the presence of a single major QTL for β-damascenone in the Mv × CS F1 population (section 6.1.3). The QTL was located on chromosome 2 and covered the region from 4.9 Mbp from the top of the chromosome to 5.4 Mbp. This region contained 54 genes as annotated on the NCBI version of the PN42004 genome sequence and these are listed in Table 16.

Table 16 List of genes underlying the QTL for β-damascenone production identified in the Mv × CS biparental cross.

NUMBER	LOCUS	DESCRIPTION	NUMBER	LOCUS	DESCRIPTION
1	LOC100267708	YABBY	28	LOC100250615	L-Lactate dehydrogenase A
2	LOC100262548	Monocopper oxidase-like protein	29	LOC100255799	Unknown (deacylase domain)
3	LOC100241232	Acid β-fructofuranosidase	30	LOC100855013	Root meristem GF6-like
4	LOC100257422	Transaldolase	31	LOC100260865	Autophagy-related protein 8f
5	LOC100852434	Trehalose-phosphate phosphatase	32	LOC100252271	Scarecrow-like protein
6	LOC100852463	INAPERTURATE POLLEN1	33	LOC100262529	Unknown (lectin receptor)
7	LOC100854486	Xyloglucan galactosyltransferase	34	LOC104882157	Unknown
8	LOC100854507	3-Oxoacyl-[acyl-carrier-protein] synthase, chloroplastic	35	LOC100255681	DEAD-box ATP-dependent RNA helicase
9	LOC100852507	PLATZ transcription factor	36	LOC104877275	Leukocyte receptor cluster member 1 homolog
10	TRNAG-GCC	Transfer RNA glycine	37	LOC100266073	Unknown (Heavy metal -associated domain)
11	LOC100852547	Flavin-containing monooxygenase	38	LOC100243757	23.6 kDa Heat shock protein, mitochondrial
12	LOC100852585	Flavin-containing monooxygenase	39	LOC100248906	Small heat shock protein, chloroplastic
13	LOC100852623	Flavin-containing monooxygenase	40	LOC100253999	Vesicle-associated protein 1-1
14	LOC100255688	Flavin-containing monooxygenase	41	LOC100260753	Phosphatidylinositol/phosphatidylcholine transfer protein SFH11
15	LOC100852661	Unknown function (DUF863)	42	LOC100852959	Aspartyl protease
16	LOC100257409	Translation initiation factor IF-2	43	LOC100265964	Bifunctional adenosine 5'-phosphosulfate phosphorylase/adenylsulfatase HINT4
17	LOC100259139	Lipid phosphate phosphatase	44	LOC100243649	Arginine N-methyltransferase
18	LOC100253994	Oxysterol-binding protein-related protein	45	LOC100259128	Chaperone protein dnaJ 13

NUMBER	LOCUS	DESCRIPTION	NUMBER	LOCUS	DESCRIPTION
19	LOC100252278	Oxysterol-binding protein-related protein	46	LOC104882149	Non-coding RNA
20	LOC100259122	NRT1/ PTR Family	47	LOC100855263	Unknown
21	LOC100247265	14 kDa Proline-rich protein	48	LOC100259241	Pectinesterase
22	LOC100252384	14 kDa Proline-rich protein	49	LOC100242021	Peroxisome biogenesis protein 19-2
23	LOC100257513	14 kDa Proline-rich protein	50	LOC104882152	Unknown
24	LOC100262648	36.4 kDa Proline-rich protein	51	LOC104882170	Plasma membrane ATPase 1
25	LOC100267804	14 kDa Proline-rich protein	52	LOC100249424	Palmitoyl-protein thioesterase 1
26	LOC100245490	NADH dehydrogenase	53	LOC100853058	Pentatricopeptide repeat-containing protein
27	LOC109123649	Non-coding RNA	54	LOC104882172	Unknown

There were no obvious candidate genes annotated as being directly involved in the production of carotenoids, which are the class of compounds from which β-damascenone is produced, nor in carotenoid cleavage. The carotenoids are produced in chloroplasts and the list does contain genes that have been implicated in the import of proteins into chloroplasts, namely the small heat shock proteins and chaperones. Interestingly, the DnaJ family of chaperones has been shown to influence carotenoid production in cauliflower (Lu, et al. 2006), and might be a good candidate. Pentatricopeptide repeat-containing proteins are mRNA binding proteins that influence the expression of organellar mRNA transcripts and have a potential role in norisoprenoid accumulation via carotenoid biosynthesis.

When this gene list was being collated, it became apparent that this locus also contains the genes that determine flower sex (Iocco-Corena, et al. 2021). The F1 population of the Mv × CS cross was segregating for flower sex as both parents are heterozygous and contain a copy of the dominant 'hermaphrodite' allele and the recessive 'female' allele. The berries from female flowers are very small and lack seeds but were also included in phenotyping via the small-scale fermentations. Previous research has shown that wines made with seeds have very low β-damascenone concentrations (Blackford, et al. 2022), and so removing this tissue from the fermentations may increase the abundance of this compound without any biochemical effect on biosynthesis pathways. To test the hypothesis that the berries from female plants would have higher concentrations of β-damascenone, the concentration of this compound was plotted against flower sex genotype (Figure 79). There is a clear relationship between the flower sex genotype and β-damascenone concentration, with the berries from female flowers having higher concentrations of this norisoprenoid. This does not rule out that there is a gene linked to the sex locus that influences β-damascenone production independent of berry size, but it is confounding. The same QTL was not identified in the Mv × R cross, where Mv is heterozygous and Riesling is homozygous for the hermaphrodite locus and so no females were present in the F1 population. Gene editing could be used to test the hypothesis about flower sex and berry size influencing β-damascenone accumulation by producing a plant with the PLATZ gene knocked-out and comparing with plants genetically identical except for this targeted mutation. If they do not show a phenotypic difference with regard to the β-damascenone concentrations then the causal gene is linked to the flower sex locus.

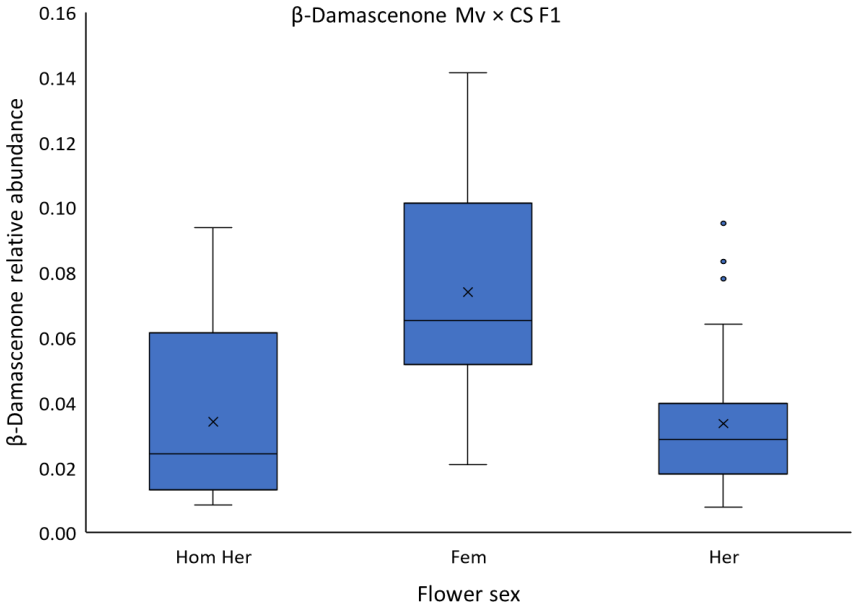


Figure 79 Box plots of β-damascenone concentration in the wines made from the Mv × CS progeny versus the flower sex genotype of the plants (Fem = female, Hom Her = homozygous hermaphrodite, Her = heterozygous hermaphrodite).

6.4 Generation and/or examination of other populations for gene mapping

6.4.1 F2 population derived from self-pollinated Mv × R F1 plant #131

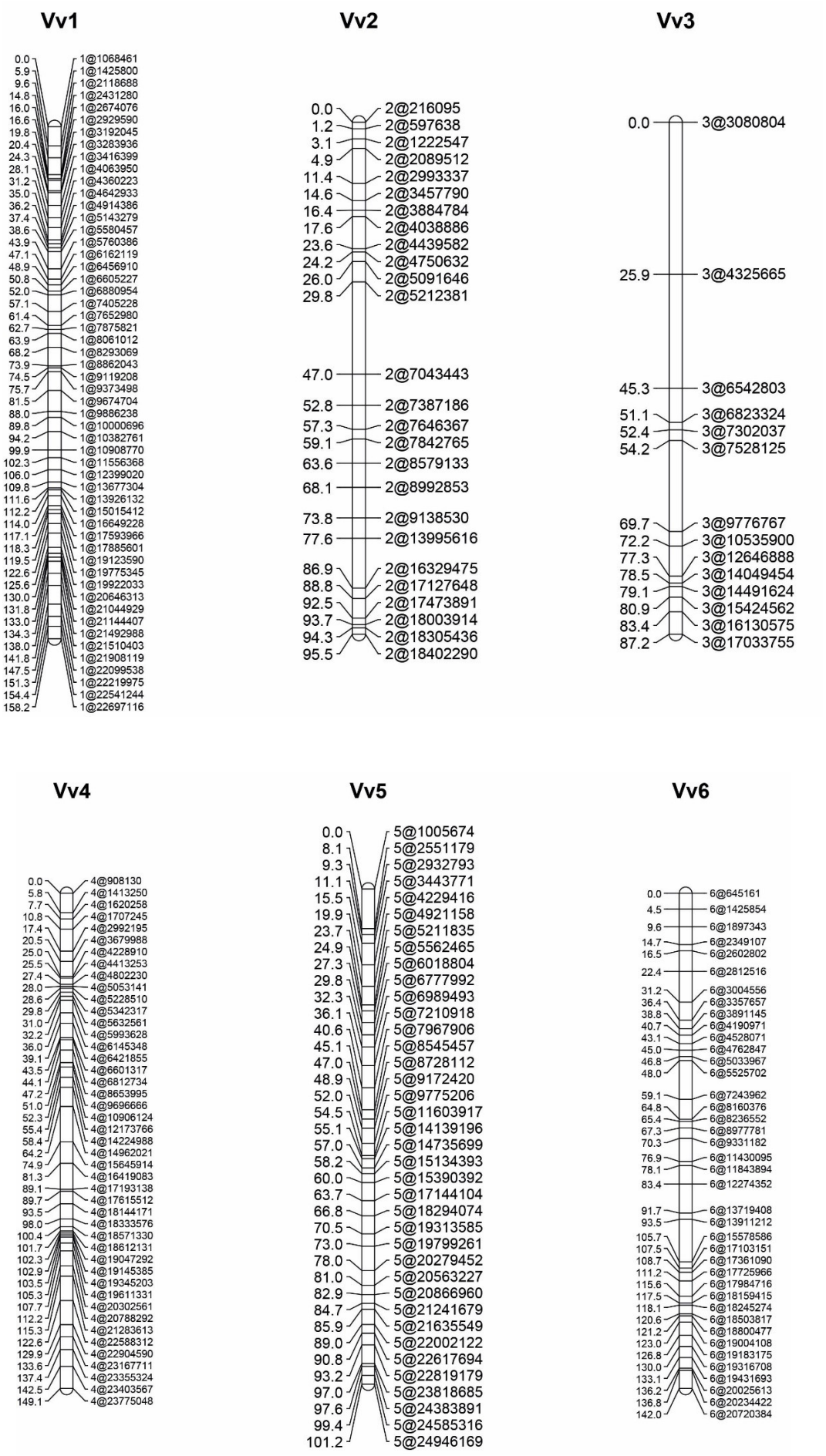
Introduction

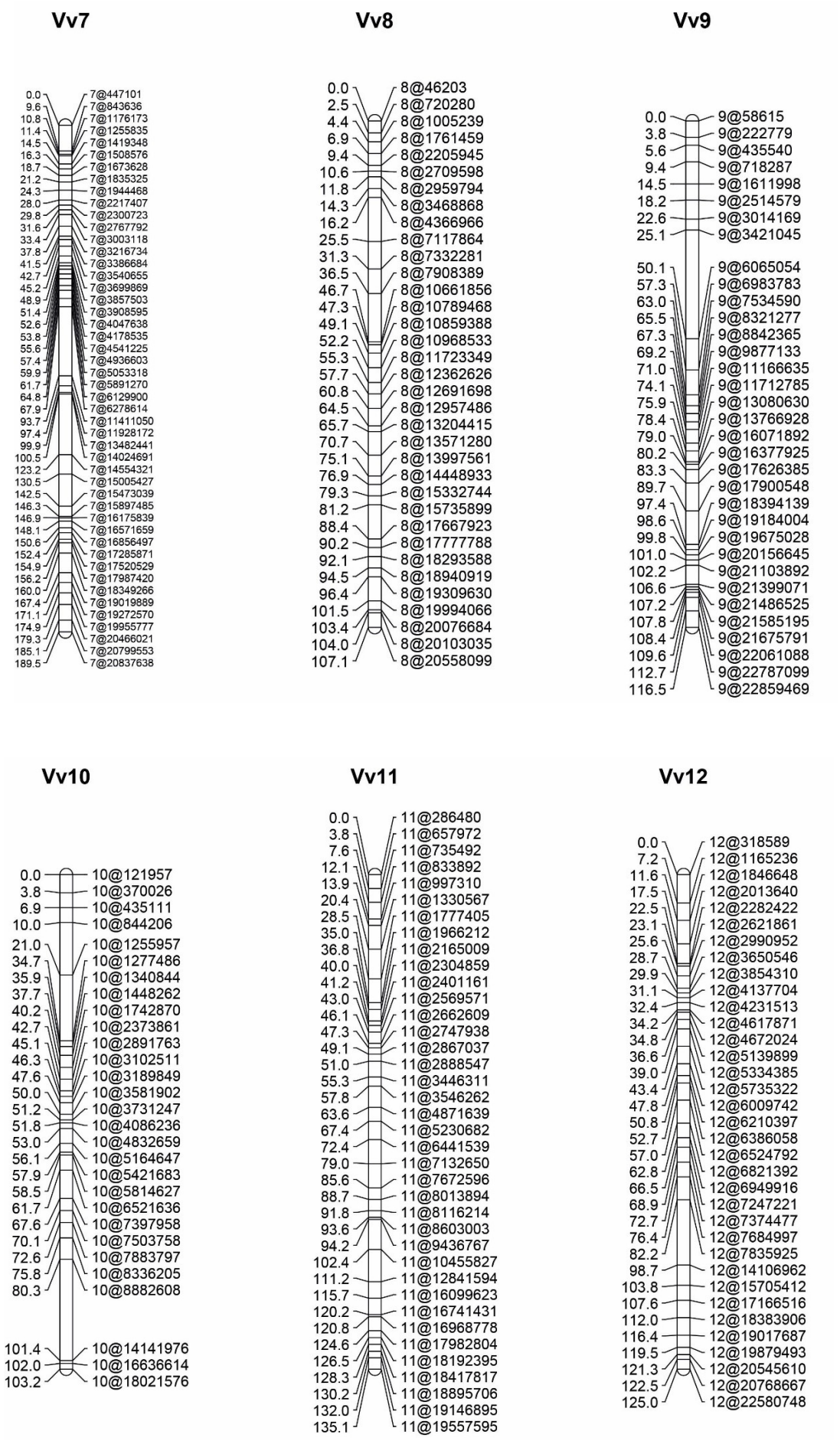
The analysis of linalool production in the Mv × R F1 population suggested that there was a genetic factor suppressing the accumulation of this compound in the berries of the progeny and that this factor was linked to the locus on chromosome 1 causing the dwarfing and early flowering phenotype (Boss and Thomas 2002). This meant that only one of the progeny, plant #131 (hereafter called Mv×R131), produced wine with concentrations of linalool that were similar to the R parent. However, the existence of other F1 progeny (e.g. plants #66 & 138, Figure 28) that had crossovers close to the dwarfing locus but did not have R-like concentrations of linalool suggested that there was a positive regulator of linalool production segregating. It was hypothesised that this should he heterozygous in Mv×R131 and, given that the suppressing factor from Mv was not present, the production of linalool should segregate in an F2 population derived from Mv×R131.

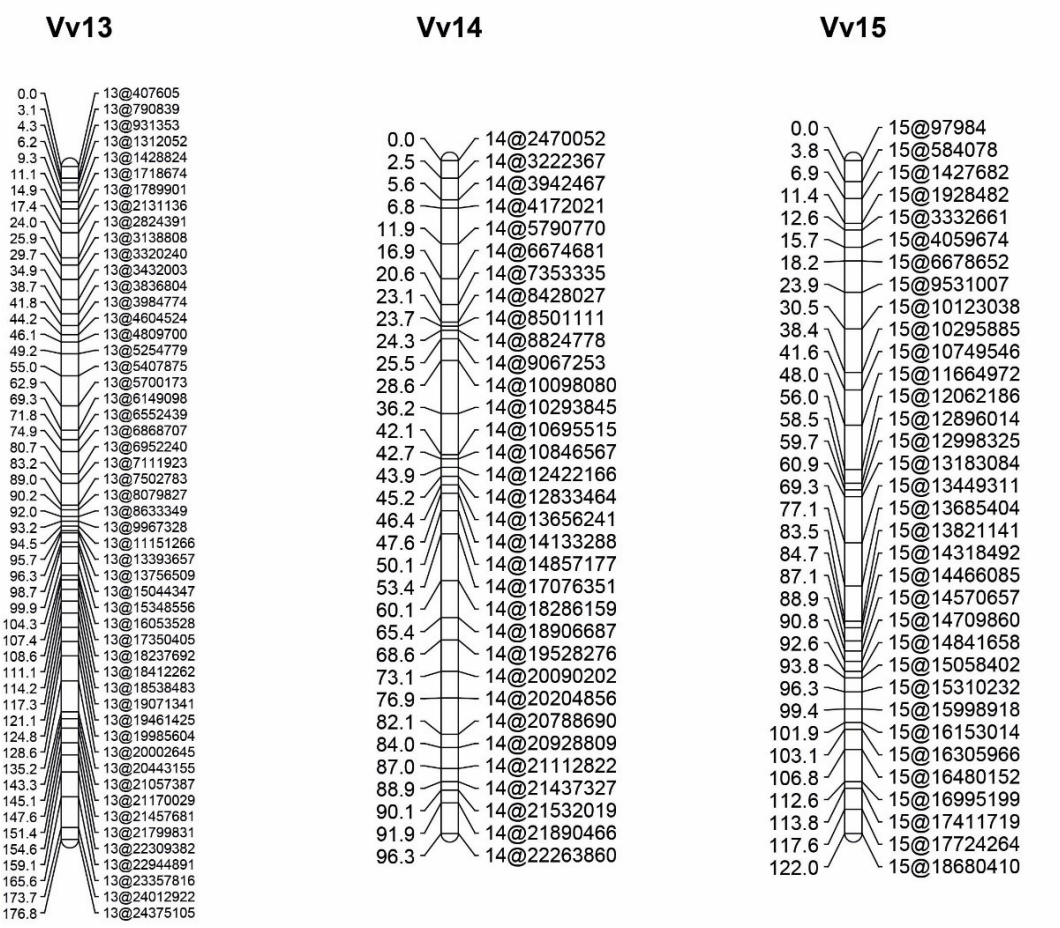
More Mv×R131 plants were propagated and seed collected from berries that we assumed to be self-pollinated. The seed were germinated and the semi-dwarf (Mv-like, heterozygous for the mutated *Vvgai* gene) and full dwarf (Picovines, homozygous for the mutated *Vvgai* gene) plants were selected, as they would be rapid flowering compared to wild type. A total of 84 plants were selected for the F2 population by the end of the project.

