The genetic basis of acid composition in developing berries of the cultivated grapevine *Vitis vinifera*

Emily Grace Higginson

School of Agriculture, Food and Wine

University of Adelaide

December 2015

Table of Contents

Abstract

Grapevines contain many different organic acids and the two most abundant are tartaric acid and malic acid. Malic acid and tartaric acid both increase in concentration up until veraison then after veraison malic acid is broken down as sugar increases. Malic acid has been studied in a variety of fruits for it is a very common acid. However tartaric acid is an uncommon primary acid in fruits and very little is known about its synthesis in grapevine. However, tartaric acid is important in providing a low pH which is important for the prevention of microbial spoilage during the winemaking process. A high pH of juice means that more tartaric acid will need to be added in the winery increasing the cost to wine makers. By discovering more about the genes involved in the synthesis of malic and tartaric acid and the breakdown of malic acid this knowledge could be used to breed vines with higher acid concentrations.

L-Idonate dehydrogenase (L-IDH) is one of only two genes known to participate in the tartaric acid synthesis pathway. Since its initial characterisation two more isoforms have been annotated in the grapevine genome based on sequence similarity. The characterisation of these isoforms was undertaken using a variety of techniques including expression of the proteins in *E. coli* and *in vitro* protein activity assays and also *in planta* expression in the microvine with the creation of transgenic microvines.

To try and discover regions of the genome that might be involved in acid metabolism in grapevine berries, malic and tartaric acid concentrations were measured from four progeny populations. The individuals of these populations were then sequenced using a genotyping

by sequencing method to find SNPs markers for a Genome Wide Association Study (GWAS). This GWAS was then verified with genetic mapping and QTL analysis.

During the process of measuring acid from these progeny populations the question of variability in acid concentration between berries from the same vine arose. A preliminary study into this variability was conducted to determine the variability of malic and tartaric acid in berries both within a bunch and between bunches from the same vine. This data was then used to predict the error in sampling subsets of berries of different sizes.

Tartaric acid concentration in tissues other than the berry was also explored. Acid concentration was measured in several tissues including root, shoots and leaves. It was found that tartaric acid was present in these tissues with varying concentrations. Tartaric acid concentration in leaves was then studied further try see if there was a link between the age of the leaf and tartaric acid concentration and also between leaf tartaric acid concentration and berry tartaric acid concentration. There was found to be no link between the two in these preliminary studies.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for an other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying subject to the provisions of the Copyright Act 1968.

The Author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has be granted by the University to restrict access for a period of time.

Acknowledgements

I would like to thank the members, past and present, of the Thomas Laboratory and the Ford Laboratory, especially, Dr Mahmood Hassan, Pat Corena, Don McKenzie, Dr. Crystal Sweetman and Yong Jia and also everyone at CSIRO agriculture Flagship at Waite Campus for their help.

I would like to thank Dr. Natoyia Lloyd from the Australian Wine Research Institute for conducting the UHPLC-MS/MS to measure acid concentration in berry samples and Dr. Olena Kravchuk for providing statistical analysis to determine the error of sampling subsets of berries.

I would like to thank Angela Lush from Lush Logic for help with spelling and editing of the thesis manuscript.

I would like to thank the Australian Grape and Wine Authority for providing me with a top-up and laboratory materials scholarship, the University of Adelaide for providing me with the C.J Everard Scholarship and CSIRO Agriculture Flagship for providing the space and materials so I could conduct my research.

I would like to thank my family and my partner Eric for providing me with support during my PhD.

Finally I would like to thank my Supervisors Dr Mark Thomas and Associate Professor Christopher Ford for their guidance and help with my research and thesis writing.

List of Abbreviations

AGRF	Australian Genome Research Facility
AWRI	Australian Wine Research Institute
EST	Expressed Sequence Tag
FPLC	Fast Protein Liquid Chromatography
GA	Gibberellins
GBS	Genotyping By Sequencing
GLM	General Linear Model
GUI	Graphical User Interface
GWAS	Genome Wide Association Study
HPLC	High Performance Liquid Chromatography
ICP	Inductively Coupled Plasma Mass Spectrometry
IPTG	Isopropyl β-D-1-thiogalactopyranoside
5KGA	5-keto-D-gluconic acid
L-IDH	L-idonate Dehydrogenase
LOD	Log Of Odds
MLM	Mixed Linear Models
MS	Malate Synthase
NAD-cyMDH	cytoplasmic NAD dependent malate dehydrogenase
NAD-mMDH	Mitochondrial NAD Dependent Malate Dehydrogenase
NADP-ME	NADP dependent malic enzyme
NCBI	The National Center for Biotechnology Information
NGS	Next Generation Sequencing
NMR	Nuclear Magnetic Resonance
PCA	Principal Component Analysis
PEP	Phosphoenolpyruvate
PEPC	Phosphoenolpyruvate Carboxylase
PEPCK	Phosphoenolpyruvate Carboxykinase
QTL	Quantitative Trait Loci
qRT-PCR	Quantitative Reverse Transcription PCR
RT-PCR	Reverse Transcription PCR
SARDI	South Australian Research and Development Institute
SDS-PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
SE	Standard Error
SNP	Single Nucleotide Polymorphism

TCA	Tricarboxylic Acid
UHPLC-MS/MS	Ultra-High Performance Liquid Chromatography Mass Spectrometry

Chapter 1

Literature Review

1.1 The importance of the grapevine as a crop

Domestication of the grapevine and the first instances of wine production are thought to have occurred in the Neolithic period (8500-4000 BC) in Transcaucasia and Anatolia (Arroyo-Garcia et al. 2006, Vouillamoz et al. 2006). The consumption of grapes had been documented at pre-historical sites across Europe (Arroyo-Garcia et al. 2006). Today, grapevines are grown all over the world including in Australia. In the 2012-2013 financial year, 154,030 ha of land was devoted to the growing of grapevines in Australia (Australian Bureau of Statistics 2015). 1.23 billion litres of wine was produced and sold both domestically and internationally equating to 2.4 billion dollars of sales (Australian Bureau of Statistics 2015).

1.2 Lifecycle and development of the grapevine

The grapevine is a woody, deciduous, perennial plant that flowers and fruits once a year in temperate climates. Berry development can be divided into two growth phases with a lag phase in between (Figure 1.1). Throughout the first growth phase, berries are hard, green and bitter tasting, due to the accumulation of organic acids, and have very little sugar content (Coombe 1960, Famiani et al. 2000). As well as the accumulation of acids, the first growth phase involves growth of the berry through cell division and then cell enlargement up until veraison (Coombe 1960). Ojeda *et al.* (1999) measured DNA content in pericarp (the skin and flesh of the berry) cells as a measure of cell division. They noted that DNA content increased from anthesis until 35 days post anthesis. After this time the growth of the berry is reliant on the enlargement of pre-existing cells.

Between the first and second growth phases is veraison, when berries begin to soften and red berries gain their colour. After veraison is the second growth stage, sometimes termed

ripening (Figure 1.1). It is in this stage that the berries accumulate sugars and lose some acidity, due to malic acid breakdown (Coombe 1960). During this phase the rate of growth of the berry increases again. There is no cell division at this stage and the berry grows through the enlargement of cells within the pericarp (Coombe 1960).

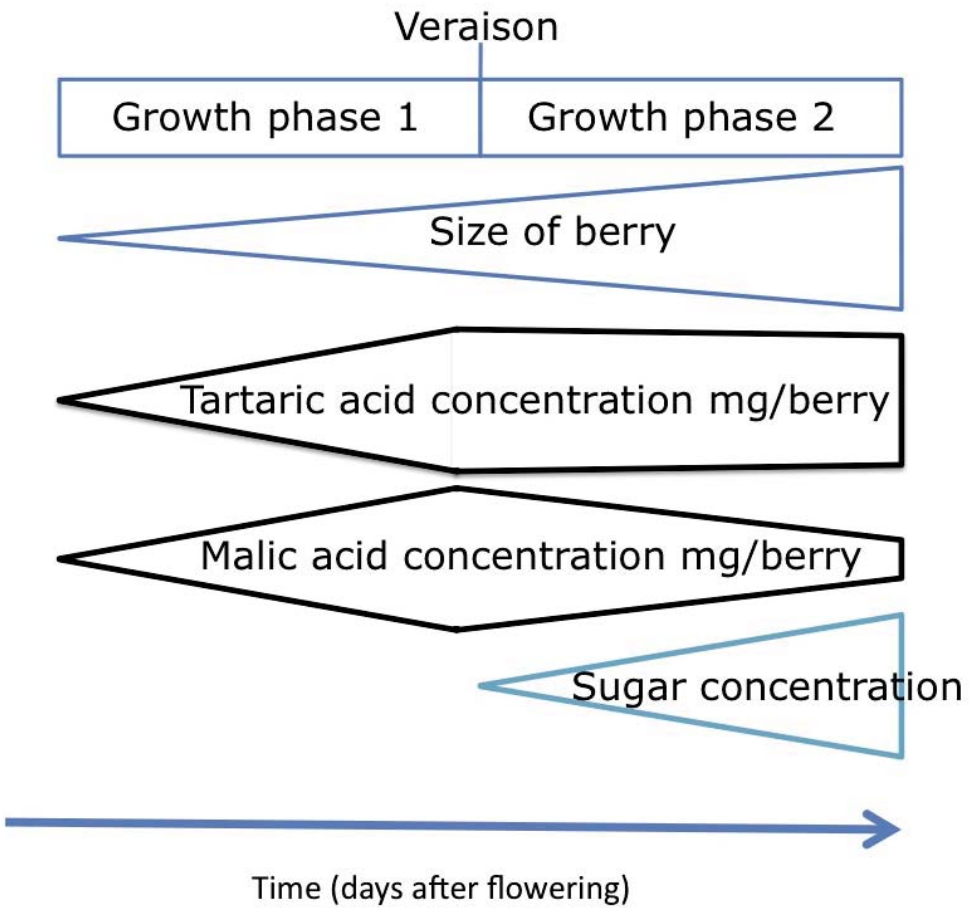


Figure 1.1: Berry development is divided into two phases separated by veraison. The size of the berry increases during both growth phases. During the first growth phase tartaric and malic acids accumulate. During the second growth phase, sugars begin to accumulate and malic acid is metabolised so the concentration of malic acid decreases. The content of tartaric acid stays relatively constant throughout the second growth phase.

1.3 Advances in grapevine molecular genetics

In the past ten years there have been several developments that have advanced the field of grapevine molecular genetics, including the release of the draft sequence of the grapevine genome (Jaillon et al. 2007) and the development of the microvine (Chaib *et al*, 2010). These advances have lessened the difficulty of performing genetic studies on grapevines such as transformation studies and genetic mapping, both with Quantitative Trait Loci (QTL) or Genome Wide Association Study (GWAS).

The genome of *V. vinifera* was sequenced in 2007 (Jaillon et al. 2007). The genome sequence data was constructed from a near-homozygous genotype known as PN40024. This genotype is often mistakenly referred to as an inbred line of Pinot Noir but DNA marker analysis indicated the original parentage might have been Helfensteiner or a cross between Pinot Noir and Helfensteiner. This genotype was chosen for sequencing as unpublished initial attempts to assemble genome sequence from standard grapevine varieties proved unsuccessful due to the highly heterozygous nature of the grapevine genome of these standard vines (M. Thomas, personal communication). The study of this genome found that the gene families encoding stilbene synthases and terpene synthases, both of which contain genes which produce aroma compounds, are greatly increased in number compared to *Arabidopsis* and rice. (Jaillon et al. 2007). An unassembled heterozygous genome of Pinot Noir was also published and provided one of the first insights into the amount of sequence polymorphism in the grapevine genome with the homologous chromosomes of this genome differing by 11.2 % (Velasco et al. 2007, Martinez-Zapater et al. 2010). Sequencing of the whole genome has provided a basis for genetic mapping, genetic identification and genetic diversity studies (Martinez-Zapater et al. 2010). Primers and probes can be designed anywhere in the genome to look at genetic

diversity of a particular gene in a variety of cultivars and gene specific markers can be developed.

Chaib *et al.*(2010) developed microvine lines for rapid grapevine genetic studies. A microvine is a vine of dwarf stature that flowers and fruits rapidly along the whole cane (Figure 1.2). This plant exploits a mutation in the *GA insensitive* gene discovered by Boss and Thomas (2002) that confers dwarf stature and continuous flowering and fruiting along the whole cane (Figure 1.2b). The microvine can be transformed with *Agrobacterium tumefaciens* allowing it to be used in transgenic studies (Chaib et al. 2010). The development of the microvine allows researchers to create genetic populations and transgenic vines that take up less space and can flower and fruit faster than wild type grapevines, greatly reducing the time taken and space needed to perform genetic studies or gene knockout or over-expression studies. However, despite these advantages there are still no published studies using transgenic microvines. Only a handful of genetic or gene expression studies have used microvines (Dunlevy et al. 2013, Fernandez et al. 2013, Rienth, Torregrosa, Kelly, et al. 2014, Rienth, Torregrosa, Luchaire, et al. 2014).

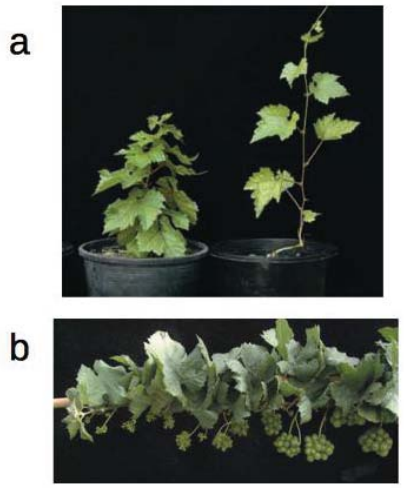


Figure 1.2: The microvine a) Picture of the microvine (left) and wild type vine (right) showing the differences in size at the young seedling stage. From: Chaib et al. (2010). b) Picture of grapevine with GA insensitive mutation showing that this mutation allows for the fruiting along the whole cane. From: Boss and Thomas (2002).

1.4 Organic acids in grape berries

Organic acids that accumulate before veraison are extremely important for the taste and organoleptic properties of the berries and wine and for preventing microbial spoilage of the juice and wine (Liu et al. 2006, Ford 2012). Kliewer (1966) found 23 organic acids in grapevine berries but most are found in minuscule concentrations. The most important and most studied are tartaric and malic acids attributing to 90% of the total acidity in the berry (Ruffner 1982).Tartaric acid has no known role in metabolism within the berry; however, malic acid is known to have several roles in carbon and sugar metabolism both within the grape berry and other fruits.

Both tartaric and malic acids accumulate in the vacuoles of pericarp cells within the berry (Fontes et al. 2011). The transporter VvALMT9 has been found to be involved in the transport of tartaric and malic acids into the vacuole (De Angeli et al. 2013). The accumulation of both tartaric acid and malic acid and the breakdown of malic acid have

been studied extensively over the whole growing season (Coombe & Jones 1983).

1.4.1 Tartaric acid

Tartaric acid is the primary acid in grapevine berries. Tartaric acid accumulates in berries before veraison and the content remains relatively stable until harvest (Figure 1.1). There is a slight drop in concentration post-veraison that is attributed to the increase in the size of the berry (Iland & Coombe 1988).

Tartaric acid accumulation to high concentrations is unusual among plants (Loewus 1999, Ford 2012). Grapevines are the only commercial fruit crop to have tartaric acid as the primary acid; most fruits accumulate malic and citric acids (Loewus & Stafford 1958, DeBolt et al. 2007). The leaves of other plants that produce tartaric acid in relatively high levels include members of the *Parthenocissus* family, including Boston Ivy and Virginia Creeper, *Pelargonium* (Geraniums) and *Phaseolus vulgaris* (bush bean) (Stafford 1959).

1.4.1.1 Tartaric acid synthesis in grapevine berries

Tartaric acid is synthesised from ascorbic acid (vitamin C). When tartaric acid synthesis was first studied in grapes, L-ascorbic acid was ruled out as its precursor. Radiolabeling studies using L-ascorbic acid-6-^{14}C had shown a low recovery of the radiolabel in tartaric acid (Loewus & Stafford 1958). Later, Saito and Kasai (1969) showed that ascorbic acid was the precursor of tartaric acid, using radiolabeling studies with L-ascorbic acid-1-^{14}C. Seventy-two percent of the radiolabeled ^{14}C from L-ascorbic acid was recovered as tartaric acid. This percentage recovery was higher than that of any of the other substrates trialled, suggesting that L-ascorbic acid was the precursor for tartaric acid synthesis. Williams and Loewus (1978) showed conclusively that the four carbon fragment C_1 to C_4 of ascorbic acid

is converted to tartaric acid in grapevines by performing similar experiments with L-ascorbic acid-4-^{14}C.

While the pathway of tartaric acid synthesis in grapevine was determined by adding L-[1-^{14}C] ascorbic acid to grape berries an inhibitory compound was added (iodoacetic acid) and the location of the ^{14}C was found to be in three intermediates L-idono-γ-lactone, 2-keto-L-gluconic acid and L-idonic acid (Saito & Kasai 1982). These three intermediates were then radiolabeled to determine the order of the pathway (Saito & Kasai 1984) (Figure 1.3). From these experiments the intermediates in the tartaric acid synthesis pathway were determined but even today very little is known about the genes and proteins that function in the pathway. Only one enzyme of the pathway, L-idonate dehydrogenase (L-IDH), has been isolated and studied (DeBolt 2006). A candidate gene for the conversation of 2-keto L-gulonic acid into L-idonate (2-keto L-gulonate reductase) has been identified and studied but may be involved in a variety of other reactions (Burbidge 2011).

The L-IDH enzyme was discovered by comparing the expression of eight candidate genes for steps in the tartaric acid pathway, using RT-PCR, from *V. vinifera* and a closely related species, *Ampelopsis aconitifolia*, which does not produce tartaric acid (DeBolt 2006). Candidate 5 (which was chosen because of its homology to *E. coli* L-idonate dehygrogenase) was produced in *V. vinifera* but not in *A. aconitifolia* (DeBolt 2006). The transcript of this gene was shown to accumulate in *V. vinifera* berries in a similar pattern to tartaric acid and the *V. vinifera* protein was shown to have a similar function to the bacterial protein in in-vitro studies, so the protein was identified as L-idonate dehydrogenase (DeBolt et al. 2006). Proteomic studies have shown that the amount of L-IDH decreases as the berry grows and the rate of this decrease increases as the berry

develops (Martinez-Esteso et al. 2011). Transcriptomic studies also show that the L-IDH mRNA decrease corresponds to the decrease in tartaric acid synthesis over development (Deluc et al. 2007, Sweetman et al. 2012). Wen *et al.* (2010, 2014) created an antibody to L-IDH and showed that the expression of the L-IDH transcript and protein over berry development corresponded to tartaric acid accumulation in several *V. vinifera* cultivars and hybrids.

1.4.2 Malic acid

Malic acid, like tartaric acid, increases in concentration during the first growth phase, up until six weeks after anthesis, but then decreases sharply after veraison (Ruffner & Hawker 1977). Malic acid takes part in a number of important metabolic processes in the berry, mainly during the second growth period, that lead to the decrease in concentration of malic acid (Sweetman et al. 2009, Ford 2012). The pathways of malate synthesis and breakdown are common to all plants and have been characterised. Sweetman *et al.* (2009) produced a comprehensive review of both malate synthesis and breakdown in fruits focusing on grapes.

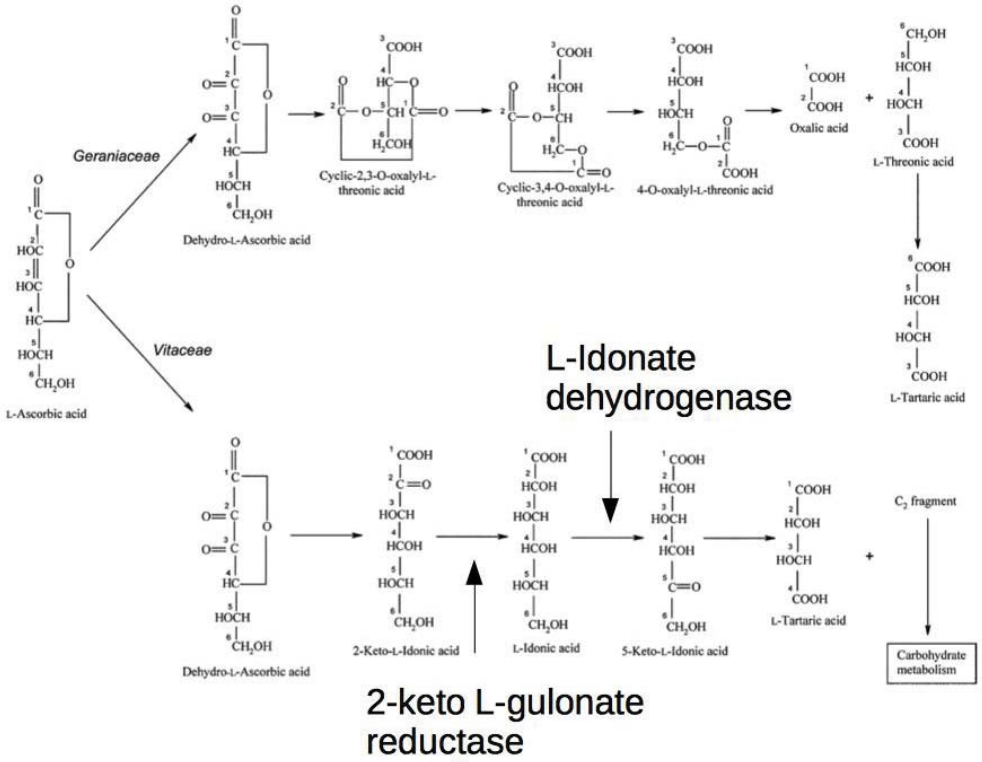


Figure 1.3: Tartaric acid synthesis in plants. Top pathway occurs in Geraniaceae (Geranium) and bottom pathway occurs in Vitaceae. The pathways begin with the same precursor, ascorbic acid, but have slightly different intermediates and end products. In Geraniums, as well as tartaric acid, oxalic acid is also produced from ascorbate however, in Vitaceae, tartaric acid is produced with a 2 carbon fragment that is hypothesised to be glycoaldehyde (DeBolt et al. 2006). L-idonate dehydrogenase is the only enzyme that has been definitively characterised from the pathway in Vitaceae. 2-keto L- gulonate is thought to participate in the pathway but may participate in other pathways as well (Burbidge 2011). To date no enzymes have been characterised from the pathway in Geraniaceae. Adapted from: Hancock and Viola (2005).

1.4.2.1 Malic acid synthesis in the berry

There are a number of pathways for the formation of malic acid, via glycolysis and carbon fixation or from the tricarboxylic acid (TCA) or glyoxylate cycles. Glycolysis is a pathway in which sugars are broken down to form ATP and pyruvate, which can enter the TCA cycle (Figure 1.4). The penultimate step in glycolysis creates phosphoenolpyruvate (PEP), which is then converted to pyruvate. PEP can also be converted to malate via oxaloacetate (Figure 1.4). This process also fixes CO_2; CO_2 is converted into the bicarbonate anion and reacts with PEP to form oxaloacetate. The enzyme responsible for the β-carboxylase reaction between PEP and oxaloacetate is PEP carboxylase (PEPC). Oxaloacetate is then converted to malate via cytoplasmic NAD-malate dehydrogenase (Figure 1.4). Despite grapevines being a C3 plant, there is evidence that grapevine berries participate in a C4 like CO_2 fixation and malate may be produced from this pathway (Figure 1.4) (Sweetman *et. al.* 2009).