Generation of a genetic map

Leaf samples were taken from each of the F2 progeny and DNA produced by AGRF as described in the methods section. The DNA was then sent to Diversity Array Technologies (University of Canberra, Australia) and markers developed using the DArTseq technology. Markers were selected for the generation of the F2 genetic map if they had no misses in any of the progeny and if they segregated as ‘a’, ‘b’ and ‘ab’ genotypes. The resulting map was more dense than the F1 maps and included 734 markers across the 19 chromosomes with a total map length of 2484 cM with an average interval length of 3.5 cM (Figure 80). The number of markers per chromosome ranged from 14 to 59, with an average linkage group size of 131 cM.


Vv1
0.0 1@1068461
5.9 1@1425800
9.6 1@2118688
14.8 1@2431280
16.0 1@2674076
16.6 1@2929590
19.8 1@3192045
20.4 1@3283936
24.3 1@3416399
28.1 1@4063950
31.2 1@4360223
35.0 1@4642933
36.2 1@4914386
37.4 1@5143279
38.6 1@5580457
43.9 1@5760386
47.1 1@6162119
48.9 1@6456910
50.8 1@6605227
52.0 1@6880954
57.1 1@7405228
61.4 1@7652980
62.7 1@7875821
63.9 1@8061012
68.2 1@8293069
73.9 1@8862043
74.5 1@9119208
75.7 1@9373498
81.5 1@9674704
88.0 1@9886238
89.8 1@10000696
94.2 1@10382761
99.9 1@10908770
102.3 1@11556368
106.0 1@12399020
109.8 1@13677304
111.6 1@13926132
112.2 1@15015412
114.0 1@16649228
117.1 1@17593966
118.3 1@17885601
119.5 1@19123590
122.6 1@19775345
125.6 1@19922033
130.0 1@20646313
131.8 1@21044929
133.0 1@21144407
134.3 1@21492988
138.0 1@21510403
141.8 1@21908119
147.5 1@22099538
151.3 1@22219975
154.4 1@22541244
158.2 1@22697116
Vv2
0.0 2@216095
1.2 2@597638
3.1 2@1222547
4.9 2@2089512
11.4 2@2993337
14.6 2@3457790
16.4 2@3884784
17.6 2@4038886
23.6 2@4439582
24.2 2@4750632
26.0 2@5091646
29.8 2@5212381
47.0 2@7043443
52.8 2@7387186
57.3 2@7646367
59.1 2@7842765
63.6 2@8579133
68.1 2@8992853
73.8 2@9138530
77.6 2@13995616
86.9 2@16329475
88.8 2@17127648
92.5 2@17473891
93.7 2@18003914
94.3 2@18305436
95.5 2@18402290
Vv3
0.0 3@3080804
25.9 3@4325665
45.3 3@6542803
51.1 3@6823324
52.4 3@7302037
54.2 3@7528125
69.7 3@9776767
72.2 3@10535900
77.3 3@12646888
78.5 3@14049454
79.1 3@14491624
80.9 3@15424562
83.4 3@16130575
87.2 3@17033755
Vv4
0.0 4@908130
5.8 4@1413250
7.7 4@1620258
10.8 4@1707245
17.4 4@2992195
20.5 4@3679988
25.0 4@4228910
25.5 4@4413253
27.4 4@4802230
28.0 4@5053141
28.6 4@5228510
29.8 4@5342317
31.0 4@5632561
32.2 4@5993628
36.0 4@6145348
39.1 4@6421855
43.5 4@6601317
44.1 4@6812734
47.2 4@8653995
51.0 4@9696666
52.3 4@10906124
55.4 4@12173766
58.4 4@14224988
64.2 4@14962021
74.9 4@15645914
81.3 4@16419083
89.1 4@17193138
89.7 4@17615512
93.5 4@18144171
98.0 4@18333576
100.4 4@18571330
101.7 4@18612131
102.3 4@19047292
102.9 4@19145385
103.5 4@19345203
105.3 4@19611331
107.7 4@20302561
112.2 4@20788292
115.3 4@21283613
122.6 4@22588312
129.9 4@22904590
133.6 4@23167711
137.4 4@23355324
142.5 4@23403567
149.1 4@23775048
Vv5
0.0 5@1005674
8.1 5@2551179
9.3 5@2932793
11.1 5@3443771
15.5 5@4229416
19.9 5@4921158
23.7 5@5211835
24.9 5@5562465
27.3 5@6018804
29.8 5@6777992
32.3 5@6989493
36.1 5@7210918
40.6 5@7967906
45.1 5@8545457
47.0 5@8728112
48.9 5@9172420
52.0 5@9775206
54.5 5@11603917
55.1 5@14139196
57.0 5@14735699
58.2 5@15134393
60.0 5@15390392
63.7 5@17144104
66.8 5@18294074
70.5 5@19313585
73.0 5@19799261
78.0 5@20279452
81.0 5@20563227
82.9 5@20866960
84.7 5@21241679
85.9 5@21635549
89.0 5@22002122
90.8 5@22617694
93.2 5@22819179
97.0 5@23818685
97.6 5@24383891
99.4 5@24585316
101.2 5@24946169
Vv6
0.0 6@645161
4.5 6@1425854
9.6 6@1897343
14.7 6@2349107
16.5 6@2602802
22.4 6@2812516
31.2 6@3004556
36.4 6@3357657
38.8 6@3891145
40.7 6@4190971
43.1 6@4528071
45.0 6@4762847
46.8 6@5033967
48.0 6@5525702
59.1 6@7243962
64.8 6@8160376
65.4 6@8236552
67.3 6@8977781
70.3 6@9331182
76.9 6@11430095
78.1 6@11843894
83.4 6@12274352
91.7 6@13719408
93.5 6@13911212
105.7 6@15578586
107.5 6@17103151
108.7 6@17361090
111.2 6@17725966
115.6 6@17984716
117.5 6@18159415
118.1 6@18245274
120.6 6@18503817
121.2 6@18800477
123.0 6@19004108
126.8 6@19183175
130.0 6@19316708
133.1 6@19431693
136.2 6@20025613
136.8 6@20234422
142.0 6@20720384


Vv7
0.0 7@447101
9.6 7@843636
10.8 7@1176173
11.4 7@1255835
14.5 7@1419348
16.3 7@1508576
18.7 7@1673628
21.2 7@1835325
24.3 7@1944468
28.0 7@2217407
29.8 7@2300723
31.6 7@2767792
33.4 7@3003118
37.8 7@3216734
41.5 7@3386684
42.7 7@3540655
45.2 7@3699869
48.9 7@3857503
51.4 7@3908595
52.6 7@4047638
53.8 7@4178535
55.6 7@4541225
57.4 7@4936603
59.9 7@5053318
61.7 7@5891270
64.8 7@6129900
67.9 7@6278614
93.7 7@11411050
97.4 7@11928172
99.9 7@13482441
100.5 7@14024691
123.2 7@14554321
130.5 7@15005427
142.5 7@15473039
146.3 7@15897485
146.9 7@16175839
148.1 7@16571659
150.6 7@16856497
152.4 7@17285871
154.9 7@17520529
156.2 7@17987420
160.0 7@18349266
167.4 7@19019889
171.1 7@19272570
174.9 7@19955777
179.3 7@20466021
185.1 7@20799553
189.5 7@20837638
Vv8
0.0 8@46203
2.5 8@720280
4.4 8@1005239
6.9 8@1761459
9.4 8@2205945
10.6 8@2709598
11.8 8@2959794
14.3 8@3468868
16.2 8@4366966
25.5 8@7117864
31.3 8@7332281
36.5 8@7908389
46.7 8@10661856
47.3 8@10789468
49.1 8@10859388
52.2 8@10968533
55.3 8@11723349
57.7 8@12362626
60.8 8@12691698
64.5 8@12957486
65.7 8@13204415
70.7 8@13571280
75.1 8@13997561
76.9 8@14448933
79.3 8@15332744
81.2 8@15735899
88.4 8@17667923
90.2 8@17777788
92.1 8@18293588
94.5 8@18940919
96.4 8@19309630
101.5 8@19994066
103.4 8@20076684
104.0 8@20103035
107.1 8@20558099
Vv9
0.0 9@58615
3.8 9@222779
5.6 9@435540
9.4 9@718287
14.5 9@1611998
18.2 9@2514579
22.6 9@3014169
25.1 9@3421045
50.1 9@6065054
57.3 9@6983783
63.0 9@7534590
65.5 9@8321277
67.3 9@8842365
69.2 9@9877133
71.0 9@11166635
74.1 9@11712785
75.9 9@13080630
78.4 9@13766928
79.0 9@16071892
80.2 9@16377925
83.3 9@17626385
89.7 9@17900548
97.4 9@18394139
98.6 9@19184004
99.8 9@19675028
101.0 9@20156645
102.2 9@21103892
106.6 9@21399071
107.2 9@21486525
107.8 9@21585195
108.4 9@21675791
109.6 9@22061088
112.7 9@22787099
116.5 9@22859469
Vv10
0.0 10@121957
3.8 10@370026
6.9 10@435111
10.0 10@844206
21.0 10@1255957
34.7 10@1277486
35.9 10@1340844
37.7 10@1448262
40.2 10@1742870
42.7 10@2373861
45.1 10@2891763
46.3 10@3102511
47.6 10@3189849
50.0 10@3581902
51.2 10@3731247
51.8 10@4086236
53.0 10@4832659
56.1 10@5164647
57.9 10@5421683
58.5 10@5814627
61.7 10@6521636
67.6 10@7397958
70.1 10@7503758
72.6 10@7883797
75.8 10@8336205
80.3 10@8882608
101.4 10@14141976
102.0 10@16636614
103.2 10@18021576
Vv11
0.0 11@286480
3.8 11@657972
7.6 11@735492
12.1 11@833892
13.9 11@997310
20.4 11@1330567
28.5 11@1777405
35.0 11@1966212
36.8 11@2165009
40.0 11@2304859
41.2 11@2401161
43.0 11@2569571
46.1 11@2662609
47.3 11@2747938
49.1 11@2867037
51.0 11@2888547
55.3 11@3446311
57.8 11@3546262
63.6 11@4871639
67.4 11@5230682
72.4 11@6441539
79.0 11@7132650
85.6 11@7672596
88.7 11@8013894
91.8 11@8116214
93.6 11@8603003
94.2 11@9436767
102.4 11@10455827
111.2 11@12841594
115.7 11@16099623
120.2 11@16741431
120.8 11@16968778
124.6 11@17982804
126.5 11@18192395
128.3 11@18417817
130.2 11@18895706
132.0 11@19146895
135.1 11@19557595
Vv12
0.0 12@318589
7.2 12@1165236
11.6 12@1846648
17.5 12@2013640
22.5 12@2282422
23.1 12@2621861
25.6 12@2990952
28.7 12@3650546
29.9 12@3854310
31.1 12@4137704
32.4 12@4231513
34.2 12@4617871
34.8 12@4672024
36.6 12@5139899
39.0 12@5334385
43.4 12@5735322
47.8 12@6009742
50.8 12@6210397
52.7 12@6386058
57.0 12@6524792
62.8 12@6821392
66.5 12@6949916
68.9 12@7247221
72.7 12@7374477
76.4 12@7684997
82.2 12@7835925
98.7 12@14106962
103.8 12@15705412
107.6 12@17166516
112.0 12@18383906
116.4 12@19017687
119.5 12@19879493
121.3 12@20545610
122.5 12@20768667
125.0 12@22580748


Vv13
0.0 13@407605
3.1 13@790839
4.3 13@931353
6.2 13@1312052
9.3 13@1428824
11.1 13@1718674
14.9 13@1789901
17.4 13@2131136
24.0 13@2824391
25.9 13@3138808
29.7 13@3320240
34.9 13@3432003
38.7 13@3836804
41.8 13@3984774
44.2 13@4604524
46.1 13@4809700
49.2 13@5254779
55.0 13@5407875
62.9 13@5700173
69.3 13@6149098
71.8 13@6552439
74.9 13@6868707
80.7 13@6952240
83.2 13@7111923
89.0 13@7502783
90.2 13@8079827
92.0 13@8633349
93.2 13@9967328
94.5 13@11151266
95.7 13@13393657
96.3 13@13756509
98.7 13@15044347
99.9 13@15348556
104.3 13@16053528
107.4 13@17350405
108.6 13@18237692
111.1 13@18412262
114.2 13@18538483
117.3 13@19071341
121.1 13@19461425
124.8 13@19985604
128.6 13@20002645
135.2 13@20443155
143.3 13@21057387
145.1 13@21170029
147.6 13@21457681
151.4 13@21799831
154.6 13@22309382
159.1 13@22944891
165.6 13@23357816
173.7 13@24012922
176.8 13@24375105
Vv14
0.0 14@2470052
2.5 14@3222367
5.6 14@3942467
6.8 14@4172021
11.9 14@5790770
16.9 14@6674681
20.6 14@7353335
23.1 14@8428027
23.7 14@8501111
24.3 14@8824778
25.5 14@9067253
28.6 14@10098080
36.2 14@10293845
42.1 14@10695515
42.7 14@10846567
43.9 14@12422166
45.2 14@12833464
46.4 14@13656241
47.6 14@14133288
50.1 14@14857177
53.4 14@17076351
60.1 14@18286159
65.4 14@18906687
68.6 14@19528276
73.1 14@20090202
76.9 14@20204856
82.1 14@20788690
84.0 14@20928809
87.0 14@21112822
88.9 14@21437327
90.1 14@21532019
91.9 14@21890466
96.3 14@22263860
Vv15
0.0 15@97984
3.8 15@584078
6.9 15@1427682
11.4 15@1928482
12.6 15@3332661
15.7 15@4059674
18.2 15@6678652
23.9 15@9531007
30.5 15@10123038
38.4 15@10295885
41.6 15@10749546
48.0 15@11664972
56.0 15@12062186
58.5 15@12896014
59.7 15@12998325
60.9 15@13183084
69.3 15@13449311
77.1 15@13685404
83.5 15@13821141
84.7 15@14318492
87.1 15@14466085
88.9 15@14570657
90.8 15@14709860
92.6 15@14841658
93.8 15@15058402
96.3 15@15310232
99.4 15@15998918
101.9 15@16153014
103.1 15@16305966
106.8 15@16480152
112.6 15@16995199
113.8 15@17411719
117.6 15@17724264
122.0 15@18680410
Vv16
0.0 16@879736
3.8 16@1713124
8.2 16@2331447
9.4 16@2548762
11.3 16@2735179
15.1 16@2941906
20.2 16@5584746
24.0 16@6274948
26.4 16@6564457
30.9 16@7163412
34.0 16@9011955
37.8 16@10107537
40.3 16@12087169
47.6 16@13062360
57.3 16@13960502
62.4 16@14558033
68.3 16@15023693
71.5 16@15642348
76.0 16@16439360
79.8 16@16956177
84.9 16@17226202
90.8 16@17692327
97.4 16@17926848
101.2 16@18821042
103.8 16@19400411
108.1 16@20682410
110.0 16@21022791
111.9 16@21346844
116.3 16@21694612
124.4 16@21924685
Vv17
0.0 17@47210
2.5 17@192402
4.3 17@764364
6.8 17@879353
8.0 17@951740
9.2 17@1214204
11.1 17@1394961
13.6 17@1780871
16.9 17@2193838
20.8 17@2763473
22.7 17@2929247
27.3 17@3081857
33.4 17@3249084
37.4 17@3527395
40.6 17@3702177
44.5 17@3822747
48.4 17@4289782
50.2 17@4357161
52.7 17@4767480
53.9 17@4931567
55.8 17@5159272
58.2 17@5232561
59.4 17@5533460
60.6 17@5742205
62.5 17@5872114
63.7 17@6164838
64.3 17@6213317
66.7 17@6421704
68.6 17@6578678
73.0 17@6670753
76.1 17@6831049
81.9 17@7067084
85.7 17@7164915
93.0 17@7511929
96.1 17@7803476
97.3 17@7916250
99.8 17@8264956
103.6 17@8559188
108.1 17@9936579
111.2 17@10328570
113.7 17@10470436
115.5 17@10654301
118.0 17@10862390
119.8 17@10955154
121.1 17@11021315
125.6 17@11671974
127.4 17@11974848
129.9 17@12097298
131.7 17@14485994
137.5 17@14981438
144.0 17@15636776
147.1 17@16344044
153.7 17@16641612
Vv18 [1]
0.0 18@352361
3.8 18@640938
5.0 18@731835
6.2 18@923348
8.0 18@1448748
11.1 18@1952553
14.2 18@2287320
17.3 18@2448688
20.4 18@2718426
26.9 18@3410294
30.6 18@3956070
33.1 18@4335656
34.9 18@4405187
36.7 18@4512850
37.9 18@5080367
38.5 18@5275624
39.1 18@6313504
47.3 18@7099436
47.9 18@7190825
49.8 18@7535676
50.4 18@8062575
54.2 18@8478660
57.4 18@9355293
58.6 18@9926472
61.7 18@10180421
66.9 18@10415055
71.4 18@10871060
73.9 18@11360225
77.7 18@11618875
85.7 18@11791718
93.7 18@12174021
Vv18 [2]
97.5 18@12608267
98.7 18@12702041
99.9 18@13009684
101.2 18@13450343
102.4 18@13584629
104.8 18@13651140
108.7 18@13999098
111.2 18@14024088
115.0 18@14227411
120.2 18@18041353
121.4 18@18137534
124.6 18@18918398
129.1 18@19682011
134.3 18@20420593
136.7 18@20992401
139.2 18@21259968
147.1 18@22611011
156.5 18@24362834
159.6 18@24813832
166.1 18@25051292
174.9 18@27369450
177.4 18@27446751
179.3 18@27769004
182.4 18@28152619
185.5 18@28654237
187.3 18@28944005
188.5 18@29189936
Vv18 [3]
190.4 18@29246091

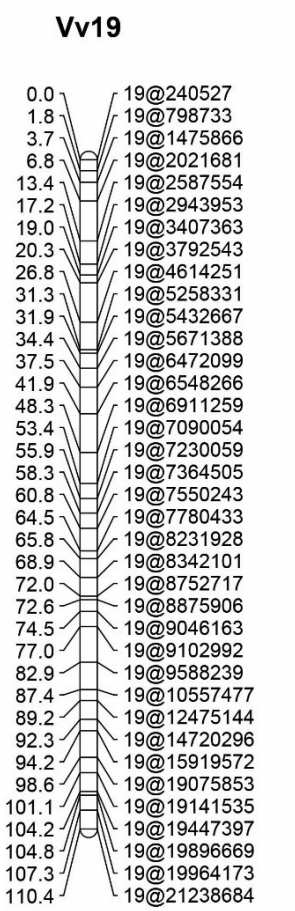


Figure 80 Linkage map of *Vitis vinifera* self-pollinated Mv×R131 F2 population. Linkage groups are numbered as originally assigned in Adam-Blondon et al. (2004) and only the consensus maps are shown. Distances of markers from the top are indicated on the left in cM.