Malate is an intermediate of the TCA cycle and is formed from fumarate via fumarase and is then converted to oxaloacetate via mitochondrial NAD dependent malate dehydrogenase (NAD-mMDH) (Figure 1.4). Malate is also an intermediate of the glyoxylate cycle, which produces succinate (which can be converted into sugars). Malate is formed in this cycle by the action of malate synthase (MS) (Figure 1.4). Malate may be sequestered from both the TCA and the glyoxylate cycle for storage in the vacuole (Figure 1.4).

Studies using microarrays, RNAseq, proteomics or immunohistochemistry have shown that the enzymes named above are expressed at the correct times during the ripening process when malic acid is synthesised (Famiani et al. 2000, Pilati et al. 2007, Martinez-Esteso et al. 2011, Sweetman et al. 2012, Agudelo-Romero et al. 2013).

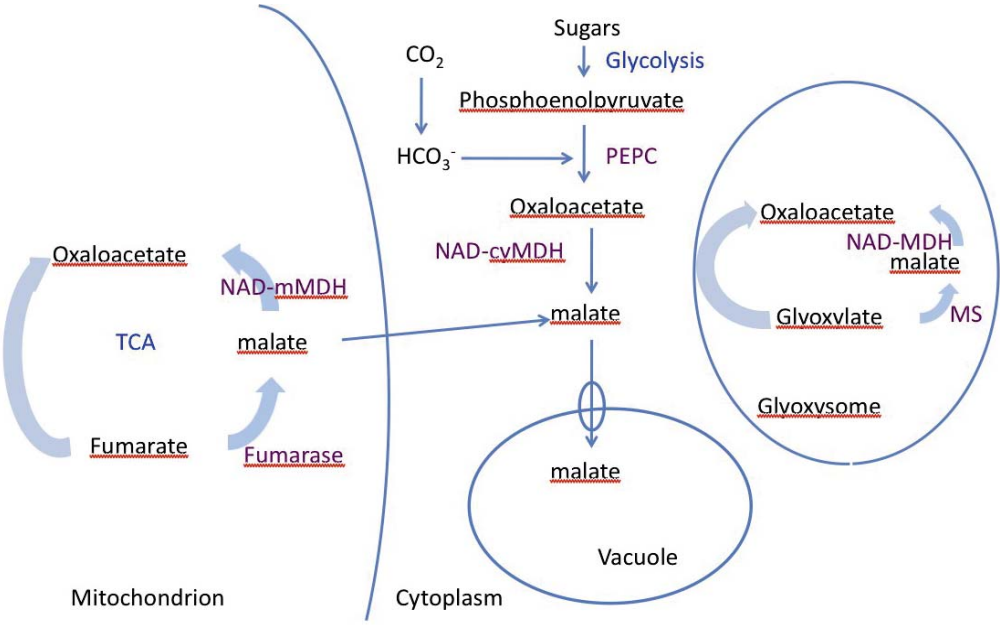


Figure 1.4: Malate synthesis in the grape berry. Phosphoenolpyruvate is created from glycolysis that can be converted into pyruvate. However it can also be converted to oxaloacetate with the action of the enzyme phosophoenolpyruvate carboxylase. Oxaloacetate is then converted to malate via cytoplasmic malate dehydrogenase (NAD-cyMDH). Malate is created in the tricarboxylic acid cycle in the mitochondria and then sequestered from this cycle for storage in the vacuole. Malate can also be created and sequestered in a similar way from the glyoxysome. Metabolic pathways are in blue, enzymes are in purple and substrates are in black. Adapted from Sweetman et al. (2009).

1.4.2.2 Malic acid breakdown in the berry

At veraison, sugars stop being used as the major carbon source for berry growth and begin to be stored within the berry. At this time, it is thought that malate becomes an important carbon source for the berry (Ruffner & Hawker 1977). Malate is exported out of the vacuole and can be imported into the mitochondria to be fed back into the TCA to maintain carbon levels during this cycle (Figure 1.5). Sugars are produced through gluconeogenesis, which is essentially a reversal of glycolysis in which pyruvate becomes glucose. Malate can be reconverted into PEP though oxaloacetate (Figure 1.5). NAD-MDH can covert malate back to oxaloacetate and then oxaloacetate is converted to PEP with the action of the enzyme PEP carboxykinase (PEPCK) releasing CO_2. Malate can be decarboxylated via NADP dependent malic enzyme into pyruvate, also releasing CO_2 (Figure 1.5). The same

studies that published expression or proteomic data for the enzymes involved in malic acid synthesis also show data on genes and proteins involved in malic acid breakdown. These genes and proteins are expressed at the correct stage of berry development corresponding to malic acid breakdown (Famiani et al. 2000, Pilati et al. 2007, Martinez-Esteso et al. 2011, Sweetman et al. 2012, Agudelo-Romero et al. 2013).

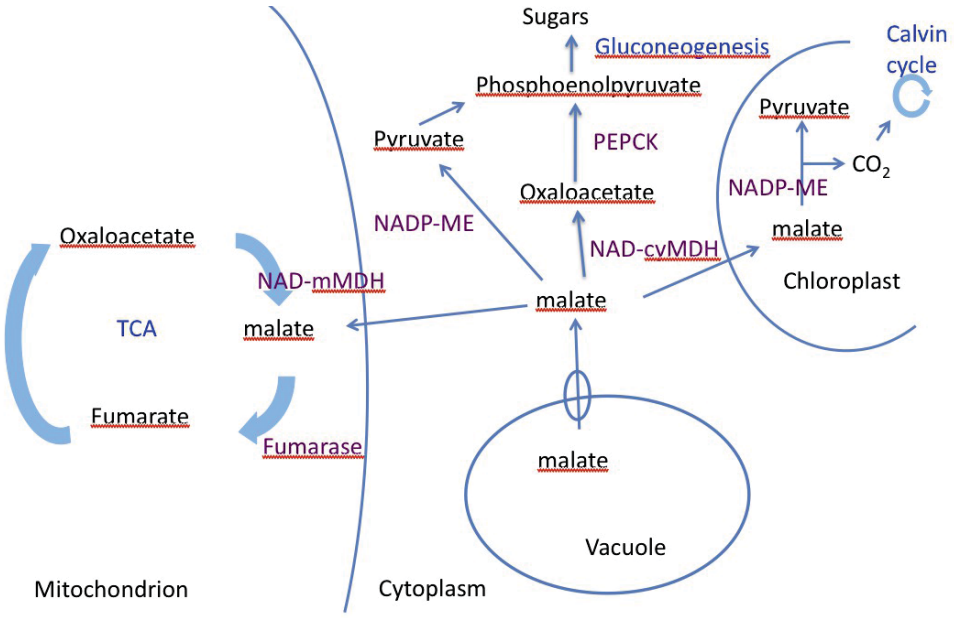


Figure 1.5: Malic acid breakdown in the grape berry. Malate can be converted back into oxaloacetate via cytoplasmic NAD dependent malate dehydrogenase (NAD-cyMDH). Oxaloacetate is then converted into phosphoenolpyruvate via the enzyme phophoenolpyruvate carboxykinase and PEPC can enter gluconeogenesis to create sugars. Malate can also be converted directly to pyruvate via NADP dependent malic enzyme (NADP-ME) and pyruvate can become phosphoenolpyruvate. Malate can also enter the tricarboxcylic acid cycle in the mitochondria to maintain carbon levels during the cycle. In the chloroplast, malate is converted into pyruvate, which releases CO_2 for photosynthesis. Metabolic pathways are in blue, enzymes are in purple and substrates are in black. Adapted from Sweetman et al. (2009).

1.5 Variation in acid concentration between cultivars and within populations

There have been several studies published which document the differences in acid concentration between cultivars of *V. vinifera,* and its hybrids, which are outlined in Table 1.1. Pavlousek and Kumsta (2011) measured the acid concentrations in eleven hybrids as well as four *V. vinifera* varieties. They list the total variation in malic and tartaric acid as well as the average variation in just the one cultivar (Table 1.1). Sabir *et al.* (2010) studied five cultivars over four stages of development and Munoz-Robredo *et al.* (2011) sampled three table grape cultivars at six time points from just before veraison to harvest (Table 1.1). While both of these studies provide a clear picture of acid concentrations after veraison, they did not incorporate time points from early development and in both cases the first time point was just before veraison. Melino *et al.* (2009) published malic and tartaric acid content over four time points including early development. DeBolt (2006) and Kliewer *et al.* (1967) published the tartaric and malic acid concentrations for a range of species from the Vitaceae family (Table 1.1). Liu *et al*. (2006) sampled 98 cultivars of both hybrid and *V. vinifera* wine and table grapes over two years (Table 1.1). Malic acid levels in berries lay between 0.38 mg/ml to 7.06 mg/ml and tartaric acid lay between 1.54 mg/ml and 9.05 mg/ml (Liu et al. 2006). Preiner *et al.* (2013) measured the tartaric and malic acid concentrations in seven native Croatian *V. vinifera* cultivars over three years and Soyer *et al.* (2003) measured the concentrations in eleven cultivars grown in Turkey (Table 1.1). Both of these studies found a wide range of acid concentrations between the cultivars.

Table 1.1: Published studies detailing the concentrations of malic and tartaric acids in berries from different grapevine species, from V. vinifera cultivars and different developmental time points.

Study	No. of cultivars	Malic acid concentration	Tartaric acid concentration	Time-points sampled
Pavlousek and Kumsta (2011)	4 *V. vinifera* 11 hybrids	All 2.68 g/L to 3.72 g/L *V. vinifera* 3.32 g/L to 5.04 g/L	All 6.85 g/L to 8.25 g/L *V. vinifera* 5.73 g/L to 9.05 g/L	N/A
Sabir *et al.* (2010)	5 *V. vinifera*	Unripe: 9.1 g/L to 15.1 g/L Veraison: 8.9 g/L to 13.8 g/L Early ripening: 4.3 g/L to 5.2 g/L Late ripening: 2.8 g/L to 3.6 g/L	Unripe: 10.3 g/L to 12.3 g/L Veraison: 7 g/L to 11 g/L Early Ripening: 5.7 g/L to 6.4 g/L Late Ripening: 3.8 g/L to 5.2 g/L	4 time points
Liu *et al.* (2006)	98 cultivars of both *V. vinifera* and hybrids	0.38 mg/ml to 7.06 mg/ml	1.54 mg/ml and 9.05 mg/ml	1 time point
Munoz-Robredo *et al.* (2011)	3 *V. vinifera* cultivars (table grapes)	0.38 g/L to 29.92 g/L	1.28 g/L to 7.45 g/L	6 time points (Close to harvest)
DeBolt (2006)	Several species in the Vitaceae family	An averaged value given for each species	An averaged value given for each species	Pre-veraison and post veraison time point
Kliewer *et al.* (1967)	Over 20 different cultivars	0.19 to 0.68 100g malic /100ml	0.52 to 1.08 100g tartaric /100ml	Early and late stages of maturity
Preiner *et al.* (2013)	Three years and seven cultivars	0.31 to 3.40 g/L	4.49 to 7.19 g/L	One time point
Soyer *et al.* (2003)	11 cultivars	1.43 to 3.40 g/L	6.61 to 10.12 g/L	One time point
(Melino et al. 2009)	2 cultivars	0.84 to 78.71 µmol/berry	2.12 to 55.4 µmol/berry	4 time points

1.6 Environmental factors that affect variation in acid concentration in berries and wines

There are many factors, in addition to physiological and biochemical features of berry metabolism that can lead to variation in acid concentration, such as cultivar, climate and soil. Van Leeuwen *et al.* (2004) studied the impact that cultivar or soil had on various parameters of grape quality, including the concentrations of tartaric and malic acids. Tartaric acid did not seem to vary significantly with different soil, climate or cultivar. However, malic acid did vary with climate and cultivar. Van Leeuwen et al (2004) used berries at the final harvest date for their analysis. This single point measurement does not capture acid synthesis during the pre-veraison period, malic acid breakdown during the ripening phase and the effect of climate and genotype on these different processes. A further study of other time points would provide this information.

1.6.1 Temperature and sun exposure

Temperature is known to affect malic acid concentration in berries with higher temperatures increasing the rate of malic acid breakdown. This is seen the most clearly post-veraison when malic acid concentrations are already decreasing (Kliewer 1968, Sadras et al. 2013, Sweetman et al. 2014). Malic enzyme is more heat stable than PEPC and this could be why the rate of degradation of malic acid is greater when temperatures are higher during the second growth phase (Lakso and Kliewer, 1978).

Temperature may also differentially affect individual berries from the same vine as berries that receive direct sunlight are likely to be hotter than berries that receive diffuse sunlight. Bergqvist *et al.* (2001) found that there was a 3-4 ºC difference in the temperature of berries from the north and south side of vines grown in the San Joaquin Valley in

California (a warm, sunny climate like Australia). Malic acid concentration has been shown to be lower in berries exposed to more direct sunlight than shaded berries (Kliewer & Lider 1968, Crippen & Morrison 1986).

While it does not appear that temperature affects tartaric acid concentration, there are mixed results when looking at sun exposure. Some studies have shown that tartaric acid concentrations in mg/berry do not differ between sun exposed or shaded berries (Kliewer & Lider 1968, Crippen & Morrison 1986). However if all sunlight is excluded, with the help of a box designed to block out sunlight without increasing the temperature of the berries, there is a significant difference between exposed and moderately shaded berries (DeBolt et al. 2008). In the studies by Kliewer and Lider (1968) and Crippen and Morrison (1986) sunlight was not completely excluded from the berries and the shaded berries of their experiments would roughly align with the moderately shaded berries of DeBolt (2008) indicating that completely excluding light from berries has an affect on tartaric acid accumulation.

1.7 Genetic variation in grapevine berries

Since the synthesis of both malic and tartaric are biological processes, there must be some genetic element that contributes to variation in acid concentration, but how much this contributes to overall acid concentration of the berries is unknown. Duchene *et al.* (2014) found significant genetic variability in tartaric and malic acids in a cross of Riesling and Gewurztraminer. This genetic variation could be used for genetic studies such as GWAS and QTL analysis. They also evaluated the acid concentrations of the individuals at the heat sum after veraison. This is a novel approach to comparing concentration between individuals with different veraison dates such as progeny populations.

1.7.1 The importance of genetic variation in the study of acid metabolism in grapevines

Acid concentration and pH of juice is linked to wine quality and in Australia tartaric acid is often added to the juice during winemaking to lower pH. The addition of tartaric acid to juice and wines is one of the biggest costs during the winemaking process. Malic acid breakdown is greater in warmer climates, leading to lower acid concentration at harvest. As temperatures in Australian wine regions are expected to rise due to climate change (Hall & Jones 2009) the pH of juice is expected to be higher at harvest (de Orduna 2010) and more acid will need to be added during the winemaking process to compensate for the lower acid concentrations, raising the cost to wineries.

As there is a genetic component to the variation in acid concentrations this can be exploited for genetic studies to find genes involved in acid metabolism. While a lot is known about the genes involved in malic acid accumulation and breakdown, almost nothing is known about their contribution to variation in acid concentration in grape berries. As for tartaric acid, almost nothing is known about the genes involved in its synthesis in the grapevine. Information discovered about the genes that affect acid concentration in grapevines can be used to produce new wine grape varieties with higher acid concentration at harvest. To do this the genes and gene placement within the genome must be known so that they can be sequenced and characterised in other cultivars, identifying variation that will aid the breeding process.

Genetic variation within the genome, especially Single Nucleotide Polymorphisms (SNPs) can be used as markers for GWAS or to construct a genetic map and perform QTL analysis to discover genes involved in tartaric and malic acid metabolism. Both these analyses detect the co-segregation of genetic markers with genes that are involved in the trait of

interest (Oraguzie & Wilcox 2007). Depending on the number of markers and the nature of the cross, the QTL may contain many markers found to associate with the trait or the two markers might be far apart, creating a large area that contains many genes which may associate with the trait or may be in linkage disequilibrium with the marker or the gene involved in the trait. The QTL area may need to be narrowed down with the use of further crosses, individuals or markers. Or for species which have a published genome sequence, all the genes contained within the area can be assessed for their likelihood for participating in acid metabolism. An example of this method can be seen in Ban *et al.* (2014) who used the published grapevine genome to look for genes that may be controlling anthocyanin content in grapevines the QTLs that they had discovered using a cross of *Vitis labruscana* and *Vitis vinifera*.

SNPs are excellent markers for GWAS and QTL analysis due to their placement across the whole genome and their abundance within the genome. A Genotyping by Sequencing (GBS) method to discover SNP markers has been developed for grapevine using next generation sequencing methods (Hyma et al. 2015). SNPs discovered using this method were used to create a genetic map and for QTL analysis and GWAS to find areas of the genome involved in powdery mildew resistance and susceptibility (Barba et al. 2014). SNP markers discovered with this method in new populations with GWAS and QTL analysis could be used to discover chromosome regions and genes that are important for determining the overall concentration of malic and tartaric acid within the berry.

1.8 Summary

While the pathway for tartaric acid synthesis in grapevine is known, only two genes have been confirmed to have some role in this pathway (DeBolt et al. 2006, Burbidge 2011). L-IDH is one of these genes. Since the initial characterisation of this gene two more isoforms of L-IDH have been annotated in the grapevine genome based on sequence similarity. While the presence of these isoforms is known, only one isoform has been characterised. Work described in this thesis attempts to understand more about the function of the isoforms of L-IDH in the grapevine. This was done by protein expression and activity assays, gene sequencing. The work detailed here also builds on the original characterisation work by DeBolt *et al.* (2006) and takes the characterisation further with the over-expression of the three isoforms in the microvine. This is the first published account of the transformation of microvines other than the original proof-of-concept publication by Chaib *et al.* (2010).

Grapevines appear to make tartaric acid in a variety of tissues but only a few studies have ever investigated tartaric acid content in other tissues of the grapevine (Stafford 1959, Kliewer 1966). In this work, the tartaric acid content and L-IDH expression was measured in a variety of grapevine tissues. While the expression of L-IDH has been measured before it is thought that L-IDH 1 and L-IDH 3 are always measured together due to the high sequence similarity especially at the 3' end of the cDNA (Wen et al. 2010, 2014, Sweetman et al. 2012). In this work a pyrosequencing approach was chosen as it could distinguish between L-IDH 1 and L-IDH 3 and this is the first time that the expression of these two isoforms have been measured separately.

We know that berry acid concentrations vary between different cultivars and this has been

well documented (Kliewer et al. 1967, Soyer et al. 2003, Liu et al. 2006, Sabir et al. 2010, Munoz-Robredo et al. 2011, Pavlousek & Kumsta 2011, Preiner et al. 2013) but the variation between individual berries from the same vine has not been so well documented. To measure this variation an Ultra-High Performance Liquid Chromatography Mass Spectrometry (UHPLC-MS/MS) method developed by the Australian Wine Research Institute (AWRI) especially for this study to accurately measure acid amounts individual berries and samples of multiple berries (Higginson et al. 2015). While the use of UHPLC-MS/MS is not a new technique this method was developed to be high-throughput taking five minuets per sample and is faster than other published methods. Of particular interest is how this variation in acid concentration in individual berries might affect the estimation of the overall mean of acid concentration when collecting subsets of berries from a vine. Using data collected in this work estimation of standard error of total acid concentration was estimated for different samples sizes. Which was of great help to the sampling studies conducted in this work.

Another goal of this work was to search for other genes that might be involved in tartaric and malic acid synthesis in grapevines. This was done using QTL analysis and GWAS on four grapevine progeny populations grown in the field. QTL studies have been performed before in grapevine including studies for seedlessness (Doligez et al. 2002) and bunch architecture (Correa et al. 2014). However there has only been one other study that combines GWAS with QTL analysis (Barba et al. 2014). It is through that the combination of these two techniques might help to negate some of the disadvantages of that the two techniques each have (Korte & Farlow 2013).

It is hoped that the genes involved in acid metabolism might be used in selected breeding

to breed vines with higher acid levels. Vines that produce higher concentrations of acid may be able to reduce the amount of tartaric acid that is added to wines in Australia and reduce costs for winemakers. However the amount known about the genes involved in tartaric acid synthesis especially is very small. This work hopes to increase the knowledge of the genes in the metabolism of malic and tartaric acids in grapevines.

Chapter 2

Characterisation of the isoforms of L-idonate dehydrogenase: a gene involved in tartaric acid synthesis

2.1 Introduction

While malic acid synthesis has been widely researched due to its role in central metabolism, there has been little research into the synthesis of tartaric acid. This is due to the rarity of tartaric acid as a major acid in fruits and the fact that it does not appear to be useful to the plant. Experiments conducted in the past using bunch feeding of radiolabelled compounds determined that tartaric acid is produced from ascorbic acid, that there are two different pathways of formation (see chapter 1 for further details and Ford (2012) for a review) and that one of these pathways appears to be unique to grapevine. From these experiments the steps of the pathways of tartaric acid synthesis were determined but little research has been conducted to discover the genes that are involved in these steps or in the regulation of the pathway. So far, only two steps of the pathway have strong candidate genes thought to be involved in tartaric acid synthesis: 2-keto L-gulonate reductase (Burbidge 2011) and L-idonate dehydrogenase (L-IDH) (DeBolt et al. 2006).

Since the discovery of L-IDH in grapevines, the grapevine genome has been sequenced (Jaillon et al. 2007, Velasco et al. 2007) and the annotation of many genes has been added to the sequence based on computer prediction. The National Center for Biotechnology Information (NCBI) has a copy of this genome sequence and all annotations, allowing for the analysis of grapevine genes. NCBI only uses the near-homozygous sequence from the clone PN40024 as the reference genome and this genome sequence was produced as part of the international Grape Genome Program. There have been two versions of the assembly, the original 8x assembly and a newer 12.0x, which is in use today.

When this study began, additional forms of L-IDH had been annotated in the draft grapevine genome in NCBI, based on sequence similarity to the L-IDH that had been

previously characterised, as well as computer prediction. It is unknown if these isoforms are translated into protein or if they have any involvement in tartaric acid synthesis. The first L-IDH to be identified was characterised by protein expression in *E.coli* and *in vitro* protein activity assays (DeBolt et al. 2006). L-idonate is rare, and a small amount was gifted to the Ford laboratory for past studies. 5-keto-D-gluconate however, is commercially available. L-idonate and 5-keto-D-gluconic acid require the use of cofactors NAD+ and NADH respectively. The absorbance of NADH can be measured at 340 nm and this provides an easy way to test the activity of both the forward and reverse reactions of the L-IDH enzymes. During the reaction with 5-keto-D-gluconic acid, NADH is converted to NAD+ and so the absorbance will decrease over time, while NAD+ will convert to NADH in the reaction with L-idonate and the absorbance should increase.

None of the L-IDH isoforms have ever been tested *in planta*. Grapevine transformations are difficult as grapevines have a long generation time, take up a lot of space and have an annual fruiting cycle. This means that the creation and characterisation of transgenic grapevines takes many years, longer than the span of a PhD. Since the discovery of the microvine (discussed in chapter 1) it has become much easier to transform grapevines due to the attributes of the microvine including dwarf stature, year round production of fruit in a growth cabinet and fruiting along the whole cane (Chaib et al. 2010). It is now possible to study the L-IDH isoforms *in planta* in the microvine.