QTL mapping of linalool concentration in the berries

The population was generated towards the end of the project and by the time the analysis had to start, samples had been obtained from 66 of the plants. As the main purpose of this population was to look for segregation of moderate linalool production, we decided to phenotype the progeny by analysing the volatile compound profiles of the berries using two extraction methods. One method conducted the extraction at 35 °C to mimic the temperature in the mouth, and the other was conducted at 95 °C to maximise volatilisation, although this does risk the production artifacts through chemical transformations at high temperature. A comparison of the two extraction techniques found some differences between the two, although an R^2 of 0.7846 suggests there is a strong relationship (Figure 81). We also examined the linalool data in relation to the parents of the original cross (Mv and R) and the Mv×R131 plant from which the F2 population was derived (Figure 82). As can be seen the abundances are quite variable and do not suggest a neat 1:1 segregation of the plants having the ability to produce moderate concentrations of linalool or not. This suggests that the phenotype is multigenic, controlled by several genes at a number of loci.

QTL analysis was conducted using the new map and linalool data generated from the low temperature SPME extraction of the samples. As predicted, multiple QTLs were detected above the LOD threshold (p <0.05); seven in all were identified across six chromosomes (Figure 83).

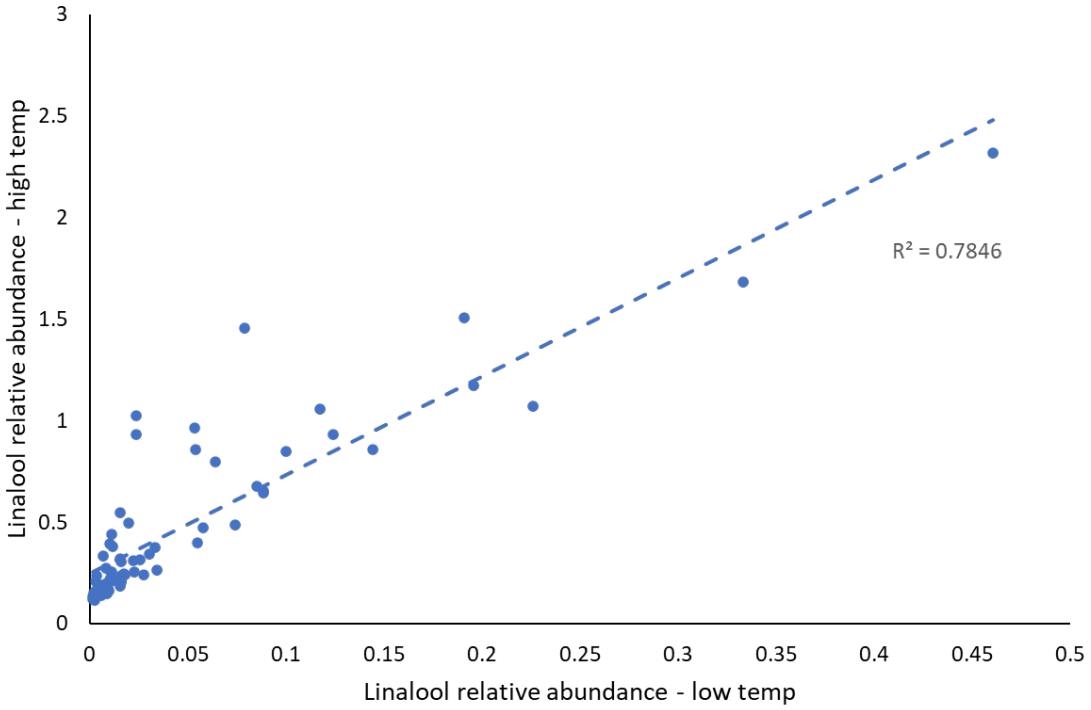


Figure 81 Relative abundances of linalool in berries extracted using high or low temperature SPME plotted against each other.

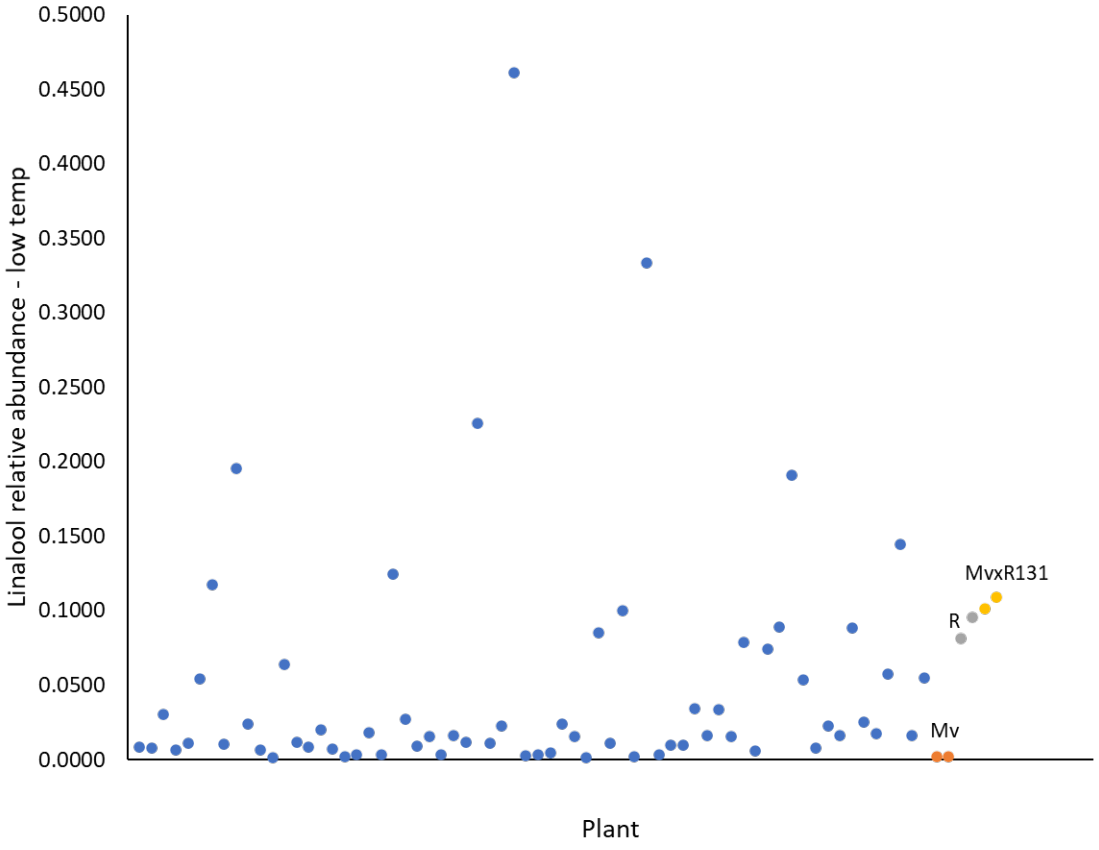


Figure 82 Relative abundances of linalool in the individual F2 progeny derived from the MvxR131 F1 plant, with concentrations in Mv, R and MvxR131 presented for comparison.

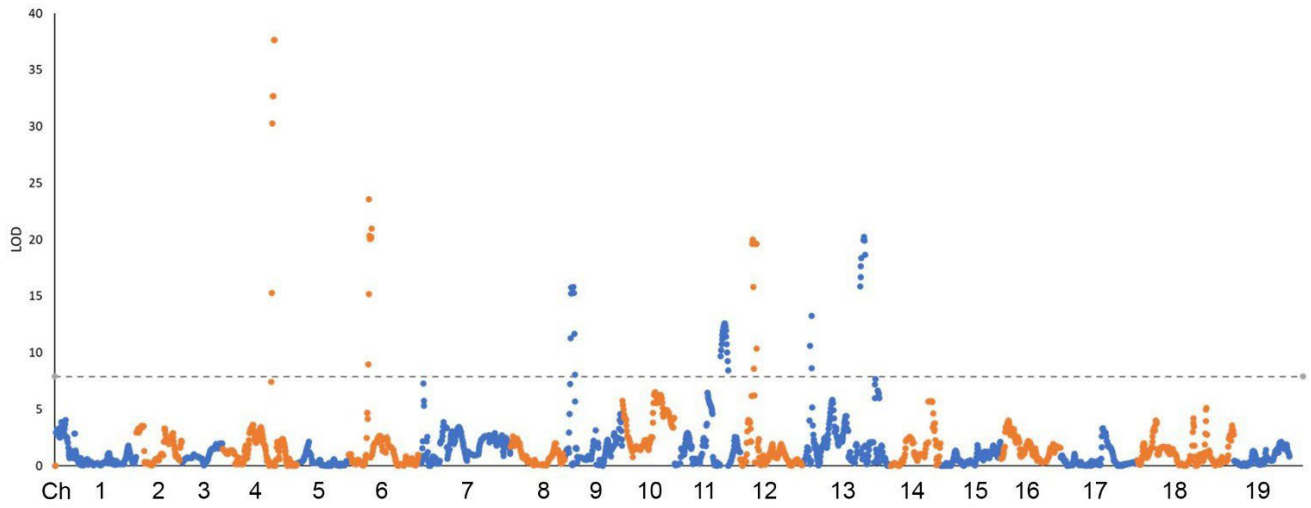


Figure 83 QTL mapping identifying genome regions significantly associated with linalool concentrations in berries from the MvxR131 F2 population. The dotted grey line represents the LOD threshold (p<0.05).

The QTLs together explained about 74% of the total variation in linalool concentration and the peaks had significantly higher LODs than the p<0.05 threshold of 7.9 (Table 17). The QTL explaining the most variance was located near the bottom of chromosome 4 and encompass a 1 Mbp region with ~100 genes. There are many candidates in this region that have roles in biochemical transformations, glycosylation, transport and chloroplast mRNA stability. But given the locus is only one of many controlling this phenotype, it is probably not a good target to want to trace in a breeding program or alter via gene editing.

Table 17 Location of QTLs from the MvxR131 F2 population associated with linalool concentration in berries and maximum LOD and % explained for each peak.

CHROMSOME	PEAK POSITION	MAX LOD	% EXPLAINED
Chr 4	103 cM	37.7	34 %
Chr 6	43 cM	23.6	10 %
Chr 9	18 cM	15.8	6 %
Chr 11	102.4 cM	12.6	2 %
Chr 12	25 cM	20.0	10 %
Chr 13	17 cM	13.2	5 %
Chr 13	117 cM	18.4	9 %

This population can be used to map other traits

The methods used to quantify linalool in the MvxR131 F2 progeny also extract other volatile compounds and these data can be used to map QTLs associated with the production of these compounds as well. While this was not the primary aim of the generation of this population, it does give us an interesting set of plants that may segregate for important odour-active compounds.

To demonstrate the potential utility of the population and the relatively dense genetic map produced, we tried to identify QTLs related to the concentration of a monoterpene called α-

phellandrene. This compound has an odour described as ‘dill-like’ (Blank, et al. 1992), and is often listed as having an aroma that is ‘peppery-minty’ and ‘slightly citrusy’. α-Phellandrene has many medicinal uses including improving wound healing, as an anti-microbial, anti-cancer and as an analgesic (Thangaleela, et al. 2022). It has also been listed as a volatile constituent of wine (Scacco, et al. 2010, Slaghenaufi, et al. 2022).

The Kruskal-Wallace non-parametric analysis of the marker:phenotype relationship in the Mv×R131 F2 populations provided strong evidence that the production of α-phellandrene in these plants is controlled by a gene on chromosome 13. Nineteen markers covering the area from 90 cM to 136 cM from the top of the chromosome were significantly associated with the phenotype at the $p < 0.001$ level. This area corresponded to a physical position between 7.5 Mbp to 20.4 Mbp from the top of chromosome 13. We chose a marker near the top and the bottom of this QTL to examine the marker:phenotype association, and it is clear that this does exist (Figure 84). The ‘b’ genotype represents the Mv parent and the ‘a’ genotype arises from the R parent with ‘h’ being the heterozygous genotype. The plants with the ‘b’ genotype at both markers often had no detectable α-phellandrene in the berries with the ‘a’ and ‘h’ genotype associated with α-phellandrene production. The marker at ~18.5 Mbp from the top of chromosome 13 suggests that homozygous ‘a’ plants produce more of the monoterpene compared to the heterozygotes (‘h’) at this locus (Figure 84) suggesting the trait is semi-dominant.

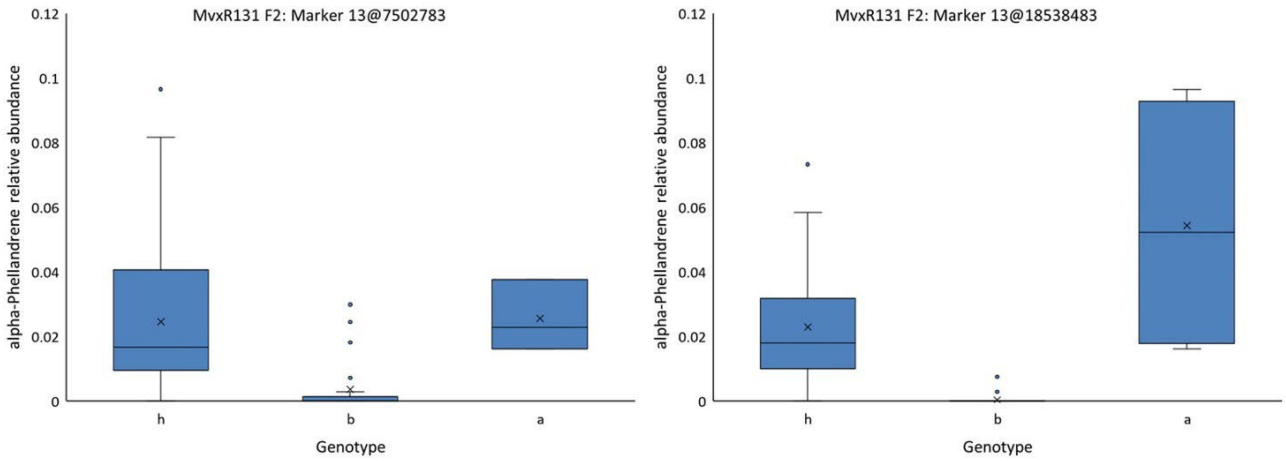


Figure 84 Box plots for two highly significant markers in chromosome 13 showing the relationship between phenotype and genotype in the MvxR131 F2 population.

When the MQM QTL analysis was conducted, a single QTL was detected on chromosome 13 (Figure 85). The region of the genome above the LOD threshold was narrowed down to between the marker at ~18.4 Mbp from the top of the chromosome and ~19.1 Mbp. The second version of the annotation of the PN42002 genome sequence has 32 genes in this interval, one of which is a monoterpene synthase and is a very strong candidate for the causal gene for this trait. More recent

annotations of the Pinot noir genome, however, indicate that there is probably a cluster of terpene synthase genes in this region of the genome, with at least 10 members of this gene family present.

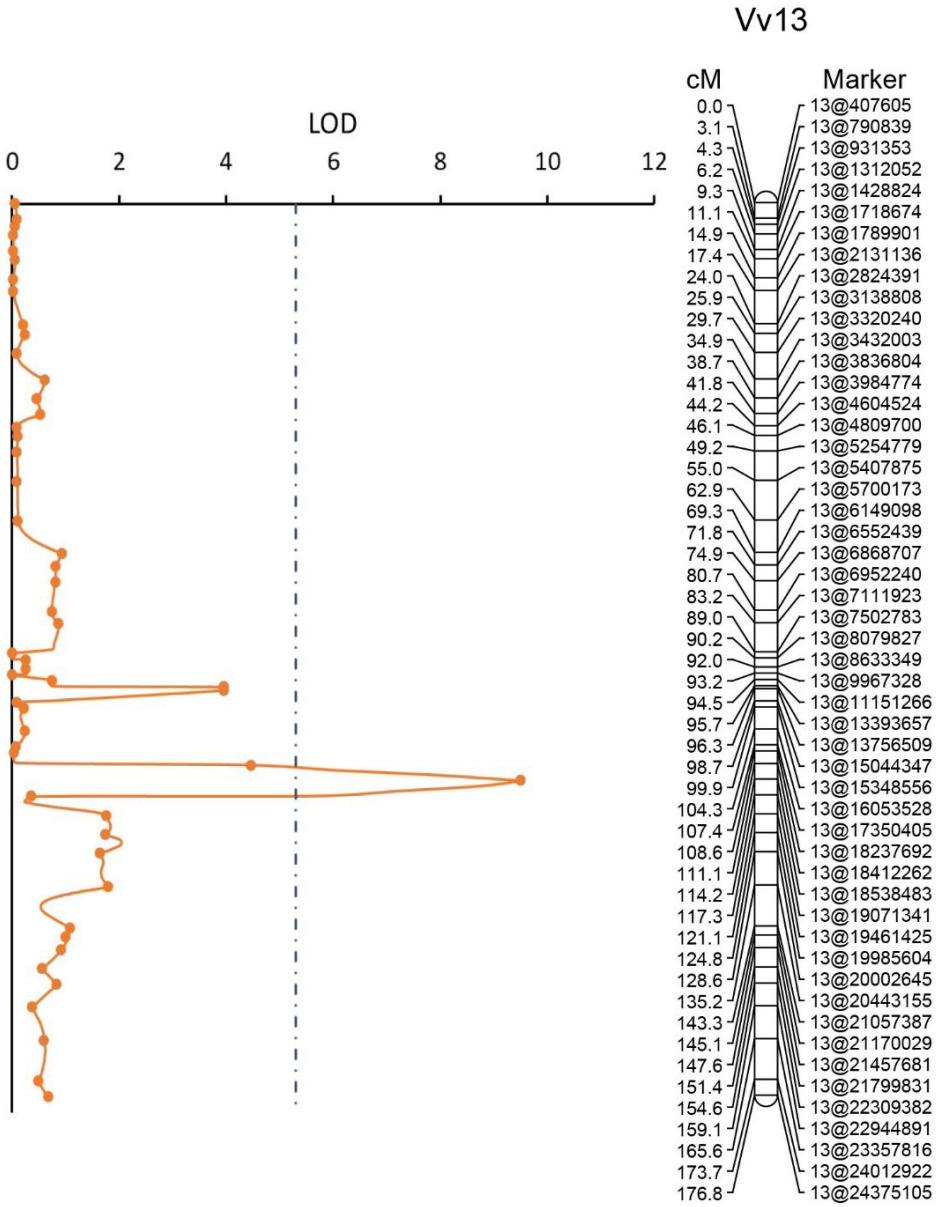


Figure 85 Schematic of the chromosome 13, α-phellandrene QTL mapping interval in the MvxR131 F2 population. Shows logarithm of odds (LOD) score against marker and map units (in cM) on the right. The vertical dashed line represents the genome-wide significance threshold ($p < 0.05$).

In summary, this population is a valuable asset for genetic studies, especially with regard to berry composition. Future projects looking at understanding or manipulating berry composition will benefit from the use of these progeny to map traits, identify causal genes and then use this information for following traits in breeding programs, gene-editing traits of interest or understanding the basis of berry compositional changes in the vineyard.

6.4.2 1st generation mildew-resistant vines

Introduction

When looking for novel germplasm to phenotype and genotype for GWAS studies, we decided to investigate the utility of the remaining plants produced as part of the production of the 1st generation mildew resistant lines from Project CSP 0904. From the original 1200 plants produced, 131 were still in the ground at Nuriootpa. They all have a plant known as 'R23' as the female parent, which harbours the resistance genes for powdery and downy mildew, and various male pollen donors were used as listed in Table 2, namely Chardonnay (5 plants), Dunkelfelder (34), Frontignac (25), Muscadelle (2), Muscat Gordo (17), Riesling (23), Sangiovese (1), Shiraz (2), Tannat (14), Tempranillo (1) and Verdelho (5).