Since the microvine is an emerging grapevine technology, only a handful of studies using microvines have been published. Most of these studies are genetic or gene expression studies (Dunlevy et al. 2013, Fernandez et al. 2013, Rienth, Torregrosa, Kelly, et al. 2014, Rienth, Torregrosa, Luchaire, et al. 2014). To date this thesis describes the only work using

transgenic microvines except for the original paper proving that they could be transformed (Chaib et al. 2010).

This chapter describes experiments in which the three isoforms of L-IDH were studied using a variety of techniques including expression in *E.coli* and *in vitro* activity assays, and *in planta* studies over-expressing the three isoforms in the microvine. The nucleic acid and amino acid sequences of the three isoforms obtained from NCBI were compared to each other. These sequence alignments show similarities between L-IDH 1 and L-IDH 3 which led to the question "are L-IDH 1 and L-IDH 3 incorrectly annotated as two separate genes and they are in fact alleles of the same gene?" It may be that one of these areas of heterozygosity in the sequenced PN40024 clone was in the region of the L-IDH isoforms on chromosome 16. To address this question, PCR amplification and sequencing of L-IDH isoforms from homozygous microvine lines (Chaib et al. 2010) were performed. Since it is unknown if the two isoforms had any involvement in tartaric acid synthesis or had any activity with L-idonate, re-expression of the previous L-IDH clone (DeBolt et al. 2006) and expression of the two new isoforms in *E. coli* were attempted to test protein activity *in vitro* with both L-idonate and 5-keto-D gluconic acid as substrates. Finally, constitutively expressed constructs containing the three L-IDH isoforms were created and transformed into the microvine to study the effects of over-expression of the isoforms on tartaric acid synthesis *in planta*.

2.2 Materials and Methods

Using NCBI to find the three isoforms of L-idonate dehydrogenase

Using the three mRNA accession numbers (L-IDH 1 XM_002269900, L-IDH 2 XM_002269859, L-IDH 3 XM_002267626) of the L-IDH isoforms from NCBI, the locus numbers of the isoforms were determined and their location found on the grapevine genome (L-IDH 1: LOC100232980, L-IDH 2: LOC100265888, L-IDH 3: LOC100257330).

Alignment of the three isoforms of L-IDH

The data listed above were downloaded from NCBI and analysed using CLC genomics workbench 7 (Qiagen). Alignments were created using a pairwise algorithm with a gap cost of 10 and a gap extension cost of 1.0 for the three mRNA sequences. These sequences were translated into amino acid sequences and aligned.

Using R to find microarray data that relates to expression of L-idonate dehydrogenase isoforms over development

Existing Affymetrix gene expression grape berry data sets (NCBI GEO accession GSE7679, GSE7680 and GSE7677) were used to identify the expression profile of L-IDH 1 (XM_002269900), L-IDH 2 (XM_002269859) and L-IDH 3(XM_002267626) using existing unpublished R scripts and an Rdata file (M.R. Thomas, unpublished).

Amplification and cloning of gDNA fragments from homozygous microvine lines

Collection of leaf material

Leaf tissue was collected from the two homozygous lines (called V3 and V10) and a heterozygous microvine L1 parent vine which was used as the control. The first few leaves

below the shoot tip were collected and stored at -20 °C. These vines were chosen because they had been specifically bred for genetic analysis and tested for homozygosity by DNA marker analysis (Chaib et al. 2010).

DNA extraction

DNA was extracted using a DNeasy DNA miniprep kit (Qiagen) following the manufacturer's instructions (DNA was eluted in 50 µl of elution buffer). DNA was visualised on a 1% (w/v) agarose gel to determine quality (see section on gel electrophoresis below). All extractions showed one clear band of DNA. DNA was also checked using a spectrophotometer at 260 nm and 280 nm. Concentrations ranges were between 30 ng/µl and 70 ng/µl.

Table 2.1: Primers used to amplify each of the L-IDH isoforms

Gene	Forward primer	Reverse primer	Size of fragment
L-IDH 1	L-IDH1 2F 5' CACCAGAGATGGGGAAAGGAGGCA 3'	L-IDH1 2R 5' ACGCTTAGAAACATAGCCTGA 3'	2.5 kbp
L-IDH 2	L-IDH2 F 5' CACCGCTGCTGTGTCTTCCGTCAC 3'	L-IDH2R 5' CCATCTAGACTGGTAAGACTGTT 3'	3.7 kbp
L-IDH 3	L-IDH3 2F 5'CACCGCTGAGACCGAGGGACAAGAGA 3'	L-IDH3 R 5' CAGCAAGCCGAAAGTCACTCCT 3'	1.7 kbp

PCR conditions

A 50 µl PCR reaction contained 1X Platinum Taq high fidelity buffer, 2.5 units Platinum Taq high fidelity DNA polymerase (Invitrogen), 0.2mM of dNTP, 2mM $MgSO_4$, 1mM of the forward and the reverse primer, between 75ng and 175 ng of DNA

template (either gDNA or a gel purified PCR product) and varying amounts of autoclaved milli-Q water. PCR mixes were assembled as a master mix with DNA and primers being added after aliquoting into 0.2 ml PCR strip tubes with domed caps (Axygen). PCR conditions are listed in table 2.2.

Table 2.2: PCR conditions for amplification of L-IDH isoforms from the homozygous lines and L1 parent

Initial	94 ºC	2 mins	
	94 ºC	30 sec	35 cycles
Annealing	58 ºC	30 sec	
Extension	68 ºC	3 min 30 sec	
Final extension	68 ºC	10 mins	

PCR machine: PCRSprint Hybaid (Thermo Scientific)

Gel electrophoresis conditions

Gels were made from 1 % (w/v) agarose gel with SYBR Safe DNA gel stain (Life Technologies) added to gel for DNA visualisation. 1X TBE was used as the running buffer. The DNA ladder GeneRuler 1 kb DNA ladder (ThermoScientific) was used for base pair length analysis. Gels were run for 30-45 minutes at 100 Volts. DNA was visualised under blue light.

Cloning process

The PCR products were purified with a QIAquick gel extraction kit (Qiagen) to the manufacturer's instructions (DNA was eluted in 30 µl of Buffer EB). The gel purified PCR products were amplified a second time as concentration was low (same conditions as above) and gel purified again with the QIAquick kit (Qiagen). 1 µg of the purified DNA fragment was A-tailed using 1 unit of Taq Polymerase (NEB) in a mix also containing 1 X Thermo Pol buffer (NEB), 1nM dATP and water in a total volume of 20 µl. This was

incubated in a PCR Sprint thermo-cycler (Thermo Scientific) for 20 minutes at 72 °C. This A-tailed mixture was cleaned up and concentrated using ethanol precipitation and re-suspended in 10 µl of autoclaved milli-Q water. Concentrated DNA fragments were then ligated to the pDriveTA cloning vector (Qiagen). Ligation mix contained 1X T4 Ligation buffer, 1 µl of concentrated T4DNA ligase (NEB), 25 ng of pDrive TA cloning vector (Qiagen), 500 ng of insert DNA and milli-Q water to make up to 20 µl. Reactions were incubated at 16 °C overnight. Ligation reactions were cleaned up and concentrated by butan-1-ol extraction and re-suspended in 5 µl of milli-Q water. 1 µl of the construct was transformed into 40 µl of *E. coli* XL1-blue cells (made in house) by electroporation. Cells were recovered by the addition of 400 µl LB and incubation at 37 °C for 1 hour with shaking at 200 rpm. They were then plated onto LB plates with 100 µg/ml ampicillin, 80 µg/ml X-gal, 13 µg/ml IPTG and 10 µg/ml tetracycline. Plates were incubated overnight at 37 °C.

Sequencing and alignments

Clones were verified using colony PCR and positive clones were sequenced using M13(-20) forward primer and M13 reverse primer (M13F 5' GTAAAACGACGGCCAGT 3', M13R 5' GGAAACAGCTATGACCATG 3'). All sequencing was performed by AGRF (Adelaide).

Alignments were created using CLC genomics workbench 7 (Qiagen) pairwise alignment algorithm with a gap cost of 10 and a gap extension cost of 1.0. Sequences were aligned with the genomic DNA sequences of L-IDH 1, 2 and 3 from NCBI (accession numbers: L-IDH 1 LOC100232980, L-IDH 2 LOC100265888 and L-IDH 3 LOC100257330).

All sequences that were obtained in this work were submitted to NCBI. NCBI accession

numbers of sequences: L-IDH 1 L1 allele 1 KT150494, L-IDH 1 L1 allele 2 KT150493, L-IDH 1 V10 KT150495, L-IDH 3 L1 allele 1 KT150496, L-IDH 3 allele 2 KT150497, L-IDH 3 V10 KT150498, L-IDH 3 V3 KT150499, L-IDH 2 L1 allele 1 KT156961, L-IDH 2 L1 allele 2 KT156962.

Cloning of L-idonate dehydrogenase isoforms into pET 52b for protein expression

Table 2.3: Primers for the amplification of L-IDH isoforms for cloning into the pET 52 b vector

Gene	Forward primer	Reverse primer	Size of fragment
L-IDH 1	L-IDH 1 pET 52b F 5' CAGGGACCCGGT GGGAAAGGAGGC AATTCT 3'	L-IDH 1-3 pET 52b R 5' GGCACCAGAGCG TTGAGATTAAACA TGACCTTGAT 3'	1.2 kbp
L-IDH 2	L-IDH 2 pET 52b F 5' CAGGGACCCGGT GGAAAAGGAGGG ATGTCT 3'	L-IDH 2 pET 52b R 5' GGCACCAGAGCG TTGAGGTTGAAC ATGACCTTAAT 3'	1.2 kbp
L-IDH 3	L-IDH 3 pET 52b F 5' CAGGGACCCGGT GGGAAAGGAGGC AACTCT 3'	L-IDH 1-3 pET 52b R 5' GGCACCAGAGCG TTGAGATTAAACA TGACCTTGAT 3'	1.2 kbp

PCR conditions

A 20 µl PCR reaction contained 1X Platinum Taq high fidelity buffer, 1 units Platinum Taq high fidelity DNA polymerase (Invitrogen), 0.2mM of dNTP, 2mM $MgSO_4$, 1mM of the forward and the reverse primer, between 75ng and 175 ng of DNA template (L-IDH-D-topo constructs made for the transformation of the L-IDH genes into the microvine) and autoclaved milli-Q water. PCR mixes were assembled as a master mix with DNA and primers being added after aliquoting into 0.2 ml PCR strip tubes with domed caps (Axygen). PCR conditions are listed in table 2.4.

Table 2.4: PCR conditions for the amplification of the L-IDH isoforms for cloning into the pET 52b vector

Initial	94 ºC	2 mins	
	94 ºC	30 sec	35 cycles
Annealing	58 ºC	30 sec	
Extension	68 ºC	3 min 30 sec	
Final extension	68 ºC	10 mins	

PCR machine: PCRSprint Hybaid (Thermo Scientific)

Cloning process

These inserts were cloned into the pET 52b (+) 3C/LIC vector (Novagen) using the manufacturer's instructions and reagents that came with the kit. The products were then transformed into giga single chemically competent cells (Novagen) for sequencing as per the manufacturer's instructions. The cells were plated onto plates containing LB with 50 µg/ml ampicillin, 80 µg/ml X-gal, 13 µg/ml IPTG and 12.5 µg/ml tetracycline. Plates were incubated overnight at 37 ºC. Clones were verified with colony PCR and positive clones were sequenced with T7 promotor primer 5' TAATACGACTCACTATAGGG 3' for beginning of sequence and the T7 Terminator primer 5' GCTAGTTATTGCTCAGCGC 3' for the end of the sequence to confirm that the vector and insert were in frame. These constructs were then used to transform BL21(DE3) pLysS or BL 21(DE3) (Novagen) chemically competent cells for protein expression in *E. coli*. They were plated on LB plates containing 50 µg/ml ampicillin for BL 21 (DE3) and 50 µg/ml ampicillin and 34 µg/ml chloramphenicol for BL 21 (DE3) pLysS cells.

Cloning of L-idonate dehydrogenase isoforms into pET 14b for protein expression

Table 2.5: Primers for the amplification of the L-IDH isoforms for cloning into the pET 14b vector

Gene	Forward primer	Reverse primer	Size of fragment
L-IDH 1	L-IDH 1 pET F 5' ATAATCATATGAG AGATGGGGAAAG GAGGCA 3’	LIDH-1 pET R 5’ CTCAAGGGATCC ACGCCTAGAAAC ATAGCCTGA 3’	1.2 kbp
L-IDH 2	L-IDH2 F: 5' ATAAATCATATGG CTGCTGTGTCTTC CGTCAC 3'	L-IDH2 R: 5' CTCAAGGATCCCC ATCTAGACTGCTA AGACTGTT 3'	1.2 kbp
L-IDH 3	L-IDH3 F: 5' ATAAATCATATGG CTGAGACCGAGG GACAAGAGG 3'	L-IDH3 R: 5' ATGGATCCGAGC AAGCCGGAACTC ACTCCT 3'	1.2 kbp

PCR conditions

A 20 µl PCR reaction contained 1X Platinum Taq high fidelity buffer, 1 units Platinum Taq high fidelity DNA polymerase (Invitrogen), 0.2 mM of dNTP, 2 mM $MgSO_4$, 1 mM of the forward and the reverse primer, between 75 ng and 175 ng of DNA template (L-IDH-D-topo constructs made for the transformation of the L-IDH genes into the microvine) and autoclaved milli-Q water. PCR mixes were assembled as a master mix with DNA and primers being added after aliquoting into 0.2 ml PCR strip tubes with domed caps (Axygen). PCR conditions are listed in table 2.6.

Table 2.6: PCR conditions for the amplification of the L-IDH isoforms for cloning into the pET 14b vector

Initial	94 °C	2 mins	
	94 °C	30 sec	35 cycles
Annealing	58 °C	30 sec	
Extension	68 °C	3 min 30 sec	
Final extension	68 °C	10 mins	

PCR machine: PCRSprint Hybaid (Thermo Scientific)

Cloning process

The PCR products were purified with a QIAquick PCR purification kit (Qiagen) to the manufacturer's instructions (DNA was eluted in 30 µl of Buffer EB). The pET14b (Novagen) vector and the purified inserts were digested in a double digest with Bam HI (NEB) and Nde I (NEB). The mix contained 1 X Buffer 3.1, 1 unit of both Nde I and Bam HI, 1 µg of purified PCR fragment or vector. The mix was incubated for 1 hour at 37 ^{o}C. As Bam HI could not be heat inactivated the reaction mixes were run on a 1 % (W/V) argarose gel and purified with a QIAquick gel extraction kit (Qiagen) to the manufacturer's instructions (DNA was eluted in 30 µl of Buffer EB). The vector was then dephosphorylated with calf alkaline phosphatase before ligation with the L-IDH isoforms by addition of the buffer and phosphatase straight to the cleaned up mix. This was incubated for 15 minutes at 37 ^{o}C and then heat inactivated at 70 ^{o}C for 5 minutes. Ligation mix contained 1X T4 ligation buffer, 2 µl of T4DNA ligase (NEB), 60 ng of pET14b vector (Novagen), 100 ng of insert DNA and milli-Q water to make up to 20 µl. Reactions were incubated at 16 ^{o}C overnight and then heat inactivated at 65 ^{o}C for 10 minutes. The products were then transformed into giga single chemically competent cells (Novagen) for sequencing as per the manufacturer's instructions. The cells were plated onto plates containing LB with 50 µg/ml ampicillin, 80 µg/ml X-gal, 13 µg/ml IPTG and 12.5 µg/ml tetracycline. Plates were incubated overnight at 37 ^{o}C. Clones were verified with colony PCR and positive clones were sequenced with T7 promotor primer 5' TAATACGACTCACTATAGGG 3' for beginning of sequence and the T7 Terminator primer 5' GCTAGTTATTGCTCAGCGC 3' for the end of the sequence to confirm that the vector and insert were in frame. These constructs were then used to transform BL21(DE3) pLysS (Novagen) chemically competent cells for protein expression in *E. coli*. They were plated on LB plates containing 50 µg/ml ampicillin

and 34 µg/ml chloramphenicol,

Expression of the L-IDH 1, 2 and 3 proteins in E. coli

Expression conditions

Five millilitre overnight cultures (in LB with appropriate antibiotics) were created from glycerol stocks (or in some cases fresh colonies). These cultures were incubated at 37 °C overnight with 200 rpm shaking. After incubation, a 1/100 dilution of these cultures were used to inoculate a culture of a larger volume (100 ml to 500 ml of LB or TB with appropriate antibiotics). This culture was grown until the OD_{600} reached 0.6 at 37 °C with shaking at 200 rpm. This culture was then induced with IPTG and incubated between 16 °C and 37 °C for between 3 and 16 hours (see tables below for the conditions of each trial) with 200 rpm shaking. These cultures were then spun down to pellet cells for protein purification. This expression method was used for almost all trials, except where a different method is stated.

Table 2.7: Trial induction conditions for expression of the L-IDH isoforms in the pET 52b vector

	Cell stain	IPTG concentration	Induction temp and length	Method of purification
Trial one	BL 21(DE3) pLysS	0.4 mM	16 °C overnight	HisTalon resin
Trial two	BL 21 (DE3) pLysS	0.4 mM	37 °C 3 hours	HisTalon resin
Trial three	BL 21 (DE3)	0.4 mM	37 °C 3 hours	HisTalon resin
Trial four*	BL 21 (DE3)	1 and 2 mM	30 °C 3 hours	HisTalon resin

* Trial four used a slightly different method to the one stated above. The overnight culture was spun down and the all the cells re-suspended in fresh LB with antibiotics before being used to inoculate 100 ml culture. This 100 ml culture was spun down once it reached an

OD_{600} 0.6. The pellet was then re-suspended in fresh LB with antibiotics before induction with IPTG. Two concentrations of IPTG were tried in separate flasks and the induction incubation temperature was lowered to 30 °C.

Table 2.8: Trial induction conditions for the expression of the L-IDH isoforms in the pET14b vector

	Cell stain	Cell culture media	IPTG concentration	Induction temp and length	Method of purification
Trial one	BL 21(DE3) pLysS	LB	0.4 mM	16 °C overnight	HisTalon kit
Trial two	BL 21(DE3) pLysS	LB	0.4 mM	30 °C 3 hours and 16 °C overnight	HisTalon kit
Trial three	BL 21(DE3) pLysS	TB	0.01 and 0.1 mM	37 °C 3 hours	HisTalon kit
Trial four re-cloned	BL 21(DE3) pLysS	LB	0.4 mM	37 °C 3 hours	HisTalon kit
Trial five*- new lab trials	BL 21(DE3) pLysS	LB	0.5 mM	16 °C overnight	FPLC method

*Trial five had different induction and purification methods as it was performed in another lab using their methods. A 20 ml overnight culture was created in LB with appropriate antibiotics. 5 ml was used in a 1/100 dilution in inoculate two 500 ml cultures which were grown till the OD_{600} reached 0.6 then inoculated with 0.5 mM IPTG and incubated at 16 °C overnight. Cultures were spun down to pellet cells and cells re-suspended in 50 ml buffer A (see below). The cells were lysed using a cell disruptor and then spun down to pellet cell debris. Supernatant was used for FPLC purification of proteins.

<u>Purification conditions</u>

Cell lysis:

Each pellet was weighed and 2ml of x-tractor buffer (Clontech) was added per 100 mg

pellet (typical weights were 500 mg to 1 g of pellet) with 1-2 µl of DNase I. This was incubated for 15 minutes on ice with mixing then centrifuged for 20 minutes 5,000 g at 4 ^{o}C. The supernatant was then used directly for protein purification.

Protein purification:

250 µl of HisTalon resin (Clontech) was equilibrated before use with a buffer containing 20 mM Tris pH 8.0, 500 mM NaCl, 10% (v/v) Glycerol, 0.05 % (v/v) Tween 20 and 3 mM imidazole. The supernatant from the cell lysis was mixed with the resin for 30 minutes on ice with shaking. Resin was washed with 5 ml of buffer containing 20 mM Tris pH 8.0, 500 mM NaCl, 10% (v/v) Glycerol, 0.05 % (v/v) Tween 20 and 10 mM imidazole, and the wash was collected in 500 µl fractions. Proteins were eluted in a buffer containing 20 mM Tris pH 8.0, 500 mM NaCl, 10% (v/v) Glycerol, 0.05 % (v/v) Tween 20 and 200 mM imidazole, and collected in 250 µl fractions.

HisTalon kit Gravity column purification kit (Clontech):

Cell lysis and column purification protocol to manufacturers instructions. The wash buffer had 10 mM imidazole added and the elution buffer had 150 mM imidazole. Elutions and final wash were collected in 500 µl fractions. Fractions were snap frozen in liquid nitrogen and stored at -80 ^{o}C.

FPLC method:

Column: mini proaffinity IMAC cartage column (Biorad) 5 ml.

FPLC: NGC chromatography system (Biorad). Buffer A: 20 mM Tris pH 8.0, 500 mM NaCl, 1 mM β-mercaptoethanol, 10 mM imidazole. Buffer B: 20 mM Tris pH 8.0, 500 mM NaCl, 1 mM β-mercaptoethanol, 250 mM imidazole. Flow rate: 2 ml/min

Table 2.9: FPLC steps for the purification of L-IDH 1 and L-IDH 3

Step	Volume	Buffer	fractions
Column equilibration	25 ml	A	
Sample loading	50 ml	none	Flow through fractions
Wash 1	35 ml	A	5-8 ml fractions
Wash 2	35 ml	97% A and 3% B	5-8 ml fractions
Protein elution	40 ml	50% A and B	5-8 ml fractions
Column wash	15 ml	A	

Fractions were stored on ice and then stored at 4 ^{o}C until protein activity assay and SDS-PAGE gel could be run. Then relevant elutions were snap frozen in liquid nitrogen and stored at -80 ^{o}C.

SDS-PAGE conditions

12% acrylamide gels (Mini-PROTEAN TGX Precast gels) were purchased from Biorad and run at 150 V. Proteins were stained with Coomassie blue stain for 10 minutes and then de-stained for 30 minutes in de-staining solution. 5 µl of Precision plus protein Western C standards (Biorad) were also run on each gel. Samples were boiled in SDS-loading buffer before loading onto the gel.

Activity assay conditions (adapted from DeBolt (2006))

Reverse reaction with 5-keto-D-gluconic acid (5KGA):

A mix of 100 mM Tris pH 8.0, 120 µM NADH was added to a 96 well plate for absorbance measurement with a spectrophotometer (total volume 300 µl) with 40 mM 5KGA and 10 µl of protein elution. The Tris-NADH mix was added first and the 5KGA was added last. The change in absorbance was measured every 10 seconds for 3 minutes at

340 nm.

Forward reaction with L-idonate:

A mix of 100 mM Tris pH 8.0, 1 mM NAD+ was added to a 96 well plate for absorbance measurement with a spectrophotometer (total volume 300 µl) with 20 mM L-idonate and 10 µl to 100 µl of protein elution. The Tris-NADH mix was added first and the L-idonate was added last. The change in absorbance was measured every 10 seconds for 3 minutes at 340 nm. All reactions were mixed at room temperature. All solutions were kept on ice.