The genotyping and phenotyping of these plants was conducted in collaboration with Project CSA 1701 - 2.2, where the focus of this project was wine volatile compounds and Project 2.2 looked at agronomic traits. Micro-scale wines were produced in both the 2020 and 2021 vintages. The volatile profiles were examined using SPME-GC-MS (as described previously) and 38 wine volatile compounds were targeted for quantification in the chromatographs. In parallel, leaf samples were taken for DNA extraction and the DNA was used to generate DArT markers, of which just under 17,000 passed the filtering process to be used for GWAS.

Volatile compound data analysis

The first step in deciding which volatile compounds to target for GWAS was to plot the results from the two vintages against each other to confirm that genotype rather than environment and/or yeast had a significant effect on the phenotype. The data were also plotted separately according to the male parent used for each line. This clearly distinguished the aromatic plants producing relatively high concentrations of monoterpenes (e.g. hotrienol) compared to those producing non-aromatic berries and the phenotype was dependent on the male parent (Figure 86). As can be seen, those plants derived from Muscat Gordo, Riesling, Frontignac and Verdelho show a spread of phenotypes in the plots, whereas those with Tannat and Dunkelfelder as the pollen donor all have very low concentrations of hotrienol (Figure 86). For other compounds there was a mixture of phenotypes regardless of which parent was used. For example, 1-octen-3-ol shows a good agreement between the wines produced from each vintage (Figure 87), which was surprising given that it is thought to be a compound that is associated with grape rot (La Guerche, et al. 2006). This relationship was better for plants produced using certain male parents – compare Muscat Gordo with Riesling (Figure 87). Following the production of these plots and the observation of the relationship between the data in both vintages, a subset of compounds was used for GWAS.

GWAS

The filtered DArT markers were used to test for marker:trait associations with some of the volatile compounds quantified in the wines. The vintages were analysed separately to test whether the same genome regions were identified as QTLs in both vintages, to give more confidence in the associations. In many cases, a significant association in one vintage was not seen in the other. This was surprising given the replication between the years was often very close, but may reflect the fact that there is structure within the genotypic data given the female parent is the same across all of the plants. Examples of how some of the targeted compounds mapped using GWAS are presented below.

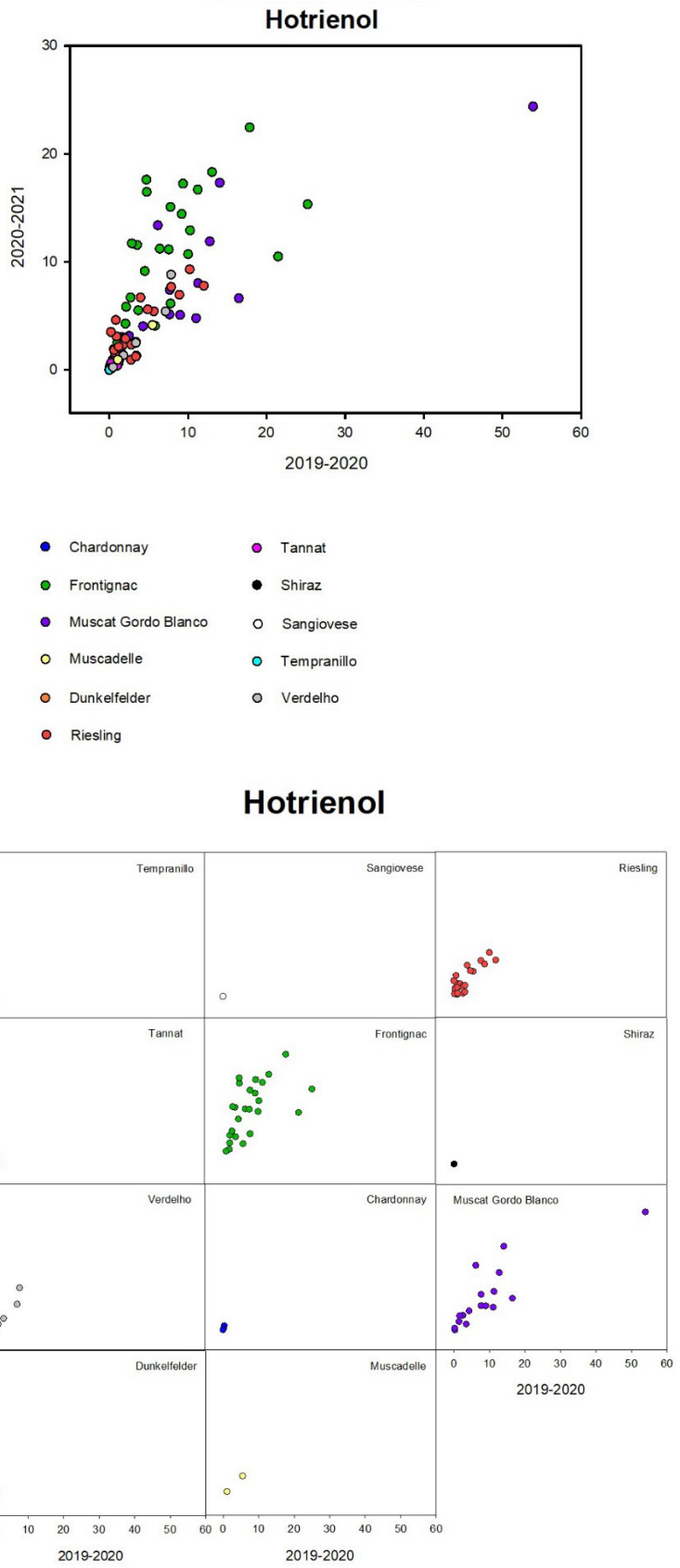


Figure 86 Plot of hotrienol abundances in replicate wines produced from lines produced for selection of the 1st generation mildew resistant varieties. The upper plot has plants colour coded by male parent and the lower plots present the data according to the male parent used to produce each plant.

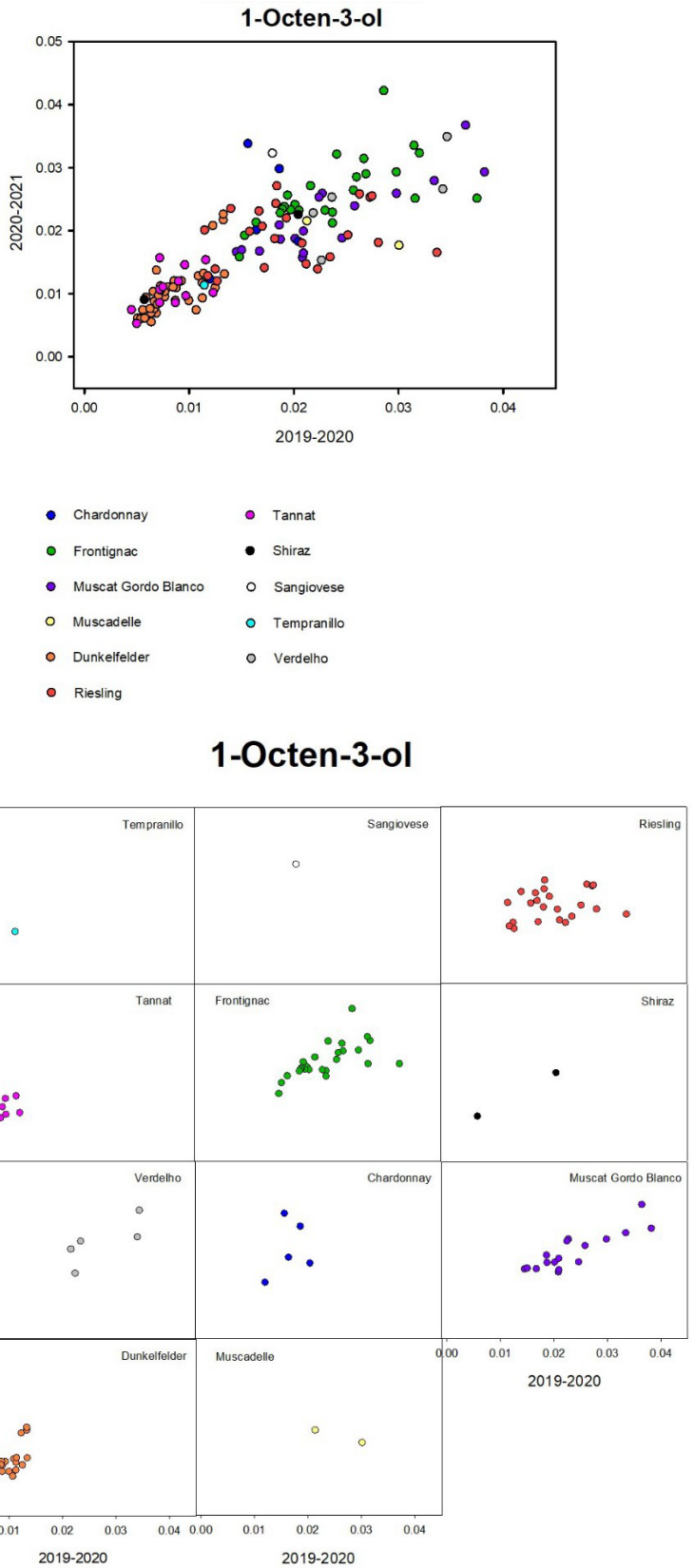


Figure 87 Plot of 1-octen-3-ol abundances in replicate wines produced from lines produced for selection of the 1st generation mildew resistant varieties. The upper plot has plants colour coded by male parent and the lower plots present the data according to the male parent used to produce each plant.

The first example is hotrienol, for which the compound abundance data are already presented in Figure 86 above. This compound has a 'citrus' and 'floral' aroma. In the 2020 vintage there was one significant QTL located on chromosome 5 (Figure 88) but no significant QTLs were detected using the 2021 vintage data. The significant marker was located at approximately 4 Mbp from the top of the chromosome and is in close proximity to the *VvDXS* gene associated with the high production of monoterpenes in berries and 'Muscat' flavour formation (Battilana, et al. 2011). Therefore, the mapping appears to be working for the 2020 vintage, but the lack of a significant association in 2021 was surprising given the relatively close relationship between the measurements in both vintages (Figure 86).

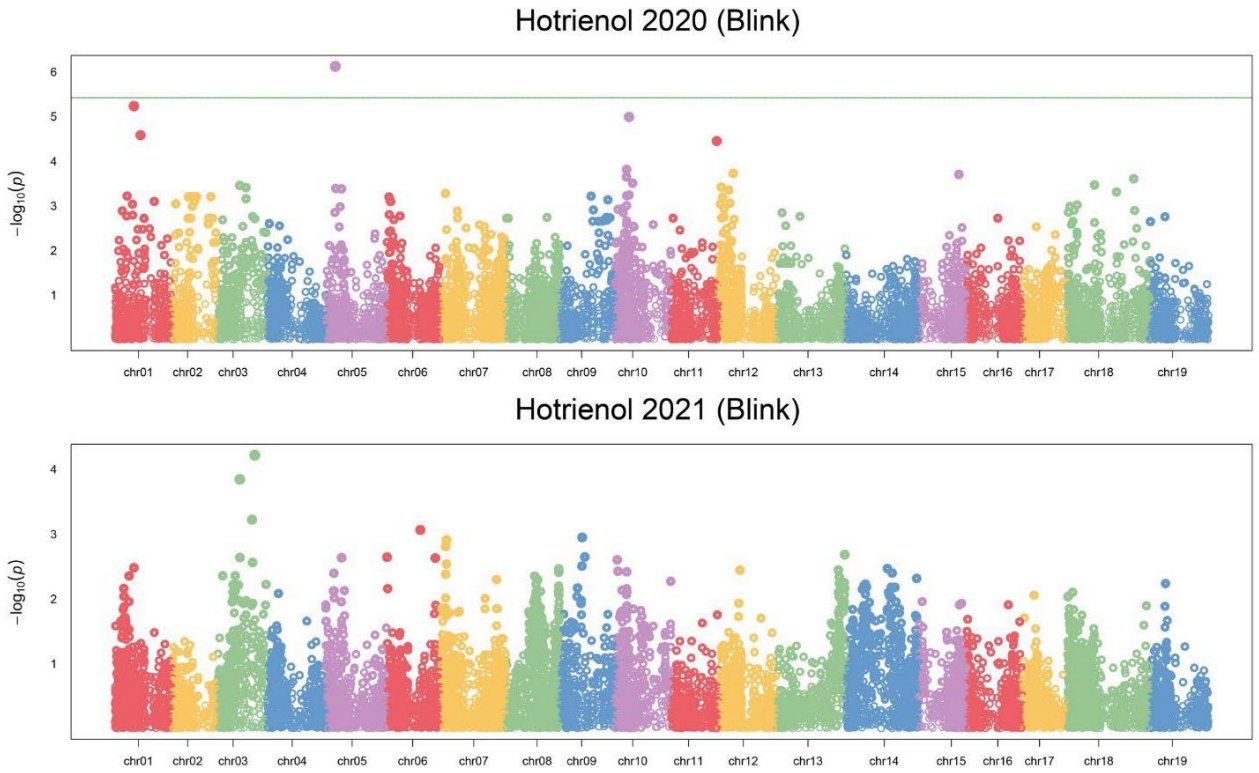


Figure 88 GWAS of hotrienol in the 1st generation mildew resistant varieties using the BLINK model in wines made from the 2020 and 2021 vintages. The horizontal line represents the significance threshold (p<0.01) and markers are grouped by chromosomes as indicated on the x-axis.

As mentioned above, the compound 1-octen-3-ol has been associated with grape rots and is seen as having a negative impact on wine aroma as it has 'mushroom' aroma and flavour characteristics (Xu, et al. 2015). Following on from our surprise that wine concentrations in both vintages were similar in the population, we also found that the trait mapped to a single QTL in both vintages (Figure 87). The QTL was located on chromosome 11 and was surrounded by potential genes of interest including one involved in wax composition and several related to leucine rich repeat receptors associated with disease resistance. The latter are interesting as the higher concentrations of 1-octen-3-ol in some lines may be related to fungal infection, the severity of which could be dependent upon disease resistance loci.

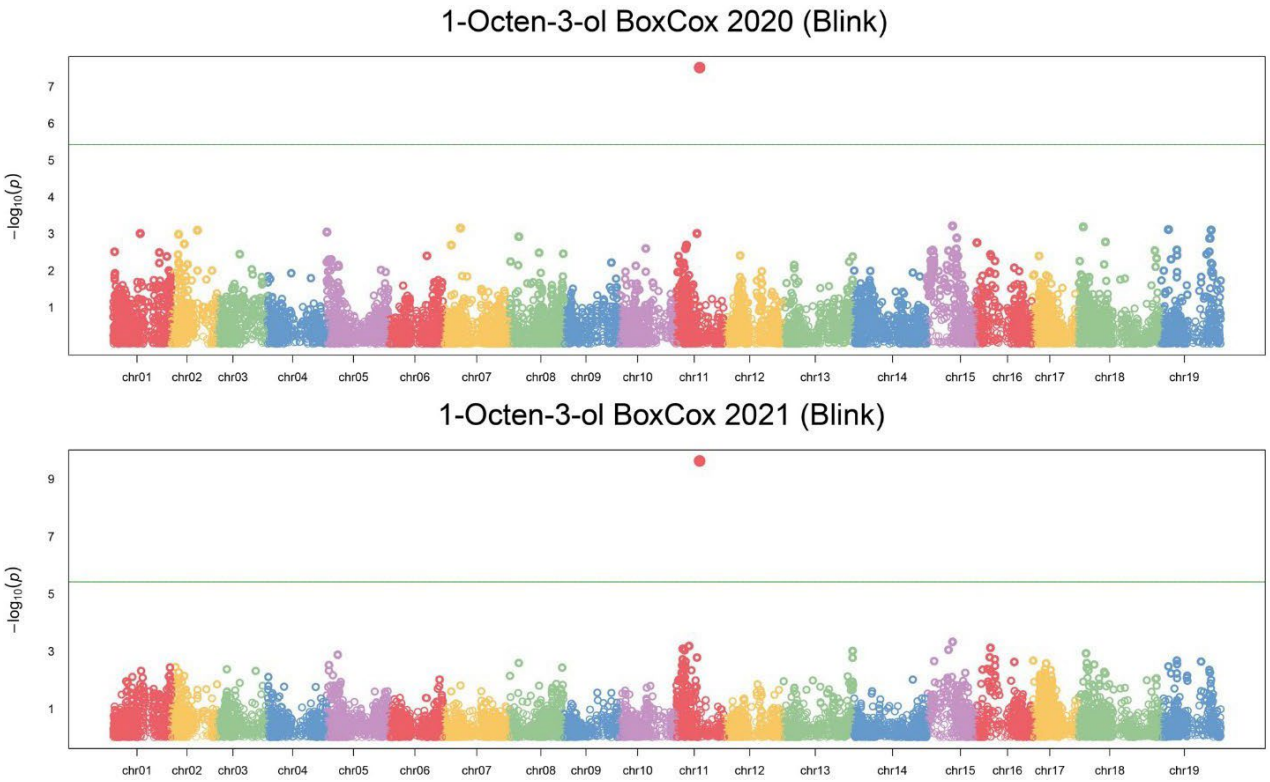


Figure 89 GWAS of 1-octen-3-ol in the 1st generation mildew resistant varieties using the BLINK model in wines made from the 2020 and 2021 vintages. The horizontal line represents the significance threshold (p<0.01) and markers are grouped by chromosomes as indicated on the x-axis.

In summary, the phenotypic analysis of the micro-scale wines from progeny of the crosses made for the production of the 1st generation disease resistant vines was promising with several wine volatile compounds showing good replication across two vintages. This suggests that the grape compositional contribution to the accumulation of these wine volatile compounds is robust enough to be maintained in different climatic conditions. We were surprised to subsequently find that the GWAS did not show the same consistency across vintages for most compounds. This anomaly should be investigated further to ensure that we make the most of these phenotypic and genotypic datasets. It is possible that the QTL mapping could be done by treating the progeny as originating from a biparental cross, given they all share a female parent, and this could solve the issues seen with GWAS.

6.5 Development of markers to genes of known function

6.5.1 Introduction

An ongoing activity in the project was to keep abreast of the latest literature concerning grapevine consumer-preferred traits and especially any molecular biology studies identifying genes that would be considered important to follow in any breeding program. It is interesting that despite the importance of this crop worldwide, there are very few examples where such genes have been identified, perhaps due to the difficulty in conducting genetic studies with grapevine. The breeding program was already using a marker that predicted the ability of the grapes to produce a 'Muscat'

flavour and this was based around the point mutation identified the *VvDXS* gene (Battilana, et al. 2011). The first two markers developed were based on previous projects conducted by PhD students at CSIRO in the grapevine research group in Adelaide. Another couple of markers were developed based on external projects, but there has been very little movement in this field of research across the five years (other than the project described this report) and so this was a less fruitful part of the project than was envisaged.

6.5.2 Marker for acylated anthocyanins

Anthocyanins are the molecules that produce the colour of red wines. They are present in several different forms in the grape skins, some of which have been shown to have more stability in wines than others and the process of acylation has been reported as increasing stability of the anthocyanin molecules (Van Buren, et al. 1968, Yonekura-Sakakibara, et al. 2008). In a previous Wine Australia funded PhD project, GWR Ph0903, Amy Rinaldo discovered the gene responsible for the acylation of anthocyanins in grape berries and, importantly, identified the mutation responsible for the lack of these acylated anthocyanins in Pinot varieties (Rinaldo, et al. 2015). This is very important for the breeding program as the Mv is derived from Pinot Meunier and therefore lacks the ability to produce acylated anthocyanins, so following this trait when stacking other genes in an Mv background is essential.

The lack of anthocyanin acylation in Pinot varieties is caused by a 'nonsense' mutation in both alleles of the *Vv3AT* gene, which encodes an anthocyanin, 3-*O*-glucoside-6''-*O*-acyltransferase, which causes it to produce an inactive truncated protein. The mutation does not disrupt a restriction enzyme site and so a dCAPS marker was designed instead where a *Pvu*II restriction enzyme site was introduced in the 3' primer and this digested the amplified wild type allele (allele 1), causing a slightly smaller band (Table 18) that could be distinguished from the mutant allele (allele 2) on a high percentage agarose gel.

Table 18 Details of dCAPS markers for predicting anthocyanin acylation in new varieties.

MARKER	5'PRIMER	3'PRIMER	ENZYME	ALLELE 1	ALLELE 2
Vv3AT	CCGTTGTTCATATATCCATCAGGG	ATTGGTGAATTCTCCGCCTGCTAACTGAATCAGCT	*Pvu*II	'C' (459bp)	'T' (494bp)

We tested the marker against a selection of varieties from the Coombe vineyard for which we had DNA. In all, 91 wine and table grape varieties were genotyped and all but nine were homozygous for the wild type allele (allele 1). As expected, the Pinot noir genotype was homozygous for the mutant allele, but eight varieties were heterozygous and carried a copy of the mutant allele (namely: Chardonnay, Traminer, Gamay, Sangiovese, Clairette, Valdiguie, Souzao and Verdicchio). This is valuable information for future breeding strategies as it tells us that we need to follow this trait in any crosses conducted with these varieties and also suggests that they all have a Pinot variety in their genetic pedigree, which is already suspected for Chardonnay, Traminer and Gamay for example.

6.5.3 Marker for methoxypyrazine production in berries

Methoxypyrazines are responsible for the 'capsicum', 'herbaceous' and 'earthy' characters in wines predominantly from varieties grown in the Bordeaux appellation (Dunlevy, et al. 2013, Koch, et al. 2010). When present in higher concentrations, these compounds can be detrimental as the 'herbaceous' characters can mask the more desirable floral and fruity sensory properties in the wine (Hein, et al. 2009). Jake Dunlevy, in a previous Wine Australia-funded PhD project, GWR Ph06/06, identified a methyltransferase gene responsible for the conversion of hydroxypyrazine precursors to methoxypyrazines (Dunlevy, et al. 2013) and showed that this gene determined whether a variety would produce these compounds in the berry or not.

We decided to design a marker that is in the gene of interest (*VvOMT3*) but is unlikely to be the genetic cause of the phenotype as this is still unknown. The marker will indicate the potential of a plant to produce methoxypyrazines in the berries and so would be best used as a way of screening out material that has this ability. As there are other modifiers that determine the total amount of methoxypyrazines in the fruit at harvest (e.g. growing region or canopy structure), some plants with the active version of this gene may still only produce low concentrations when grown under certain conditions. But the potential is still there and so it would be best to select against this possibility.

The marker is designed to a SNP that distinguishes the *VvOMT3* alleles present in Cabernet Sauvignon from those in Pinot noir. The mutation did not disrupt or create a restriction enzyme site in the sequence and so a dCAPs marker was designed. The forward primer was designed to introduce an *Nco*I site when the Cabernet Sauvignon allele is amplified, producing a slightly smaller fragment from the Pinot noir allele when digested with this enzyme (Table 19).

Table 19 Details of dCAPS marker for predicting potential methoxypyrazine production in the berries of new varieties.

MARKER	5'PRIMER	3'PRIMER	ENZYME	ALLELE 1	ALLELE 2
VvOMT3	AAGTGTGCCATTGAGCTCGGCATAGCTGATGCCAT	GCGTTCCCTTTCCCTAGAAG	*Nco*I	'C' (284bp)	'T' (316bp)

The marker was tested across a range of varieties and those carrying the active allele included Cabernet Sauvignon, Cabernet Franc and Merlot (all homozygous for the active allele), Semillon, Sauvignon blanc, Ruby Cabernet, Rousanne, Aleatico, Tinta Cão, Biancone, Durif, Crouchon, Terret noir, Raboso Piave, Souzão and Nebbiolo (all heterozygous). It is important to know those varieties that carry the active allele so this trait can be monitored using this marker in future breeding programs or in selecting superior breeding lines.

6.5.4 Marker for potential rotundone biosynthesis

Rotundone is the sesquiterpene responsible for the 'pepper' character in Shiraz and some other grape varieties (Mattivi, et al. 2011, Wood, et al. 2008). A paper reporting the identification of a terpene synthase gene responsible for the biosynthesis of the precursor of rotundone (α-guaiene) was published in 2016 (Drew, et al. 2016). More specifically, two specific polymorphisms were

required in the Shiraz gene sequence to reduce the percentage of α-guaiene produced from 45% of the products to 5%, with one mutation (V530M) reducing the percentage to 24% and the other (T414S) having a lesser effect. We decided to design markers to detect both of these SNPs, as the presence of either T414 or V530 (or both) would suggest that a variety has the potential to produce rotundone and these alleles have been denoted as 'R'.

The T414S polymorphism represents the effect of changing the 'C' nucleotide in the middle of the 'ACC' codon, which makes residue 414 in the protein sequence a threonine, into a 'G' which then encodes a serine ('AGC' codon). The presence of the 'G' nucleotide introduces an *Alu*I restriction enzyme site into the sequence and so primers were designed either side of this SNP to amplify the region by PCR and digest with this enzyme. Those with the larger band encode an enzyme with the ability to produce a higher percentage of α-guaiene than plants with the smaller band (Table 20). The V530M polymorphism changes the 530 residue from a valine to a methionine and this is caused by a 'G' to 'A' mutation that changes the codon from 'GTG' to 'ATG'. The presence of the 'A' nucleotide introduces a *Bcc*I restriction enzyme site into the DNA and again primers were designed around this region to amplify the region, which is then digested with this enzyme to detect the SNP. The lack of the restriction enzyme site results in a larger band and denotes the form of the enzyme which produces a higher percentage of α-guaiene than those where the DNA is digested to form a smaller band (Table 20).

Table 20 Details of dCAPS markers for predicting potential rotundone production in the berries of new varieties. Those with the 'R' allele contain the form of the enzyme that produces a higher percentage of α-guaiene than the 'r' alleles.

MARKER	5'PRIMER	3'PRIMER	ENZYME	ALLELE 1 (R)	ALLELE 2 (r)
VvTPS24_414	TGTAATTTGAAAAGGTTCCAAGGATATG	GAAAATCAAAAGAATTTCAGTATACACAAAAC	*Alu*I	'C' (493 + 114bp)	'G' (389 + 114 +104bp)
VvTPS24_530	CGTTTTGTGTATACTGAAATTCTTTTGAT	TAGAACCATCTCCTGGCCAT	*Bss*I	'G' (358bp)	'A' (241 + 117bp)

The marker was tested against a selection of varieties for the presence of one or both SNPs. This revealed one drawback of the markers in that we are not able to determine if an allele has both mutations, unless at least one is homozygous for one SNP and heterozygous for the other. We were surprised at how widespread the 'R' alleles were in the germplasm collection (Table 21) with several varieties being homozygous 'R' for the T414S SNP, but none for the V530M SNP, which had a greater effect on the activity of the enzyme (Drew, et al. 2016). It is suspected that these SNPs in *VvTPS24* are not the sole determinant of whether the plant will be able to produce rotundone in the berry. A gene has been identified that encodes an enzyme capable of oxidising α-guaiene to rotundone called *VvSTO2* (Takase, et al. 2015). However, there have been no analyses linking gene sequence differences to function in grape varieties and so at present there is not a pathway to producing a marker to this gene at present.

Table 21 Genotypic data from a selection of grapevine varieties testing the two markers designed against the *VvTPS24* gene. Those with the 'R' allele contain the form of the enzyme that produces a higher percentage of α-guaiene than the 'r' alleles.

VARIETY	MARKER		VARIETY	MARKER	
	VVTPS24_414	VVTPS24_530		VVTPS24_414	VVTPS24_530
Aleatico	RR	rr	Arneis	Rr	Rr
Baco noir	rr	rr	Barbera	rr	rr
Barlinka	rr	rr	Bastado	rr	rr
Beauty Seedless	rr	rr	Black Frontignac	Rr	Rr
Bonvedro	Rr	Rr	Cabernet Franc	rr	rr
Calmeria	rr	rr	Canocazo	Rr	Rr
Cardinal	rr	rr	Carignan	rr	rr
Carina	Rr	rr	Chardonnay	rr	Rr
Chasselas	rr	rr	Chenin blanc	rr	rr
Cinsaut	Rr	Rr	Colombard	rr	rr
Concord	rr	rr	Crouchen	RR	rr
Dawn Seedless	rr	rr	Dolcetto	Rr	Rr
Doradillo	rr	rr	Durif	Rr	Rr
Fer	rr	rr	Flame Seedless	rr	rr
Gamay	rr	rr	Gold	Rr	Rr
Grenache	rr	rr	Kyoho	rr	rr
Malbec	Rr	Rr	Malvasia Bianca	Rr	rr
Marroo Seedless	rr	rr	Marsanne	rr	rr
Mataro	Rr	rr	Merbein Seedless	rr	rr
Merlot	rr	rr	Mondeuse	Rr	Rr
Monukka	rr	rr	Muller Thurgau	Rr	rr
Muscadelle	rr	rr	Muscat Gordo	RR	rr
Muscat Hamburg	Rr	rr	Nebbiolo	RR	rr
Ondenc	Rr	Rr	Palomino	Rr	Rr
Parthenocissus	rr	rr	Pedro Ximenes	Rr	rr
Perlette	rr	rr	Petit Verdot	rr	rr
Pinot noir	Rr	Rr	Red Globe	rr	rr
Riesling	Rr	Rr	Rkaziteli	rr	rr
Rousanne	rr	rr	Rubired	rr	Rr
Sangiovese	Rr	rr	Sauvignon blanc	RR	Rr
Semillon	rr	rr	Shiraz	Rr	Rr
Souzao	rr	rr	Sultana	rr	rr
Tannat	Rr	Rr	Tempranillo	rr	rr
Terret noir	RR	Rr	Tinta Cao	RR	rr

VARIETY	MARKER		VARIETY	MARKER	
	VVTPS24_414	VVTPS24_530		VVTPS24_414	VVTPS24_530
Touriga	RR	Rr	Traminer	Rr	rr
Trebbiano	rr	rr	*Vitis candicans*	rr	rr
Vitis riparia	rr	rr	Valdiguie	Rr	rr
Verdelho	Rr	Rr	Verdicchio	RR	Rr
Viognier	rr	rr	Waltam Cross	rr	rr
Zante Currant	rr	rr	Zindfandel	Rr	Rr

6.5.5 Marker for red berry flesh

A marker for red berry flesh was created in collaboration with Project 1.1. It has recently been shown that the red flesh phenotype is caused by a 408bp region of the promoter of the *VvMybA1* gene (which is responsible for berry skin colour (Walker, et al. 2007)) being tandemly repeated (Röckel, et al. 2020), with more repeats having a greater effect on anthocyanin accumulation. In the Dunkelfelder variety that has been the source of the red flesh phenotype in the CSIRO breeding program, the promoter has two repeats.

The marker designed for the red flesh phenotype is, in essence, a Sequence Characterised Amplified Region (SCAR) marker which amplifies a band using PCR when there are two repeats in the promoter but does not when there is only a single copy of this 408bp promoter region. The main issue with SCAR markers is that a null result can mean that, in this case, there is only a single copy of the 408bp region, or that the amplification has failed. To eliminate the second possibility, another gene is also amplified in the same PCR mix to ensure the DNA is not degraded and that the reaction has worked. The grapevine ubiquitin gene was used for this purpose. This marker was tested and validated as part of Project 1.1 and more information can be obtained in that final report.

MARKER	5'PRIMER	3'PRIMER	ALLELE 1 (RED FLESH)	ALLELE 2 (WHITE FLESH)
VvMybA1_RF	CGTAATGTCCCATTCAGTGAG	CGGCATCTATTCAACCCTGC	976bp	no band
VviUbi (control)	AGTAGATGACTGGATTGGAGGT	GAGTATCAAAACAAAAGCATCG	174bp	174bp

6.6 GC-MS analysis of 1st generation mildew-resistant red wines

6.6.1 Introduction

In a parallel SRA project (CSA 1701-1.6), a selection of 1st generation mildew resistant varieties was being evaluated for their vine performance and fruit and wine quality/style. We wanted to investigate the possibility of using SPME-GC-MS to classify these wines based on their volatile compound profiles and to determine whether it would be feasible as a screening method for the new 2nd generation mildew resistant varieties being developed in CSA 1701-1.1.

This pilot study was conducted with a series of wines from the 20 selected red lines grown at the Charles Sturt University vineyard at Wagga Wagga. The parentage of these selections is relatively narrow with R23 (see section 5.2.2 for pedigree) as the female parent and the pollen donors being Dunkelfelder (R1-15), Shiraz (R16) and Tannat (R17-20). Descriptions of winemaking and harvest parameters can be found in the final report for CSA 1701-1.6. As controls, Shiraz and Cabernet Sauvignon wines from grapes grown in the same vineyard were analysed in parallel. The wines were analysed using a non-targeted SPME-GC-MS protocol to quantify a broad range of compounds (section 5.3.5) and the final profiles contained 177 compounds (Table 3).

6.6.2 Comparison of volatile compound profiles

Principal component analysis (PCA) was used to compare the wines with regard to their overall volatile compound profiles. This methodology is good for visualising the relationship between the different samples and to identify those that appear to be unique with regard to the variables used to create the score plot. The PCA plot in Figure 90 shows the relative position of the wine samples based on the first two principal components (PCs), which describe 20% and 14% of the total variation respectively. By this analysis we can see that the volatile compound profile of R14 is similar to that of Shiraz and R13 is close to that of Cabernet Sauvignon. Some other mildew resistant lines (R15 and R17) were located in a region of the PCA relatively close to that of Cabernet Sauvignon, and R10, R11 and R7 were close to Shiraz on the plot. The PCA suggests that R16 and R5 are outliers and have more unique profiles compared to the other wines. R5 would also be the most dissimilar to Shiraz and Cabernet Sauvignon being located both on the opposite side of PC1 to Shiraz and PC2 to Cabernet Sauvignon.

A concern of the PCA is that the first two PCs only describe 34% of the total variation, so we are only visualising the relationship between the samples based on one-third of the variation. The third PC describes another 13%, but visualising the three-dimensional plot is difficult on paper. So, to compliment the PCA, a hierarchal clustering of the samples was also conducted to examine the relatedness of the wine volatile composition (Figure 91). A logical interpretation of this analysis is that there appear to be six groups of wine based on their volatile compound profiles, and the six groups appear to be paired into three larger groups (Figure 91). For example, Cabernet Sauvignon and R13 make one group, but these are more closely related to the group that includes R8, R15, R17 and R20 than the other wines. In the next larger group, there is one set of similar wines made up of R4, R5 and R6, and a group of five that includes R3, R9, R12, R18 and R19. Interestingly, the cluster

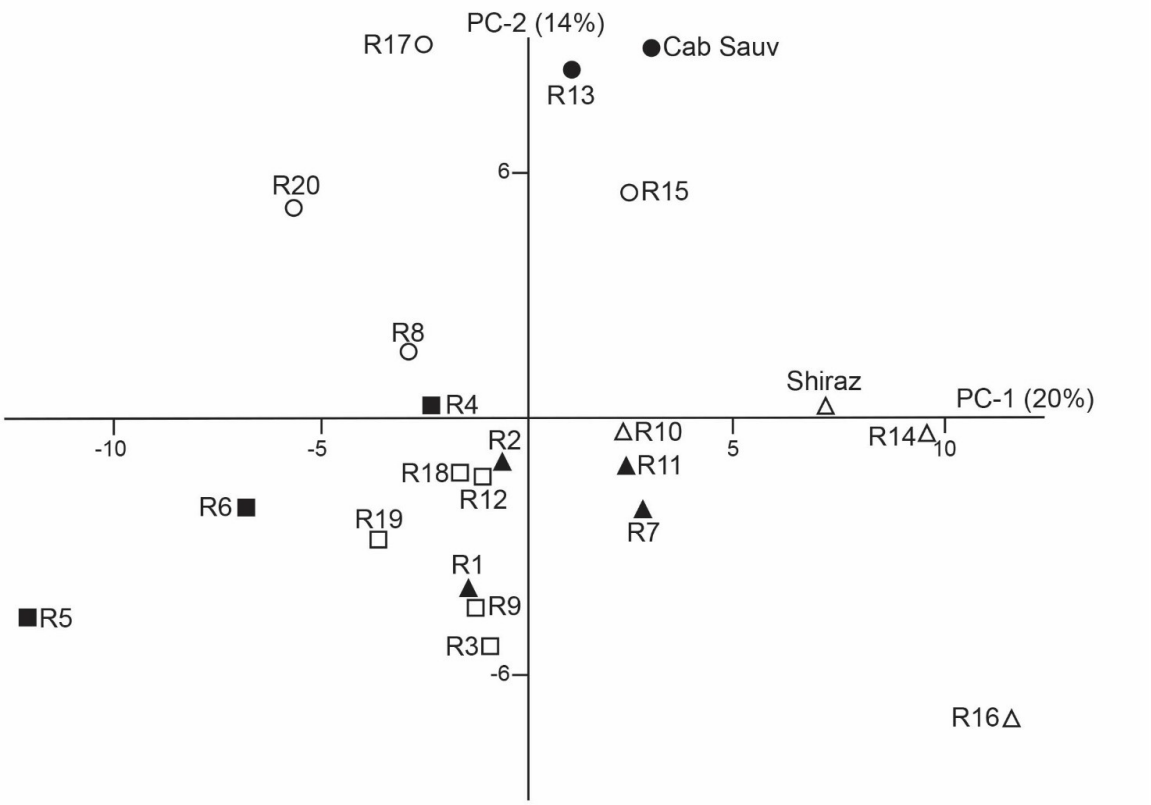


Figure 90 Score plot from PCA of 20 red wines made from 1st generation mildew resistant selections alongside Cabernet Sauvignon and Shiraz grown in the same vineyard. The same shapes are used for those wines grouped together by cluster analysis (see Figure 91).