Cloning of L-idonate dehydrogenase isoforms 1 and 2 from ESTs into D-topo vector

Sequencing to establish which ESTs contain full length L-IDH1 and L-IDH2

Table 2.10: ESTs used for amplification of L-IDH 1 and L-IDH 2

mRNA	Name (NCBI)	Stage	Contains full length cDNA
L-IDH1	CSECS079F09	Flower stage 12	Yes
	CSECS024E07	Pre veraison berries	No
	CSECS111B02	Pre veraison berries	Failed to obtain good sequence
L-IDH2	CSECS178C02	Pre veraison berries	Yes

EST plasmids were grown from glycerol stocks and the plasmids extracted using a PureLink Quick plasmid mini prep kit (Invitrogen). These were then sequenced using the following primers to establish if they contained the full length cDNA: M13F 5’ GTAAAACGACGGCCAGT 3’ M13R 5’ GGAAACAGCTATGACCATG 3’ for L-IDH1 and triplEx2 F 5’ TAATACGACTCACTATAGGGC 3’ and triplEx2R 5'CTCCGAGATCTGGACGAGC 3’ for L-IDH 2. One clone of L-IDH 1 (CSECS079F09) and the only clone of L-IDH2 (CSECS178C02) contained the full length sequence.

Amplification of L-IDH 1 and L-IDH2 from ESTs and cloning into D-topo vector

Table 2.11: Primers for the amplification of L-IDH 1 and L-IDH 2 from ESTs

Gene	Forward primer	Reverse primer	Size of fragment
L-IDH 1	L-IDH1 2F 5' CACCAGAGATGG GGAAAGGAGGCA 3'	L-IDH1 2R 5' ACGCTTAGAAAC ATAGCCTGA 3'	1.2 kbp
L-IDH 2	L-IDH2 F 5' CACCGCTGCTGT GTCTTCCGTCAC 3' L-IDH2R 5' CCATCTAGACTGG TAAGACTGTT 3'	L-IDH2R 5' CCATCTAGACTGG TAAGACTGTT 3'	1.2 kbp

PCR conditions

A 25 µl PCR reaction contained 1X Pfx 50 buffer, 1 units Pfx 50DNA polymerase (Invitrogen), 1mM of the forward and the reverse primer, between 100 ng and 150 ng of DNA template (EST constructs) and autoclaved milli-Q water. PCR mixes were assembled as a master mix with DNA and primers being added after aliquoting into 0.2 ml PCR strip tubes with domed caps (Axygen). PCR conditions are listed in table 2.12.

Table 2.12: PCR conditions for the amplification of L-IDH 1 and L-IDH 2 from ESTs

Initial	94 °C	2 mins	
	94 °C	15 sec	35 cycles
Annealing	58 °C	30 sec	
Extension	68 °C	2 mins	
Final extension	68 °C	30 mins	

PCR machine: PCRSprint Hybaid (Thermo Scientific)

The PCR products were purified with a QIAquick PCR purification kit (Qiagen) to the manufacturer's instructions (DNA was eluted in 30 µl of Buffer EB). The purified PCR fragments were cloned into the D-topo vector (Invitrogen) according to the protocol. 6 µl

of the L-IDH-D-topo construct was transformed into Top Ten chemically competent cells (Invitrogen) and plated onto LB plates containing 100 µg/ml ampicillin. The transformation was verified using colony PCR and positive colonies were sequenced using M13 primers (M13F 5' GTAAAACGACGGCCAGT 3', M13R 5' GGAAACAGCTATGACCATG 3'). NCBI accession numbers of sequences: L-IDH 1 cDNA KT150500, L-IDH 2 cDNA KT156963.

Cloning of L-idonate dehydrogenase isoforms 3 from cDNA into D-Topo vector

The final L-IDH (L-IDH3) did not have any corresponding ESTs and was amplified from cDNA.

RNA and cDNA synthesis

Pre-veraison Cabernet Sauvignon RNA created from another project was used to create cDNA to amplify L-IDH 3. cDNA was created with Superscript III reverse transcriptase (Invitrogen) following manufacturer's instructions.

Amplification of L-IDH 3 from cDNA and cloning into D-topo vector

Table 2.13: Primers for the amplification of L-IDH 3 from cDNA

Gene	Forward primer	Reverse primer	Size of fragment
L-IDH 3	L-IDH3 2F 5'CACCGCTGAGACCGAGGGACAAGAGA 3'	L-IDH3 R 5' CAGCAAGCCGAAAGTCACTCCT 3'	1.2 kbp

PCR conditions

A 25 µl PCR reaction contained 1X Pfx 50 buffer, 1 units Pfx 50DNA polymerase (Invitrogen), 1mM of the forward and the reverse primer, between 100 ng and 150 ng

of DNA template (EST constructs) and autoclaved milli-Q water. PCR mixes were assembled as a master mix with DNA and primers being added after aliquoting into 0.2 ml PCR strip tubes with domed caps (Axygen). PCR conditions are listed in table 2.14.

Table 2.14: PCR conditions for the amplification of L-IDH from cDNA

Initial	94 °C	2 mins	
	94 °C	30 sec	35 cycles
Annealing	58 °C	30 sec	
Extension	68 °C	1 min 30 sec	
Final extension	68 °C	10 mins	

PCR machine: PCRSprint Hybaid (Thermo Scientific)

The PCR products were purified with a QIAquick PCR purification kit (Qiagen) to the manufacturer's instructions (DNA was eluted in 30 µl of Buffer EB). The purified PCR fragments were cloned into the D-topo vector (Invitrogen) according to the protocol. 6 µl of the L-IDH-D-topo construct was transformed into Top Ten chemically competent cells (Invitrogen) and plated onto LB plates containing 100 µg/ml ampicillin. The transformation was verified using colony PCR and positive colonies were sequenced using M13 primers (M13F 5’ GTAAAACGACGGCCAGT 3’ M13R 5’ GGAAACAGCTATGACCATG 3’).

NCBI accession numbers of sequencing: L-IDH 3 cDNA: KT150501

Cloning of the three L-IDH isoforms into pBMTh2 vector for transformation into the microvine

The pBMTh2 vector was created (in the Thomas Lab) for transformation of constructs into the microvine using Gateway technology for over-expression of genes. This vector contains a 35s promotor for constitutively active gene expression and a GFP reporter gene

(See Figure A.4 for the vector map).

The L-IDH constructs in D-topo were linearised with a PvuI restriction enzyme digest. Reaction mix consisted of 1 ug of L-IDH -D-topo construct, 1 X Buffer 3 (NEB) 1 X BSA and 2 units of PvuI enzyme (NEB). This was incubated at 37 ^{o}C for 3 hours and then inactivated at 65 ^{o}C for 10 minutes. The mix was then cleaned up using a QIAquick PCR cleanup kit (Qiagen) (DNA was eluted in 30 µl of buffer EB). L-IDH isoforms were transferred to pBMTh2 using an LR recombination reaction. Mix consisted of 20 µl of the cleaned up linearised fragment, 80 ng of pBMTh2 vector and 2 µl of the LR clonase. This reaction was incubated at 25 ^{o}C overnight. After incubation 1 µl of proteinase K was added to stop the reaction and incubated at 37 ^{o}C for 10 minutes. These constructs were transformed into Top Ten chemically competent cells (Invitrogen) and plated on LB plates containing 100 µg/ml kanamycin. Colonies were verified by Colony PCR and positive colonies were sequenced with primers pBMTh2 F 5' GTGCGAGCTCCTTAAGCCAT 3', pBMTh2 R 5' CGAGACGCCTATGATCGCAT 3'. Recombination junctions were also sequenced to tell if there was a good quality exchange of the gene into pBMTh2. Primers were designed in conserved sections of the L-IDH gene so that one primer pair could sequence out from all three genes. The primer for sequencing out from the beginning of the genes was 5' GTGATGAACATCACTTCCAC 3' and the primer for sequencing out from the end of the genes was 5' GAAGCCTTTGAAACCCGTG 3'.

Transformation into Agrobacterium

The constructs were transformed into *Agrobacterium tumefaciens* strain EHA101 cells via electroporation. Colony PCR was performed to check the presence of the constructs in the Agrobacterium.

Creation of grapevine embryogenic callus

Embryogenic cultures were created from anthers of immature flowers from the microvine line V6 which was grown in the glasshouse as described by Franks *et al. (1998)* and Iocco *et al.* (2001) and plated onto PIV media. V6 microvine line is described in the publication by Chaib *et al.* (2010). It is called 04C23V0006 in the publication, which was simplified to V6.

Transfection of microvine callus

Callus was transfected with the Agrobacterium containing the L-IDH isoforms by the method described in (Iocco et al. 2001). When embryos began to appear they were moved to plates containing SM media with 10 μM of BAP for one week, then onto plates containing SM with 5 μM of BAP (Iocco et al. 2001). At this stage the presence of the GFP reporter gene was checked under a UV microscope. GFP expression was checked again for the leaves of plants in SM media in MAGENTA™ GA7-3 containers (Life Technologies) before transferal to soil.

Growth cabinet conditions

The plants were grown in a growth cabinet with 16 hour days at 27 ^{o}C and nights at 20 ^{o}C with 60% humidity. Average light intensity was 600 micromol m^{-2} sec^{-1}.

Collection of leaves and berries from transgenic microvines over-expressing L-IDH isoforms

Pre-veraison berries were collected when the first berries of a bunch began to soften. Mature berry samples were collected three to four weeks after the veraison stage. Berries were snap frozen in liquid nitrogen and stored at -80 ^{o}C. Ten berries from the two time

points were collected from the non-transformed (control) V6 plants for individual berry acid measurement. Leaves were collected for both acid extraction and RNA (for cDNA) extraction. The youngest leaves of a shoot tip were collected. They were snap frozen in liquid nitrogen and stored at -80 ºC.

Acid extraction of LC-MS/MS to determine tartaric acid concentrations in leaves and berries from transgenic microvines over-expressing L-IDH isoforms

For details of the UHPLC-MS/MS method used here to measure the tartaric acid concentrations, see Higginson *et al.* (2015) or Chapter 3 of this thesis. Berries from transgenic vines were ground in samples of 5 to 10 berries and tartaric acid concentration was measured with UHPLC-MS/MS. For V6 control samples, 10 individual berry samples were ground and tartaric acid concentration was measured in each sample with UHPLC-MS/MS.

RNA extraction and cDNA synthesis from leaves from transgenic plants over-expressing L-IDH isoforms to determine the expression of the trans-gene

RNA extraction

RNA from leaf tissue was extracted using the Spectrum plant total RNA kit (Sigma), using protocol A, with a DNAse digest (Qiagen).

To extract RNA from berries a modified version of the process published by Rezaian and Krake (1987) was used. Berries were ground in liquid nitrogen in a mill (IKA, Germany) and then ground to a fine powder with a mortar and pestle. 8 ml of Na-Perchlorate buffer was added to the ground tissue (Rezaian & Krake 1987) and incubated for 1 hour at room temperature before being filtered through a custom made filter of miracloth (Calbiochem)

and silane treated glass wool. This filter was put into a syringe that was inserted into a Falcon tube and centrifuged at 100 g for 10 minutes at 4 ^{o}C. The filtrate was then ethanol precipitated by adding 2.5 volumes of 100% ethanol and incubated at -20 ^{o}C for 1 hour. The pellet was collected by centrifugation at 5,000 g for 20 minutes at 4 ^{o}C. The pellet was then washed in 70% ethanol and centrifuged at 4 ^{o}C for 5 minutes. All ethanol was removed and pellet was left to air dry. This pellet was then used in the Spectrum plant total RNA kit (Sigma) using protocol A with a DNAse digest (Qiagen).

cDNA synthesis

1 µg of RNA in 12.5 µl of H_2O was added to 10 mM dNTPs and 100 µl of oligo-dT primer. This mix was incubated at 65 ^{o}C for 5 minutes then on ice for 1 minute. To this mix was added 5x first strand buffer, 0.1 M DTT and 1 µl of superscript III (Invitrogen). The mix was incubated for 1 hour at 50 ^{o}C and then the superscript III was deactivated at 70 ^{o}C for 10 minutes. cDNA was diluted in 180 µl of H_2O.

qRT-PCR

The pipetting into the 384 well plate was done using a Zephyr pipetting robot.
For each reaction 5 µl of SYBR green mix (Thermo Fisher)with 10 mM of each primer was added and 2.5 µl of standards or cDNA. Standards were diluted in a dilution series starting with 1 µg/µl. Standards were cloned copies of L-IDH 1, 2 or 3 CDS. Primer sets used as standards are actin, ubiquitin and elongation factor 7. qRT-PCR was performed using a Roche light cycler 480 II and all analysis was done with software Biogazelle q Base +.