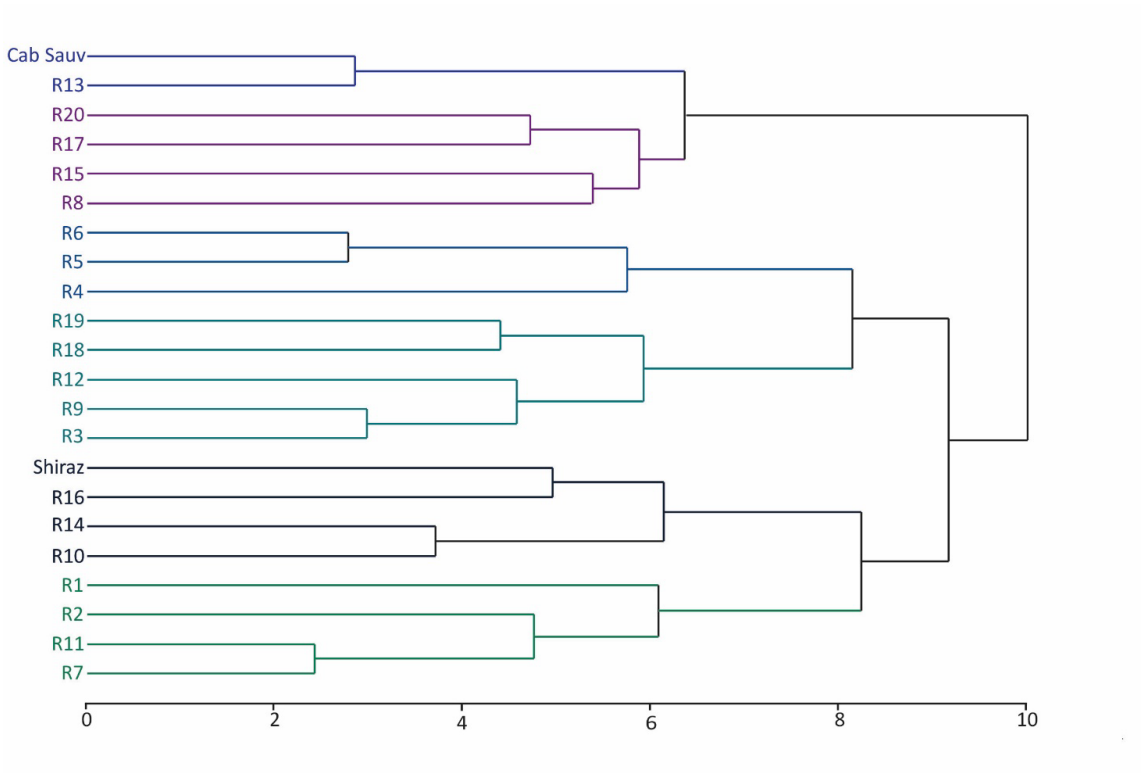


Figure 91 Complete linkage clustering using uncentered correlation of the volatile profiles from each wine. The scale represents relative distance.

analysis suggests that R16 (with a Shiraz male parent) is the closest to Shiraz, with R10 and R14 also in that group. Finally, the R1, R2, R7 and R11 wine volatile compound profiles are clustered together and are relatively similar to the group containing Shiraz (Figure 91).

6.6.3 Volatile compounds associated with the groups from the cluster analysis

We decided to dive deeper into the data to identify the volatile compounds associated with each of the six groups of wines as assigned by the cluster analysis (Figure 91). This analysis may give some insight into the wine style that each group of wines would produce. To determine those compounds that are present in relatively higher or lower concentrations in each of the six groups, partial least squares – discriminant analysis was conducted on the volatile compound profiles. Those that were significantly positively or negatively associated with each cluster are listed in Table 22.

The cluster that contained R1, R2, R7 and R11 was located near the origin of the PCA but was negatively projected on PC2 (Figure 90). Compounds that were positively associated with these wines included three medium chain fatty acids, three esters with octanoate as the acid group and 2-phenylethyl alcohol (Table 22). The fatty acids are known to be suppressors of 'fruity' wine characters and have a 'cheesy' odour (Espinase Nandorfy, et al. 2022). The esters would be contributing to fruity characters and the 2-phenylethyl alcohol would impart 'rose' sensory attributes. Compounds in relatively lower concentrations in these wines include a number of aliphatic alcohols that are associated with negative 'chemical' sensory characters. They also tended to be lower in some acetate esters ('fruity') and some monoterpenes ('citrus'/'floral'). Two compounds associated with 'green' or 'earthy' characters, (*Z*)-3-hexenol and 3-(methylthio)-1-propanol (and associated esters), were also relatively lower in these wines compared to other clusters (Table 22).

The second cluster analysed contained R3, R9, R12, R18 and R19 and these were located in a similar region of the PCA as the cluster described above but were more negatively projected on PC1 than the first one analysed (Figure 90). These wines had relatively higher concentrations of some aliphatic alcohols and 3-(methylthio)-1-propanol and its ethyl ester than those in other clusters (Table 22). Interestingly, these wines were more abundant in compounds that should impart 'minty' or 'pine-like' characters to the wine such as borneol, methyl salicylate and thymol. In contrast, these wines were negatively associated with some fatty acids, TDN ('kerosene') and a series of ethyl esters of unsaturated fatty acids and some succinate esters (Table 22).

R4, R5 and R6 made up the next cluster examined, and these were negatively projected on PC1 and neutral to negatively projected on PC2 (Figure 90). These wines were interesting in that they had higher concentrations of a number of norisoprenoids that arise from the cleavage of carotenoids – namely, 6-methyl-5-hepten-2-one ('citrus', 'green'), β-isophorone ('mint'), theaspirane A & B ('tea-like'), vitispirane ('floral'), β-cyclocitral ('saffron', 'herbal') and TDN (Table 22). They were also more abundant in some fusel oils, eucalyptol and the sesquiterpenes cabreuva oxide ('woody') and α-calacorene ('woody'). They also had higher concentrations of the methyl ketones 2-heptanone ('fruity'/'spicy') and 2-nonanone ('fruity') than the other wines. Amongst the negatively associated compounds were the aliphatic alcohols and some long chain ethyl esters. We would predict that these wines would have less 'fruit-driven' sensory characters but have some interesting 'herbal' notes.

The next cluster consisted of R8, R15, R17 and R20 and these were somewhat related to the Cabernet Sauvignon R13 cluster in that they were all found in the top half of the PCA being positively projected on PC2 (Figure 90). These wines were notable for having higher concentrations of (*Z*)-3-hexenol and related esters that would impart 'fresh green, fruity' characters to the wine (Table 22). They also had a greater abundance of other esters such as ethyl propanoate ('pineapple'), ethyl valerate ('fruity') and ethyl cinnamate ('cinnamon', 'cherry') compared to other clusters and so would potentially have interesting 'fresh fruit' sensory characters. They were negatively associated with a number of compounds that would tend to have detrimental effects on wine such as fatty acids and fusel alcohols (Table 22).

The cluster containing Shiraz also included three of the 1st generation mildew resistant lines – R10, R14 and R16 – and these wines were positively projected on PC1 and were near the origin or negatively projected on PC2 (Figure 90). Several monoterpenes ('citrus'/'floral' were positively associated with these wines as well as some esters that have propyl as the alcohol group ('fruity'). Three other notable compounds that were more abundant in these wines compared to those in other clusters were β-ionone, which is the main odorant in violet flowers, furfural, which is 'almond-like' and 1-octen-3-ol, which smells like mushrooms (Table 22). Only three compounds were negatively associated with the wines in this cluster, the 'minty' or 'pine-like' thymol, borneol and 2-nonanone (Table 22).

Finally, the last cluster consisting of Cabernet Sauvignon and R13 were located at the top of the PCA plot and were slightly positively projected on PC1 as well (Figure 90). Compounds that tended to be in high concentrations in these wines included some fusel alcohols as well as some other branched chain alcohols that have descriptors that include 'nutty' and 'fruity' (Table 22). Butanoic and isovaleric acids have very unpleasant aromas and are positively associated with these wines as are a number of 'fruity' esters and γ-nonalactone, which is described as smelling like 'coconut' or 'peach'. Some compounds with 'green' or 'citrus' aromas are negatively associated with these wines, such as monoterpenes and esters of unsaturated C_6 alcohols (Table 22).

Table 22 Results of PLS-DA on clusters of wines based on volatile compounds with indication of model strength and the compounds positively or negatively associated with the wines in each cluster

1ST GENERATION RED WINE CLUSTER	R^2_{CAL}	R^2_{VAL}	POSITIVELY ASSOCIATED COMPOUNDS	NEGATIVELY ASSOCIATED COMPOUNDS
R1, R2, R7 & R11	0.94	0.93	Isoamyl octanoate 2-Phenylethyl alcohol Octanoic acid Decanoic acid 2-Phenylethyl octanoate Dodecanoic acid Ethyl octanoate	Propyl acetate Isobutyl acetate Ethyl 2-methylbutyrate 1-Butanol 1-Pentanol 3-Hydroxy-2-butanone Ethyl (Z)-3-hexenoate 2,2,6-Trimethyl-cyclohexanone (E)-3-Hexenyl acetate Unknown monoterpene 2 3-Ethoxy-1-propanol (Z)-3-Hexen-1-ol

				2-Nonanone β-Isophorone 2-Methyltetrahydrothiophen-3-one Linalool 1-Octanol Ethyl 3-(Methylthio)propionate 1-Nonanol 3-(Methylthio)-1-propanol Propyl decanoate Ethyl phenylacetate Isoamyl acetate
R3, R9, R12, R18 & R19	0.96	0.95	Ethyl propanoate Isoamyl propanoate 3-Hydroxy-2-butanone 3-Ethoxy-1-propanol 2-Nonanol 2-Methyltetrahydrothiophen-3-one 1-Octanol Ethyl 3-(Methylthio)propionate 1-Nonanol α-Terpineol Borneol 3-(Methylthio)-1-propanol Methyl salicylate Ethyl 4-hydroxybutanoate Thymol	Ethyl crotonate 1-Pentanol Propyl hexanoate Isopentyl hexanoate Acetic acid Unknown monoterpene1 Ethyl 7-octenoate Propyl octanoate Ethyl methyl succinate Ethyl 9-decenoate Ethyl (E)-4-decenoate TDN Hexanoic acid Phenylethyl butyrate Decanoic acid 9-Decenoic acid Ethyl hydrogen succinate
R4, R5 & R6	0.94	0.92	Propyl acetate 1-Butanol 2-Heptanone Eucalyptol 2-Methyl-1-butanol 3-Methyl-1-butanol Styrene 2-Heptanol 6-Methyl-5-hepten-2-one 2-Nonanone β-Isophorone Theaspirane B Vitispirane Theaspirane A β-Cyclocitral (*Z*)-β-Farnesene TDN Cabreuva oxide α-Calacorene 2,6-Ditert-butyl-4-methylphenol	2,3-Butanedione 2,3-Pentanedione Methyl octanoate 1-Heptanol 2-Ethyl-1-hexanol 2-Methyltetrahydrothiophen-3-one 1-Octanol Ethyl decanoate 1-Nonanol 3-(Methylthio)-1-propanol 1-Decanol Ethyl dodecanoate (*E*)-Geraniol Isoamyl pentadecanoate 2-Ethyl hexanoic acid Ethyl tetradecanoate 2,4-Di-tert-butylphenol Dodecanoic acid
R8, R15, R17 & R20	0.90	0.88	Ethyl propanoate Ethyl valerate 1-Pentanol	2,3-Butanedione Isobutyl alcohol 2-Methyl-1-butanol

			Ethyl (*Z*)-3-Hexenoate Ethyl (*E*)-3-Hexenoate (*E*)-3-Hexenyl acetate (*Z*)-3-Hexen-1-ol 2-Ethyl-1-hexanol Diethyl succinate Ethyl cinnamate 2,4-Di-tert-butylphenol	3-Methyl-1-butanol 2-Heptanol 6-Methyl-5-hepten-2-ol Geranyl ethyl ether 2-Nonanol Ethyl decanoate (*Z*)-β-Farnesene Isoamyl octanoate Isovaleric acid Borneol Methyl salicylate Octanoic acid Decanoic acid
R10, R14, R16 & Shiraz	0.94	0.93	Propyl acetate Methyl hexanoate 1-Pentanol m-Cymene Propyl hexanoate Unknown monoterpene 1-Octen-3-ol Furfural Octyl acetate Unknown monoterpene Propyl octanoate Ethyl nonanoate 1-Nonanol Citronellyl acetate β-Ionone	2-Nonanone Borneol Thymol
R13 & Cab Sauv	0.93	0.91	Ethyl isobutyrate 2,3-Pentanedione Isobutyl alcohol Ethyl 4-methyl pentanoate 2-Methylbutyl isobutyrate 2-Methyl-1-butanol 4-Methyl-1-pentanol 3-Methyl- 1-pentanol Isobutyl hexanoate Furfural 3-Ethyl-4-methyl pentanol 2-Methyltetrahydrothiophen-3-one Isoamyl lactate Butanoic acid Isovaleric acid γ-Nonalactone Isoamyl alcohol	Propyl acetate 2,6,6-Trimethyl-2-vinyl-tetrahydropyran 1-Pentanol 3-Hydroxy-2-butanone Ethyl (*Z*)-3-hexenoate 2,2,6-Trimethyl-cyclohexanone (*E*)-3-Hexenyl acetate 3-Ethoxy-1-propanol Propyl octanoate Citronellol Hexanoic acid 2,6-Ditert-butyl-4-methylphenol

6.6.4 Summary

The volatile compound profiling of the wines allowed us to group similar wines based on the abundance of compounds extracted by the SPME fibre. Two methods of looking for relationships between the 1st generation mildew resistant wines gave similar results (Figures 90 & 91). The PCA plot (Figure 90) has the wines labelled according to their positioning in the cluster analysis (Figure 91) to see how the techniques overlap. Looking at both analyses, it appears that the cluster analysis accounts for more of the total variation as the positioning of the wines on the score plot from the first two PCs does not show as distinct a separation between some wines and groups as the cluster analysis. The fact that the first two PCs only account for a third of the total variation supports this.

The linking of the compound profiling to wine sensory attributes requires more study. Unfortunately, these wines had not undergone a full sensory descriptive analysis, which would have nicely complemented this study. However, there are some deficiencies in the volatile compound profiling methodology that will restrict our ability to predict wine style. Compounds present in very low concentrations in wines may not be detected even though they may have a sensory impact (e. g. rotundone and IBMP). Perhaps the methodology is better suited to detecting outliers that have a unique volatile profile and therefore may be of interest in an oenological or germplasm context. It was also found that although software for the identification of compounds and integration of peaks in the GC-MS chromatograms has improved, there is still a large amount of user time required to confirm the results of the automated analysis. These drawbacks, we believe, limit the use of volatile compound profiling by SPME-GC-MS as a selection tool for newly bred winegrape varieties.

7 Outcomes/Conclusions

There is a concerted effort from researchers around the world to increase our knowledge of how chemical compounds contribute to the sensory attributes of wine. However, there is a large gap in our understanding of how compounds in the grape berries contribute to the final aroma and flavour characteristics of the wine and, in turn, the genetic basis for the control of the production of these compounds. The project described in this report tested our ability to produce molecular markers of consumer-preferred sensory traits using a variety of methods and germplasm populations and has made considerable progress. The outputs listed below have the potential to greatly benefit the Australian wine sector, provide a solid basis for future work to determine the genetic basis of consumer-preferred traits, and indicate areas where future research will be beneficial.

- Linkage maps have been generated of established F1 populations from Mv × CS and Mv × R crosses and a high-density genetic map of an F2 population derived from an F1 progeny plant of the Mv × R cross. These maps can continue to be used while the germplasm exists and can be used for the mapping of any traits scored in the populations. In the same vein, we have also genotyped 177 varieties from the SARDI germplasm collection and 131 of the 1st generation mildew-resistant varieties and have identified thousands of markers that can be used in any future GWAS studies on any phenotype scored on these vines or a subset of them.
- Replicated wine volatile compound profiles from the same plants for which we have genetic linkage maps or SNP data have been produced. These data and chromatograms can be revisited as more becomes known about the role played by specific compounds in wine sensory characters to enable genes relating to their production to be mapped, if not already done so in the current study. We can also interrogate the data in different ways to identify varieties or progeny that produce wine with unique volatile compound profiles. Such plants can be useful in multiple ways. Such germplasm could be useful for a breeding strategy that aims to increase the production of a certain compound or groups of related compounds. It would also identify material for targeted crosses to understand the genetics of a certain trait. These plants may also be useful in understanding the role certain compounds play in specific sensory attributes if, for example, you could compare wines produced from plants that produce high concentrations of a compound with those where the compound is absent.
- A family of genes most likely to be involved in determining the concentration of (*Z*)-3-hexenol in wine has been identified. These genes were previously thought to encode seed storage proteins but were recently found to produce (*Z*)-3:(*E*)-2 hexenal isomerases (HI) that catalyse the isomerisation of (*Z*)-3-hexenal to (*E*)-2-hexenal (Kunishima, et al. 2016). Genetic markers from the genomic region where a number of these genes are located can predict (*Z*)-3-hexenol concentrations in wine from F1 progeny in crosses or the varieties in the germplasm collection. Expression analysis suggests that *VvHI2* is the most important member of this gene family in determining the amount of hexenal isomerase activity in ripe berries. We now have gene targets that can be investigated to explore the influence of viticultural treatments on the expression of these genes, which will predict the effect of vineyard or environmental variables on the concentration of (*Z*)-3-hexenol in wine.

- Results from these studies have confirmed that grape composition can greatly influence the production of wine acetate esters. Our research here takes this link one step further in identifying a region of the genome on the top of chromosome 18 that is associated with this phenomenon in the two biparental mapping experiments. As yet, we have not identified the causal gene for this trait. Its discovery would improve our ability to follow this trait in breeding strategies and allow us to monitor its expression in response to vineyard management and predict how these would affect the production of acetate esters during fermentation. It would also open opportunities for manipulation of this trait using gene-editing.
- Our experiments suggest that there is a genetic factor linked to the dwarf locus from the microvine that inhibits the production of monoterpenes. This is important for breeding strategies looking to produce varieties with moderate monoterpene production like Riesling and Viognier, and we have identified a microvine from the Mv × R F1 population that appears to have lost this inhibitory factor, which could be used in future breeding programs. In looking to identify the causal gene for this inhibition, we have shown that two grape NUDX hydrolases have the ability to degrade the precursors of both monoterpenes and sesquiterpenes which may make them interesting targets for gene-editing.
- QTLs have been identified for a range of other wine volatile compounds with our GWAS studies using the germplasm collection and the 1st generation mildew-resistant varieties. These can be used to generate SNP markers for screening of germplasm for desired traits. They can also be used to interrogate genomic data to create lists of potential candidate genes, from which the causal gene(s) can be discovered. This, in turn, can provide us with targets for gene editing to modify traits, or genes to be measured for changes in transcription during viticultural studies to gain insight into how to modulate selected traits in the vineyard. The key will be to prioritise those traits of most value to the Australian wine industry and decide which varieties to work with.
- CAPS and dCAPS markers have been generated for straightforward lab-based scoring of germplasm for the following traits: acylated (stable) anthocyanins, methoxypyrazine production in berries, α-guaiene production (rotundone precursor), red-fleshed berries and (*Z*)- 3-hexenol concentration in wine.
- Wine volatile compound profiling of the 20 selected red wine 1st generation mildew-resistant varieties was conducted to see if this technique would be useful in helping to predict wine style of breeding progeny without the need for micro-scale winemaking and sensory analysis. The results enabled clustering of the wines based on their composition and the grouping of some new varieties with Shiraz or Cabernet Sauvignon. However, the data analysis required for the profiling is labour-intensive at present and not all odour active compounds can be quantified with one method. Thus, this is not considered to be a high- throughput method for classifying new varieties with regard to their wine style. Such profiling may be more useful in identifying outlier plants producing wines with unusual compositions for further investigation.

- This project has enhanced the Australian wine industry's capability to conduct research into grape genetics and consumer-preferred sensory traits. We have established a methodological pipeline for genetic and genomic studies in grapevine that can be used to target consumer-preferred traits in any germplasm selection. The approach we have developed is unique in that we have phenotyped the plants according to the composition of the wine produced, not the grape, which adds to the existing research into wine flavour and aroma.
- The project has maintained capability in grape and wine flavour and aroma research and expanded our research skills. As mentioned above, datasets collected as part of this project are available to be used for subsequent research and will be mined further in the years to come. Furthermore, during the project the CSIRO laboratory on the Waite Campus has maintained a suite of analytical instruments for grape and wine metabolite studies which have also been used in collaborative studies with other researchers working on projects relevant to the Australian wine sector.

8 Recommendations

1. Continue scientific research

Our current research has generated a wealth of data and has identified a number of genome regions where genetic differences are associated with changes in the production of specific wine volatile compounds. There are a number of potential future research streams that could build on this work to provide a greater understanding of the role grape composition plays in consumer-preferred traits, improve germplasm and expand the methodology to include agronomic traits of interest:

Marker development and germplasm improvement

- Confirm GWAS QTL data with samples from other vintages or locations and design easy use markers (e.g. CAPS or iPLEX SNPs) for future breeding programs.
- Investigate more candidate genes in loci of interest (e.g. acetate ester concentration) and look to use next generation sequencing to more rapidly select or eliminate targets for characterisation and marker development.
- Use the knowledge gained from the microvine crosses and the markers to establish a program to stack positive consumer-preferred traits into superior breeding lines using the rapid generation times inherent in the microvine system.

Gene discovery and insight into traits

- Prioritise traits based on wine sector consultation (see below).
- Develop a gene characterisation pipeline using transformable microvines, hairy-root cultures and gene-editing technologies to produce lines overexpressing genes or with gene targets knocked-out to rapidly assess phenotype.
- Use our strong capability in analytical chemistry to phenotypically characterise germplasm with altered gene expression and identify those that are causing phenotypic differences.
- Combine gene knowledge with field-based viticulture trials to develop datasets that are less observational and, instead, generate robust knowledge as to how management and environment alter grape and wine composition.

Explore new traits/phenotypes

- Mine the wine volatile compound data obtained from the SARDI germplasm collection and 1st generation mildew-resistant varieties to identify plants with extreme phenotypes for the generation of new F1 populations by crossing with the microvine. These populations can then be used in the QTL mapping pipeline established in this project for gene discovery.
- Use the agronomic data generated in Project CSA 1701-2.2 to identify QTLs and candidate genes associated with priority agronomic traits and feed them into the pipeline for gene discovery as outlined above.

- Continue to keep abreast of the latest plant science literature and apply knowledge from other species to improve grapegrowing through germplasm improvement or novel management technologies.

2. Achieve sector impact

Sector impact can be achieved through two main pathways: 1) the creation of improved germplasm and 2) a more informed sector with tools to use to predict the effects of environmental change and management on grape and wine composition.

The development of CRISPR/Cas9 gene editing technology offers the possibility of genetically modifying specific traits in existing premium varieties without these plants having to be labelled as GMOs. This requires the use of 'DNA-free' gene editing which has two components: the efficient delivery of CRISPR/Cas9 editing machinery into grape protoplasts so that gene editing can take place without integration of the CRISPR/Cas9 transgene into the plant genome and the regeneration of whole plants from gene-edited protoplasts. It should also be remembered that gene editing and trait discovery methodologies do not exist in isolation and having integrated projects looking at both will give us the capacity to be agile and speed up delivery depending on the priorities of the wine sector and the maturity of the science for each trait. The vision is to develop a pipeline of novel clones of existing wine grape varieties with improved traits that deliver significant impact to the Australian wine sector.

Additional benefits accrue from determining the genetics of consumer-preferred traits, as knowledge of the genes that regulate traits will allow us to understand the biochemistry behind the production of important molecules in grapes and wine. This, in turn, leads to strategies to manipulate the production of these compounds, both in the vineyard and the winery, to control wine style in existing cultivars. The same is true for agronomic traits, which can be better understood through identification of causal genes whose expression can be examined in vineyard trials. We would aim to partner with wine sector collaborators or other research providers to add value to existing experiments and increase wine sector awareness of research outcomes.

3. Consult with industry on research priorities

As outlined above, an exciting extension of the work in this report is the use of the knowledge, data and materials generated to inform strategies to improve germplasm through the use of DNA-free gene-editing technologies. The concept of gene editing for the production of improved clones of existing varieties has been presented to the Research Advisory Committee of Australian Grape & Wine and subsequent positive feedback led to the drafting of a new research proposal to pursue this strategy. The key is to place research resources into traits that will deliver the best return on investment for the Australian wine sector, and this will involve extensive consultation with all stakeholders. There is also the need to balance expectations for this technology with what can be achieved in the short term given our knowledge of gene targets for important traits and the fact that it is simpler to produce gene knockouts than other outcomes such as gene over-expression or altered protein function.

9 Communication

Seminars and conference participation

Boss PK, Hassan A, Dunlevy JD, Nicholson EL, Davies C. 'Developing a new generation of grapevine cultivars: markers for consumer traits'. Queenstown Molecular Biology (QMB) Plant Satellite meeting, Queenstown, NZ, Aug 2018. Invited speaker.

Boss PK. 'Developing a new generation of grapevine cultivars: markers for consumer traits'. Waite Campus Seminar, Adelaide, South Australia, Sep 2018. Invited speaker.

Boss P, Kalua C, Nicholson E, Maffei S, Böttcher C, Davies C. 'Developmental changes in grape composition and the implications for wine volatile profiles. 17th Australian Wine Industry Technical Conference, Adelaide, Australia, Jul 2019. Poster.

Hassan A. Identification of a grapevine gene family that influences (Z)-3-hexenol concentration in wine. Crush; The Grape and Wine Science Symposium, June 2021, Adelaide, Australia. Speaker.

Publications arising from collaborations associated with this project

Armstrong C, Gilmore A, Boss PK, Pagay V, Jeffery DW (2022) Machine learning for classifying and predicting grape maturity indices using absorbance and fluorescence spectra. Food Chemistry 134321

Sanders R, Boss PK, Capone DL, Kidman C, Bramley R, Nicholson EL, Jeffery DW (2022) Rootstock, vine vigor, and light mediate methoxypyrazine concentrations in the grape bunch rachis of *Vitis vinifera* L. cv. Cabernet Sauvignon. Journal of Agricultural and Food Chemistry 70: 5417-5426

Lang TA, Walker ME, Boss PK, Jiranek V (2022) Effect of 'loss of function' mutation in SER1 in wine yeast: Fermentation outcomes in co-inoculation with non-*Saccharomyces*. Oeno One 56: 47-61.

Capone DL, Francis IL, Clingeleffer RR, Maffei SM, Boss PK (2022) Evidence that methoxypyrazine accumulation is elevated in Shiraz rachis grown on Ramsey rootstock, increasing 'green' flavour in wine. Australian Journal of Grape and Wine Research 28: 304-315.

Lin MMH, Boss PK, Walker ME, Sumby KM, Jiranek V (2022) Influence of *Kazachstania* spp. on the chemical and sensory profile of red wines. International Journal of Food Microbiology 362: 109496.

Gardner JM, Walker ME, Boss PK, Jiranek V (2021) The effect of grape juice dilution and complex nutrient addition on oenological fermentation and wine chemical composition. Journal of Food Composition and Analysis 105: 104241.

Davies C, Böttcher C, Nicholson EL, Burbidge CA, Boss PK (2021) Timing of auxin treatment affects grape berry growth, ripening timing, and the synchronicity of sugar accumulation. Australian Journal of Grape and Wine Research 28: 232-241.

Böttcher C, Johnson TE, Burbidge CA, Nicholson EL, Boss PK, Maffei SM, Bastian SEP, Davies C (2021) Use of auxin to delay ripening – sensory and biochemical evaluation of Cabernet Sauvignon and Shiraz. Australian Journal of Grape and Wine Research 28: 208-217.

Armstrong C, Ristic R, Boss PK, Pagay V, Jeffery DW (2021) Investigating the effect of grape heterogeneity on wine chemical composition and sensory attributes for *Vitis vinifera* cv. Cabernet Sauvignon. Australian Journal of Grape and Wine Research 27: 206-218.

Dal Santo S, Tucker MR, Tan HT, Burbidge CA, Fasoli M, Böttcher C, Boss PK, Pezzotti M, Davies C (2020) Auxin treatment of grapevine (*Vitis vinifera* L.) berries delays ripening onset by inhibiting cell expansion. Plant Molecular Biology 103: 91-111.

Lin MMH, Boss PK, Walker ME, Sumby KM, Grbin PR, Jiranek V (2020) Evaluation of indigenous non-*Saccharomyces* yeasts isolated from a South Australian vineyard for their potential as wine starter cultures. International Journal of Food Microbiology 312: 108373.

Böttcher C, Boss PK, Harvey KE, Nicholson, EL, Burbidge CA, Davies C (2019) Post-veraison restriction of phloem import into Riesling (*Vitis vinifera* L.) berries induces transient and stable changes to fermentation-derived and varietal wine volatiles. Australian Journal of Grape and Wine Research 25: 286-292.

Hranilovic A, Gambetta J, Schmidtke L, Boss PK, Grbin P, Masneuf-Pomarede I, Bely M, Albertin W, Jiranek V (2018) Oenological traits of *Lachancea thermotolerans* show signs of domestication and allopatric differentiation. Scientific Reports 8: 14812.

Liang C, Boss PK, Jeffery DW (2018) Extraction properties of new polymeric sorbents applied to wine. Journal of Agricultural and Food Chemistry 66: 10086-10096.

Schelezki OJ, Suklje K, Boss PK, Jeffery DW (2018) Comparison of consecutive harvests versus blending treatments to produce lower alcohol wines from Cabernet Sauvignon grapes: Impact on wine volatile composition and sensory properties. Food Chemistry 259: 196-206.

Hranilovic A, Li S, Ristic R, Boss PK, Bindon K, Grbin P, van der Westhuizen T, Jiranek V (2018) Chemical and sensory profiling of Shiraz wines co-fermented with commercial non-*Saccharomyces* inocula. Australian Journal of Grape and Wine Research 24: 166-180.

Böttcher C, Boss PK, Harvey KE, Burbidge CA, Davies C (2018) Peduncle-girdling of Shiraz (*Vitis vinifera* L.) bunches alters amino acid and hormone content of grapes and affects wine volatile compounds depending on berry sugar levels. Australian Journal of Grape and Wine Research 24: 206-281.

10 Intellectual Property

- Knowledge of genes that control pathways that influence the specific wine volatile compounds.
- Lists of candidate genes associated with traits that influence consumer preference.
- A mapping population consisting of 84 progeny from the self-pollination of F1 plant #131 from a microvine x Riesling cross.
- DArT genomic analysis data from i) microvine x Riesling F1 population; ii) microvine x Cabernet Sauvignon F1 population; iii) self-pollinated microvine x Riesling F1 plant #131 F2 population; iv) 1st generation tall progeny listed in SRA IP register in tab named "BIP – 1st Gen TALLs"; v) 175 red and white grape cultivars from the SARDI collection at Nuriootpa.

11 References

Anaraki, Z.E., S.A.H. Tafreshi, and M. Shariati (2018) Transient silencing of heat shock proteins showed remarkable roles for HSP70 during adaptation to stress in plants. Environmental and Experimental Botany 155, 142-157.

Barkan, A. and I. Small (2014) Pentatricopeptide repeat proteins in plants. Annual Review of Plant Biology 65, 415-442.

Battilana, J., L. Costantini, F. Emanuelli, F. Sevini, C. Segala, S. Moser, R. Velasco, G. Versini, and M.S. Grando (2009) The 1-deoxy-d-xylulose 5-phosphate synthase gene co-localizes with a major QTL affecting monoterpene content in grapevine. Theoretical and Applied Genetics 118, 653-669. doi: 10.1007/s00122-008-0927-8.

Battilana, J., F. Emanuelli, G. Gambino, I. Gribaudo, F. Gasperi, P.K. Boss, and M.S. Grando (2011) Functional effect of grapevine 1-deoxy-D-xylulose 5-phosphate synthase substitution K284N on Muscat flavour formation. Journal of Experimental Botany 62, 5497-5508. doi: 10.1093/jxb/err231.

Bi, C., Y. Ma, X.-F. Wang, and D.-P. Zhang (2017) Overexpression of the transcription factor NF-YC9 confers abscisic acid hypersensitivity in Arabidopsis. Plant Molecular Biology 95, 425-439.

Blackford, C.L., E.G. Dennis, R.A. Keyzers, C. Schueuermann, R.D. Trengove, and P.K. Boss (2018) Fermentation-guided natural products isolation of a grape berry triacylglyceride that enhances ethyl ester production. Molecules 23, 152.

Blackford, C.L., R.D. Trengove, and P.K. Boss (2022) Exploring the influence of grape tissues on the concentration of wine volatile compounds. Australian Journal of Grape and Wine Research 28, 218-231.

Blank, I., A. Sen, and W. Grosch (1992) Sensory study on the character-impact flavour compounds of dill herb (*Anethum graveolens* L.). Food Chemistry 43, 337-343.

Bonk, M., M. Tadros, J. Vandekerckhove, S. Al-Babili, and P. Beyer (1996) Purification and characterization of chaperonin 60 and heat-shock protein 70 from chromoplasts of *Narcissus pseudonarcissus* (Involvement of heat-shock protein 70 in a soluble protein complex containing phytoene desaturase). Plant Physiology 111, 931-939.

Boss, P.K., C. Bottcher, and C. Davies (2014) Various influences of harvest date and fruit sugar content on different wine flavor and aroma compounds. American Journal of Enology and Viticulture 65, 341-353. doi: 10.5344/ajev.2014.13137.

Boss, P.K., C. Davies, and S.P. Robinson (1996) Analysis of the expression of anthocyanin pathway genes in developing *Vitis vinifera* L cv Shiraz grape berries and the implications for pathway regulation. Plant Physiology 111, 1059-1066.

Boss, P.K., C. Kalua, E. Nicholson, S. Maffei, C. Böttcher, and C. Davies (2018) Fermentation of grapes throughout development identifies stages critical to the development of wine volatile composition. Australian Journal of Grape and Wine Research 24, 24-37.

Boss, P.K., A.D. Pearce, Y.J. Zhao, E.L. Nicholson, E.G. Dennis, and D.W. Jeffery (2015) Potential grape-derived contributions to volatile ester concentrations in wine. Molecules 20, 7845-7873. doi: 10.3390/molecules20057845.

Boss, P.K. and M.R. Thomas (2002) Association of dwarfism and floral induction with a grape 'green revolution' mutation. Nature 416, 847-850.

Böttcher, C., C.A. Burbidge, P.K. Boss, and C. Davies (2013a) Interactions between ethylene and auxin are crucial to the control of grape (*Vitis vinifera* L.) berry ripening. BMC Plant Biology 13, 1-14.

Böttcher, C., C.A. Burbidge, P.K. Boss, and C. Davies (2015) Changes in transcription of cytokinin metabolism and signalling genes in grape (*Vitis vinifera* L.) berries are associated with the ripening-related increase in isopentenyladenine. BMC Plant Biology 15, 1-15.

Böttcher, C., K.E. Harvey, P.K. Boss, and C. Davies (2013b) Ripening of grape berries can be advanced or delayed by reagents that either reduce or increase ethylene levels. Functional Plant Biology 40, 566-581. doi: 10.1071/fp12347.

Böttcher, C., R.A. Keyzers, P.K. Boss, and C. Davies (2010) Sequestration of auxin by the indole-3-acetic acid-amido synthetase GH3-1 in grape berry (*Vitis vinifera* L.) and the proposed role of auxin conjugation during ripening. Journal of Experimental Botany 61, 3615-3625.

Chaib, J., L. Torregrosa, D. Mackenzie, P. Corena, A. Bouquet, and M.R. Thomas (2010) The grape microvine - a model system for rapid forward and reverse genetics of grapevines. Plant Journal 62, 1083-1092. doi: 10.1111/j.1365-313X.2010.04219.x.

Churchill, G.A. and R. Doerge (1994) Empirical threshold values for quantitative trait mapping. Genetics 138, 963-971.

Cleary, M., H. Chong, N. Ebisuda, N. Dokoozlian, N. Loscos, B. Pan, D. Santino, Q. Sui, and C. Yonker (2015) Objective chemical measures of grape quality. In: Advances in Wine Research (ACS Publications) pp. 365-378.

Cortés-Diéguez, S., R. Rodriguez-Solana, J.M. Domínguez, and E. Díaz (2015) Impact odorants and sensory profile of young red wines from four Galician (NW of Spain) traditional cultivars. Journal of the Institute of Brewing 121, 628-635.

Costantini, L., J. Battilana, F. Lamaj, G. Fanizza, and M.S. Grando (2008) Berry and phenology-related traits in grapevine (*Vitis vinifera* L.): from quantitative trait loci to underlying genes. BMC Plant Biology 8, 1-17.

D'Onofrio, C., A. Cox, C. Davies, and P.K. Boss (2009) Induction of secondary metabolism in grape cell cultures by jasmonates. Functional Plant Biology 36, 323-338. doi: 10.1071/fp08280.

De-La-Fuente-Blanco, A., M.-P. Sáenz-Navajas, and V. Ferreira (2016) On the effects of higher alcohols on red wine aroma. Food Chemistry 210, 107-114.

de la Torre, F., R.A. Cañas, M.B. Pascual, C. Avila, and F.M. Cánovas (2014) Plastidic aspartate aminotransferases and the biosynthesis of essential amino acids in plants. Journal of Experimental Botany 65, 5527-5534.

Dennis, E.G., R.A. Keyzers, C.M. Kalua, S.M. Maffei, E.L. Nicholson, and P.K. Boss (2012) Grape contribution to wine aroma: Production of hexyl acetate, octyl acetate, and benzyl acetate during yeast fermentation is dependent upon precursors in the must. Journal of Agricultural and Food Chemistry 60, 2638-2646. doi: 10.1021/jf2042517.

Drew, D.P., T.B. Andersen, C. Sweetman, B.L. Møller, C. Ford, and H.T. Simonsen (2016) 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. Journal of Experimental Botany 67, 799-808.

Dunlevy, J.D., E.G. Dennis, K.L. Soole, M.V. Perkins, C. Davies, and P.K. Boss (2013) A methyltransferase essential for the methoxypyrazine-derived flavour of wine. The Plant Journal 75, 606-617. doi: 10.1111/tpj.12224.

Dunlevy, J.D., C.M. Kalua, R.A. Keyzers, and P.K. Boss (2009) The production of flavour and aroma compounds in grape berries. In: Grapevine Molecular Physiology and Biotechnology, Ed. K.A. Roubelakis-Angelakis (Springer: Dordrecht) pp. 293-340.

Dunwell, J.M., J.G. Gibbings, T. Mahmood, and S.M. Saqlan Naqvi (2008) Germin and Germin-like Proteins: Evolution, Structure, and Function. Critical Reviews in Plant Sciences 27, 342-375. doi: 10.1080/07352680802333938.

Earley, K.W., J.R. Haag, O. Pontes, K. Opper, T. Juehne, K. Song, and C.S. Pikaard (2006) Gateway-compatible vectors for plant functional genomics and proteomics. The Plant Journal 45, 616-629.

Elsharif, S.A. and A. Buettner (2017) Influence of the chemical structure on the odor characters of β-citronellol and its oxygenated derivatives. Food Chemistry 232, 704-711.

Emanuelli, F., J. Battilana, L. Costantini, L. Le Cunff, J.M. Boursiquot, P. This, and M.S. Grando (2010) A candidate gene association study on muscat flavor in grapevine (*Vitis vinifera* L.). BMC Plant Biology 10, 241. doi: 10.1186/1471-2229-10-241.

Emanuelli, F., M. Sordo, S. Lorenzi, J. Battilana, and M. Grando (2014) Development of user-friendly functional molecular markers for *VvDXS* gene conferring muscat flavor in grapevine. Molecular Breeding 33, 235-241.

Escudero, A., P. Hernández-Orte, J. Cacho, and V. Ferreira (2000) Clues about the role of methional as character impact odorant of some oxidized wines. Journal of Agricultural and Food Chemistry 48, 4268-4272.

Espinase Nandorfy, D., T. Siebert, F. Watson, R. Keast, and I. Francis (2022) Understanding the interactive effects of volatile compounds contributing to 'stone fruit'aroma nuances in white wines. Australian Journal of Grape and Wine Research 28, 424-438.

Fang, Y. and M. Qian (2005) Aroma compounds in Oregon Pinot Noir wine determined by aroma extract dilution analysis (AEDA). Flavour and Fragrance Journal 20, 22-29.

Genovese, A., L. Moio, R. Sacchi, and P. Piombino (2015) Sip volume affects oral release of wine volatiles. Food Research International 77, 426-431.

Goh, L. and V.B. Yap (2009) Effects of normalization on quantitative traits in association test. BMC Bioinformatics 10, 1-8.

Gómez-Míguez, M.J., J.F. Cacho, V. Ferreira, I.M. Vicario, and F.J. Heredia (2007) Volatile components of Zalema white wines. Food Chemistry 100, 1464-1473.

Gruber, B., P.J. Unmack, O.F. Berry, and A. Georges (2018) dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Molecular Ecology Resources 18, 691-699.

Guillaumie, S., A. Ilg, S. Rety, M. Brette, C. Trossat-Magnin, S. Decroocq, C. Leon, C. Keime, T. Ye, R. Baltenweck-Guyot, P. Claudel, L. Bordenave, S. Vanbrabant, E. Duchene, S. Delrot, P. Darriet, P. Hugueney, and E. Gomes (2013) Genetic analysis of the biosynthesis of 2-methoxy-3-isobutylpyrazine, a major grape-derived aroma compound impacting wine quality. Plant Physiology 162, 604-615. doi: 10.1104/pp.113.218313.

Hatanaka, A. (1993) The biogeneration of green odor by green leaves. Phytochemistry 34, 1201-1218.

Hein, K., S.E. Ebeler, and H. Heymann (2009) Perception of fruity and vegetative aromas in red wine. Journal of Sensory Studies 24, 441-455. doi: 10.1111/j.1745-459X.2009.00220.x.

Henry, L.K., S.T. Thomas, J.R. Widhalm, J.H. Lynch, T.C. Davis, S.A. Kessler, J. Bohlmann, J.P. Noel, and N. Dudareva (2018) Contribution of isopentenyl phosphate to plant terpenoid metabolism. Nature Plants 4, 721-729.

Hernández-Orte, P., M. Ibarz, J. Cacho, and V. Ferreira (2006) Addition of amino acids to grape juice of the Merlot variety: Effect on amino acid uptake and aroma generation during alcoholic fermentation. Food Chemistry 98, 300-310. doi: 10.1016/j.foodchem.2005.05.073.

Hjelmeland, A.K., J. Zweigenbaum, and S.E. Ebeler (2015) Profiling monoterpenol glycoconjugation in *Vitis vinifera* L. cv. Muscat of Alexandria using a novel putative compound database approach, high resolution mass spectrometry and collision induced dissociation fragmentation analysis. Analytica Chimica Acta 887, 138-147.

Huang, M., X. Liu, Y. Zhou, R.M. Summers, and Z. Zhang (2019) BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. Gigascience 8, giy154.

Iocco-Corena, P., J. Chaïb, L. Torregrosa, D. Mackenzie, M.R. Thomas, and H.M. Smith (2021) VviPLATZ1 is a major factor that controls female flower morphology determination in grapevine. Nature Communications 12, 1-10.

Ito, D., T. Kato, T. Maruta, M. Tamoi, K. Yoshimura, and S. Shigeoka (2012) Enzymatic and molecular characterization of Arabidopsis ppGpp pyrophosphohydrolase, AtNUDX26. Bioscience, Biotechnology, and Biochemistry 76, 2236-2241.

Kalua, C.M. and P.K. Boss (2010) Comparison of major volatile compounds from Riesling and Cabernet Sauvignon grapes (*Vitis vinifera* L.) from fruitset to harvest. Australian Journal of Grape and Wine Research 16, 337-348. doi: 10.1111/j.1755-0238.2010.00096.x.

Keyzers, R.A. and P.K. Boss (2010) Changes in the volatile compound production of fermentations made from musts with increasing grape content. Journal of Agricultural and Food Chemistry 58, 1153-1164. doi: 10.1021/jf9023646.

Kobayashi, S., N. Goto-Yamamoto, and H. Hirochika (2004) Retrotransposon-induced mutations in grape skin color. Science 304, 982-982. doi: 10.1126/science.1095011.

Koch, A., C.L. Doyle, M.A. Matthews, L.E. Williams, and S.E. Ebeler (2010) 2-Methoxy-3-isobutylpyrazine in grape berries and its dependence on genotype. Phytochemistry 71, 2190-2198. doi: 10.1016/j.phytochem.2010.09.006.

Korte, A. and A. Farlow (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9, 1-9.

Kumar, M., D.T. Le, S. Hwang, P.J. Seo, and H.U. Kim (2019) Role of the INDETERMINATE DOMAIN Genes in Plants. International Journal of Molecular Sciences 20, 2286.

Kunishima, M., Y. Yamauchi, M. Mizutani, M. Kuse, H. Takikawa, and Y. Sugimoto (2016) Identification of (Z)-3:(E)-2-Hexenal Isomerases Essential to the Production of the Leaf Aldehyde in Plants. Journal of Biological Chemistry 291, 14023-14033. doi: 10.1074/jbc.M116.726687.

La Guerche, S., B. Dauphin, M. Pons, D. Blancard, and P. Darriet (2006) Characterization of some mushroom and earthy off-odors microbially induced by the development of rot on grapes. Journal of Agricultural and Food Chemistry 54, 9193-9200.

Leonardi, R. and S. Jackowski (2007) Biosynthesis of pantothenic acid and coenzyme A. EcoSal Plus 2.

Liu, X., M. Huang, B. Fan, E.S. Buckler, and Z. Zhang (2016) Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genetics 12, e1005767.

Lu, S., J. Van Eck, X. Zhou, A.B. Lopez, D.M. O'Halloran, K.M. Cosman, B.J. Conlin, D.J. Paolillo, D.F. Garvin, and J. Vrebalov (2006) The cauliflower *Or* gene encodes a DnaJ cysteine-rich domain-containing protein that mediates high levels of β-carotene accumulation. The Plant Cell 18, 3594-3605.

Luan, F. and M. Wüst (2002) Differential incorporation of 1-deoxy-D-xylulose into (3S)-linalool and geraniol in grape berry exocarp and mesocarp. Phytochemistry 60, 451-459.

Lytra, G., M. Cameleyre, S. Tempere, and J.-C. Barbe (2015) Distribution and organoleptic impact of ethyl 3-hydroxybutanoate enantiomers in wine. Journal of Agricultural and Food Chemistry 63, 10484-10491.

Lytra, G., S. Tempere, A. Le Floch, G. de Revel, and J.-C. Barbe (2013) Study of sensory interactions among red wine fruity esters in a model solution. Journal of Agricultural and Food Chemistry 61, 8504-8513.

Magnard, J.-L., A. Roccia, J.-C. Caissard, P. Vergne, P. Sun, R. Hecquet, A. Dubois, L. Hibrand-Saint Oyant, F. Jullien, and F. Nicolè (2015) Biosynthesis of monoterpene scent compounds in roses. Science 349, 81-83.

Marigorta, U.M., J.A. Rodríguez, G. Gibson, and A. Navarro (2018) Replicability and prediction: lessons and challenges from GWAS. Trends in Genetics 34, 504-517.

Marrano, A., G. Birolo, M.L. Prazzoli, S. Lorenzi, G. Valle, and M.S. Grando (2017) SNP-discovery by RAD-sequencing in a germplasm collection of wild and cultivated grapevines (*V. vinifera* L.). PLOS ONE 12, e0170655.

Mattivi, F., L. Caputi, S. Carlin, T. Lanza, M. Minozzi, D. Nanni, L. Valenti, and U. Vrhovsek (2011) Effective analysis of rotundone at below-threshold levels in red and white wines using solid-phase microextraction gas chromatography/tandem mass spectrometry. Rapid Communications in Mass Spectrometry 25, 483-488.

Mendel, G. (1866) Versuche über Plflanzen-hybriden. Verhandlungen des naturforschenden Ver-eines in Brünn, Bd. IV für das Jahr 1865, Abhand-lungen,, 3-47. doi: www.esp.org/foundations/genetics/classical/gm-65.pdf (English translation).

Migicovsky, Z., J. Sawler, K.M. Gardner, M.K. Aradhya, B.H. Prins, H.R. Schwaninger, C.D. Bustamante, E.S. Buckler, G.-Y. Zhong, and P.J. Brown (2017) Patterns of genomic and phenomic diversity in wine and table grapes. Horticulture Research 4.

Myles, S., A.R. Boyko, C.L. Owens, P.J. Brown, F. Grassi, M.K. Aradhya, B. Prins, A. Reynolds, J.M. Chia, D. Ware, C.D. Bustamante, and E.S. Buckler (2011) Genetic structure and domestication history of the grape. Proceedings of the National Academy of Sciences of the United States of America 108, 3530-3535. doi: 10.1073/pnas.1009363108.

Pauquet, J., A. Bouquet, P. This, and A.-F. Adam-Blondon (2001) Establishment of a local map of AFLP markers around the powdery mildew resistance gene *Run1* in grapevine and assessment of their usefulness for marker assisted selection. Theoretical and Applied Genetics 103, 1201-1210.

Perez-Cacho, P.R. and R. Rouseff (2008) Processing and storage effects on orange juice aroma: A review. Journal of Agricultural and Food Chemistry 56, 9785-9796.

Pineau, B., J.C. Barbe, C. Van Leeuwen, and D. Dubourdieu (2007) Which impact for β-damascenone on red wines aroma? Journal of Agricultural and Food Chemistry 55, 4103-4108.

R Core Team (2019) 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria: Available at: https://www. R-project. org/.

Rinaldo, A.R., E. Cavallini, Y. Jia, S.M. Moss, D.A. McDavid, L.C. Hooper, S.P. Robinson, G.B. Tornielli, S. Zenoni, and C.M. Ford (2015) A grapevine anthocyanin acyltransferase, transcriptionally regulated by VvMYBA, can produce most acylated anthocyanins present in grape skins. Plant Physiology 169, 1897-1916.

Robinson, A.L., P.K. Boss, P.S. Solomon, R.D. Trengove, H. Heymann, and S.E. Ebeler (2014) Origins of grape and wine aroma. Part 1. Chemical components and viticultural impacts. American Journal of Enology and Viticulture 65, 1-24. doi: 10.5344/ajev.2013.12070.

Robinson, J., J. Harding, and J. Vouillamoz (2013) Wine grapes: a complete guide to 1,368 vine varieties, including their origins and flavours. (Penguin: UK).

Röckel, F., C. Moock, U. Braun, F. Schwander, P. Cousins, E. Maul, R. Töpfer, and L. Hausmann (2020) 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 11, 891.

Rodríguez-Flores, M.S., S.I. Falcão, O. Escuredo, M.C. Seijo, and M. Vilas-Boas (2021) Description of the volatile fraction of Erica honey from the northwest of the Iberian Peninsula. Food Chemistry 336, 127758.

Scacco, A., C.M. Lanza, A. Mazzaglia, G. Tripodi, G. Dima, and A. Verzera (2010) Correlation between aroma compounds and sensory properties of Passito Malvasia wines produced in Sicily. American Journal of Enology and Viticulture 61, 260-265.

Segura, V., B.J. Vilhjálmsson, A. Platt, A. Korte, Ü. Seren, Q. Long, and M. Nordborg (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nature Genetics 44, 825-830.

Segurel, M.A., A.J. Razungles, C. Riou, M. Salles, and R.L. Baumes (2004) Contribution of dimethyl sulfide to the aroma of Syrah and Grenache Noir wines and estimation of its potential in grapes of these varieties. Journal of Agricultural and Food Chemistry 52, 7084-7093.

Slaghenaufi, D., L. Vanzo, G. Luzzini, P. Arapitsas, M. Marangon, A. Curioni, F. Mattivi, P. Piombino, L. Moio, and A. Versari (2022) Monoterpenoids and norisoprenoids in Italian red wines. Oeno One 56, 185-193.

Spyropoulou, E.A., H.L. Dekker, L. Steemers, J.H. van Maarseveen, C.G. de Koster, M.A. Haring, R.C. Schuurink, and S. Allmann (2017) Identification and characterization of (3*Z*):(2*E*)-hexenal isomerases from cucumber. Frontiers in Plant Science 8, 1342.

Srouji, J.R., A. Xu, A. Park, J.F. Kirsch, and S.E. Brenner (2017) The evolution of function within the Nudix homology clan. Proteins: Structure, Function, and Bioinformatics 85, 775-811.

Sumby, K.M., P.R. Grbin, and V. Jiranek (2010) Microbial modulation of aromatic esters in wine: Current knowledge and future prospects. Food Chemistry 121, 1-16. doi: 10.1016/j.foodchem.2009.12.004.

Takase, H., K. Sasaki, H. Shinmori, A. Shinohara, C. Mochizuki, H. Kobayashi, G. Ikoma, H. Saito, H. Matsuo, S. Suzuki, and R. Takata (2015) Cytochrome P450 *CYP71BE5* in grapevine (*Vitis vinifera*) catalyzes the formation of the spicy aroma compound (−)-rotundone. Journal of Experimental Botany 67, 787-798. doi: 10.1093/jxb/erv496.

Tao, Y. and L. Zhang (2010) Intensity prediction of typical aroma characters of Cabernet Sauvignon wine in Changli County (China). LWT-Food Science and Technology 43, 1550-1556.

Thangaleela, S., B.S. Sivamaruthi, P. Kesika, T. Tiyajamorn, M. Bharathi, and C. Chaiyasut (2022) A Narrative Review on the Bioactivity and Health Benefits of Alpha-Phellandrene. Scientia Pharmaceutica 90, 57.

Tibbs Cortes, L., Z. Zhang, and J. Yu (2021) Status and prospects of genome-wide association studies in plants. The Plant Genome 14, e20077.

Van Buren, J., J. Bertino, and W. Robinson (1968) The stability of wine anthocyanins on exposure to heat and light. American Journal of Enology and Viticulture 19, 147-154.

Van Ooijen, J. (2006) JoinMap 4. Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherlands 33.

Van Ooijen, J. (2009) MapQTL® 6, Software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma BV, Wageningen, Netherlands 64.

Viana, F., J.V. Gil, S. Vallés, and P. Manzanares (2009) Increasing the levels of 2-phenylethyl acetate in wine through the use of a mixed culture of *Hanseniaspora osmophila* and *Saccharomyces cerevisiae*. International Journal of Food Microbiology 135, 68-74.

Walker, A.R., E. Lee, J. Bogs, D.A.J. McDavid, M.R. Thomas, and S.P. Robinson (2007) White grapes arose through the mutation of two similar and adjacent regulatory genes. Plant Journal 49, 772-785. doi: 10.1111/j.1365-313X.2006.02997.x.

Wang, J. and Z. Zhang (2021) GAPIT Version 3: boosting power and accuracy for genomic association and prediction. Genomics, Proteomics & Bioinformatics 19, 629-640.

Wang, Y., J. O'Reilly, Y. Chen, and J. Pawliszyn (2005) Equilibrium in-fibre standardisation technique for solid-phase microextraction. Journal of Chromatography A 1072, 13-17.

Watson, J.D. and F.H. Crick (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. Nature 171, 737-738.

Weingart, G., B. Kluger, A. Forneck, R. Krska, and R. Schuhmacher (2012) Establishment and application of a metabolomics workflow for identification and profiling of volatiles from leaves of *Vitis vinifera* by HS-SPME-GC-MS. Phytochemical Analysis 23, 345-358.

Williams, S.J., L. Yin, G. Foley, L.W. Casey, M.A. Outram, D.J. Ericsson, J. Lu, M. Boden, I.B. Dry, and B. Kobe (2016) Structure and function of the TIR domain from the grape NLR protein RPV1. Frontiers in Plant Science 7, 1850.

Winterhalter, P. and R. Gök (2013) TDN and β-Damascenone: Two important carotenoid metabolites in wine. In: Carotenoid cleavage products (ACS Publications) pp. 125-137.

Wood, C., T.E. Siebert, M. Parker, D.L. Capone, G.M. Elsey, A.P. Pollnitz, M. Eggers, M. Meier, T. Vossing, S. Widder, G. Krammer, M.A. Sefton, and M.J. Herderich (2008) From wine to pepper: Rotundone, an obscure sesquiterpene, is a potent spicy aroma compound. Journal of Agricultural and Food Chemistry 56, 3738-3744.

Xu, P., F. Zhu, G.K. Buss, and W.S. Leal (2015) 1-Octen-3-ol – the attractant that repels. F1000Research 4, 156.

Yano, K., S. Yoshida, J. Müller, S. Singh, M. Banba, K. Vickers, K. Markmann, C. White, B. Schuller, and S. Sato (2008) CYCLOPS, a mediator of symbiotic intracellular accommodation. Proceedings of the National Academy of Sciences of the United States of America 105, 20540-20545.

Yao, X., J. Nie, R. Bai, and X. Sui (2020) Amino acid transporters in plants: Identification and function. Plants 9, 972.

Yonekura-Sakakibara, K., T. Nakayama, M. Yamazaki, and K. Saito (2008) Modification and stabilization of anthocyanins. In: Anthocyanins (Springer) pp. 169-190.

Yoshimura, K., T. Ogawa, Y. Ueda, and S. Shigeoka (2007) AtNUDX1, an 8-oxo-7, 8-dihydro-2′-deoxyguanosine 5′-triphosphate pyrophosphohydrolase, is responsible for eliminating oxidized nucleotides in *Arabidopsis*. Plant and Cell Physiology 48, 1438-1449.

Yoshimura, K. and S. Shigeoka (2015) Versatile physiological functions of the Nudix hydrolase family in *Arabidopsis*. Bioscience, Biotechnology, and Biochemistry 79, 354-366.

Yu, H., T. Xie, J. Xie, L. Ai, and H. Tian (2019) Characterization of key aroma compounds in Chinese rice wine using gas chromatography-mass spectrometry and gas chromatography-olfactometry. Food Chemistry 293, 8-14.

12 Staff

All staff below spent periods of time on the project:

Emily L. Nicholson

Dr Ali Hassan

Dr Christopher Davies

Dr Paul K. Boss



As Australia's national science agency and innovation catalyst, CSIRO is solving the greatest challenges through innovative science and technology.

CSIRO. Unlocking a better future for everyone.

Contact us

1300 363 400
+61 3 9545 2176
csiro.au/contact
csiro.au

For further information

Agriculture & Food
Paul Boss
+61 8 8303 8614
paul.boss@csiro..au
https://people-my.csiro.au/b/p/paul-boss