2.3 Results

2.3.1 There are three isoforms of L-idonate dehydrogenase annotated in the draft grapevine genome

When this work began, only one characterised form of L-IDH had been found in the grapevine genome, along with the possibility that another two isoforms had been annotated based on sequence similarity to the first. An initial keyword search of 'L-idonate dehydrogenase' and '*Vitis vinifera*' on NCBI found three separate L-IDH isoforms present in the draft grapevine genome. From this information on NCBI the position in the genome of all three L-IDH isoforms was determined. The three isoforms are found in tandem on chromosome 16. L-IDH 1 and L-IDH 3 are similar in length while L-IDH 2 is slightly longer and annotated in the opposite direction (Figure 2.1). NCBI estimates the size of the L-IDH 1, L-IDH 2 and L-IDH 3 genes to be 2548bp, 2304bp and 2689bp respectively.

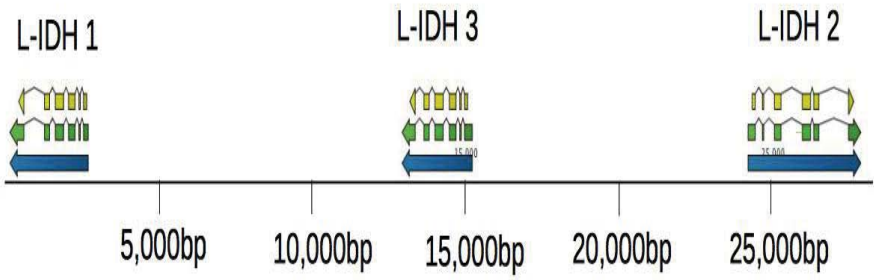


Figure 2.1: The organisation of the three L-idonate dehydrogenase (L-IDH) genes in the draft grapevine genome. NCBI shows that the three isoforms are in tandem on chromosome 16. L-IDH 1 and 3 are similar in length while L-IDH 2 is slightly longer and is facing in the opposite direction.

An alignment was created to determine the similarity of the three genes. This alignment shows that L-IDH 1 and L-IDH 3 are very similar in nucleic acid sequence. L-IDH 2 does have similarity to L-IDH 1 and 3 although it is much more diverse compared to the other two. These sequences were translated into the amino acid sequences and aligned. This alignment showed even more similarity between L-IDH 1 and L-IDH 3. There were three amino acid substitutions between the two isoforms (Figure 2.2).

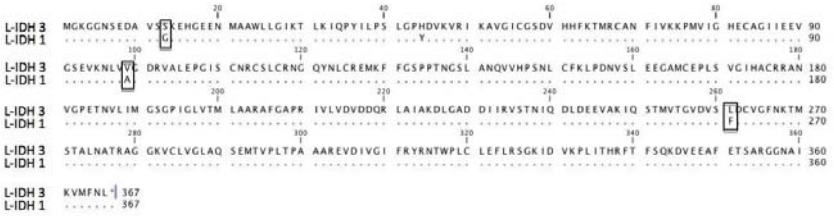


Figure 2.2: An amino acid sequence alignment of L-IDH 1 and L-IDH 3. There are three amino acids different between the two sequences. SER_{12} → GLY, which are both hydrophilic, VAL_{99} → ALA, which are both hydrophobic, and LEU_{261} → PHE, which are both hydrophobic.

Existing Affymetrix gene expression grape berry data sets (NCBI GEO accession GSE7679, GSE7680 and GSE7677) were used to find the expression pattern of the three L-IDH isoforms over berry development. An R script and an Rdata file (M.R. Thomas, unpublished) were used to separate the data for L-IDH 1 (XM_002269900), L-IDH 2 (XM_002269859) and L-IDH 3(XM_002267626) and to compare them to each other. L-IDH 1 and L-IDH 3 had very similar expression patterns matching the expression of tartaric acid accumulation over berry development. L-IDH 2, on the other hand, had a different expression pattern that was high from early development right through to late development, inconsistent with tartaric acid accumulation in grapevine berries (Figure 2.3).

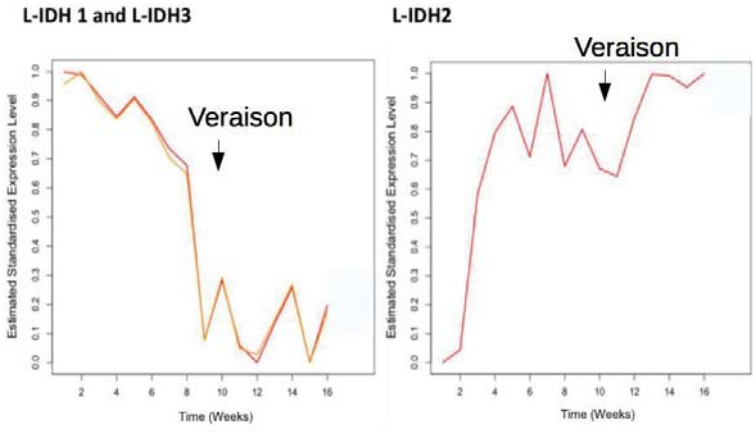


Figure 2.3: The expression of the three L-IDH isoforms from existing Affymetrix gene expression grape berry data sets (NCBI GEO accession GSE7679, GSE7680 and GSE7677). L-IDH 1 and L-IDH 3 have very similar expression patterns which mirror tartaric acid accumulation in the berry, while L-IDH 2 has a different expression pattern which is high throughout berry development.

The nucleic acid and amino acid sequences of the three L-IDH isoforms were aligned to the sequence of the original L-IDH isoform. At the time that the original L-IDH was characterised it was unknown that two more isoforms were present in the genome. This alignment showed that the original L-IDH is more similar to L-IDH 3 then L-IDH 1 (Figure 2.4) with two amino acid differences between L-IDH 1 and the original compared to the one between L-IDH 3 and the original. Therefore we believe L-IDH 3 was characterised by DeBolt *et al.* (2006) and the two uncharacterised isoforms are L-IDH 1 and L-IDH 2.

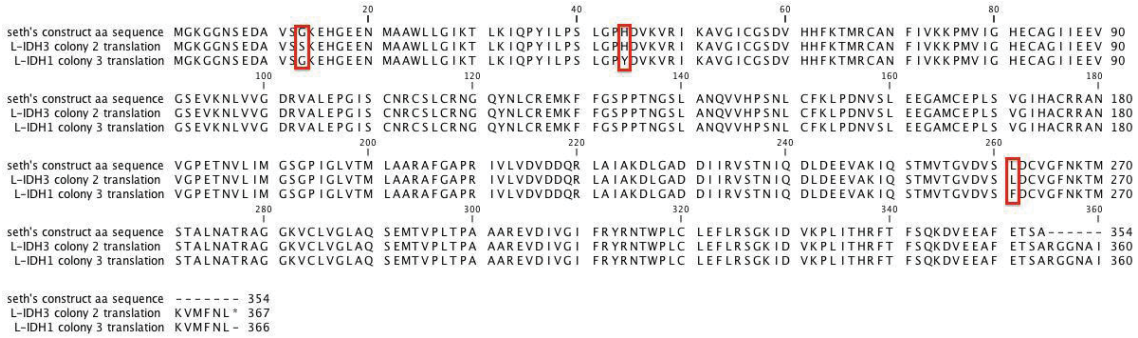


Figure 2.4: Amino acid sequence alignment of L-IDH 1 and L-IDH 3 compared to the original construct of L-IDH that was characterised previously (DeBolt 2006). L-IDH 3 and the original construct are the closest in sequence with only one amino acid difference. Boxes show the three amino acid differences.

2.3.2 Sequencing of L-idonate dehydrogenase isoforms from genomic DNA of microvine homozygous lines to determine if L-IDH 1 and L-IDH 3 are alleles at the same locus or two separate genes

Several homozygous lines of microvine have been developed for genetic studies (Chaib et al. 2010). These vines have been tested with an extensive set of DNA markers to show that they are homozygous across the whole of the genome except for the GA insensitive locus and the flower sex locus (Chaib et al. 2010). Leaves were collected from two of these homozygous lines V3 and V10 and the L1 heterozygous microvine parent, which was used as a control. Before they were used in this study, the previous marker analysis was verified to determine that the two homozygous lines were indeed homozygous in the region around the L-IDH isoforms on chromosome 16 (data not shown).

DNA from the chosen homozygous lines was extracted from young leaf tissue. This DNA was then used in PCR to amplify the three isoforms using primers for the beginning and end of each gene. These fragments were run on a 1% (w/v) agarose gel and compared to the GeneRuler 1 kb DNA ladder (ThemoScientific) to determine the size of the fragments (Figure 2.5). L-IDH 1 was estimated to be 2.5 kbp, L-IDH 2 was estimated to be 3.7 kbp and L-IDH 3 was estimated to be 1.7 kbp. Sizes of the fragments were consistent across the two homozygous lines and the L1 heterozygous parent.

L-IDH 1 and *3* were present in both homozygous lines showing that they are two genes and not alleles of the same gene. There was also a difference in size between the *L-IDH 1* and *L-IDH 3* sequences amplified from the homozygous lines and the L1 parent, indicating that the *L-IDH 1* primers did not amplify *L-IDH 3* and vice versa.

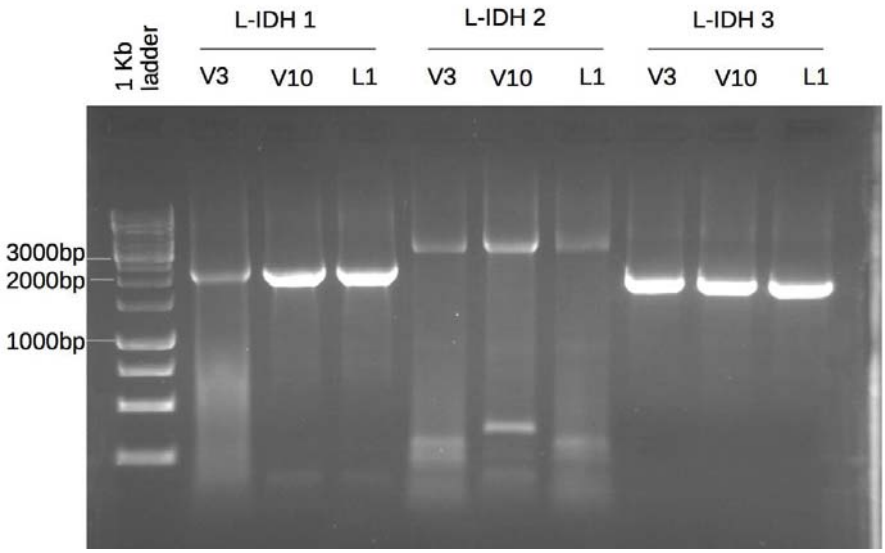


Figure 2.5: A 1% agarose gel showing the three L-IDH isoforms amplified from genomic DNA from two homozygous microvines (V3 and V10) and their heterozygous L1 parent. Non-specific amplification appeared in the L-IDH 1 V10 and L1 lanes and in the L-IDH 2 lanes. These bands are smaller than 500 bp.

The amplified *L-IDH 1*, *2* and *3* isoforms from the L1 microvine line were cloned and sequenced. For *L-IDH 1*, the gene from V10 was cloned and sequenced as well the L1 sequence. For *L-IDH 3*, the genes from both V10 and V3 were cloned and sequenced as well as the L1 sequence. Sequences were aligned to the genomic *L-IDH* sequences from NCBI using CLC workbench 7 (see materials and methods for details) (see supplementary data for sequence alignments). There were many differences between the NCBI sequences and the sequences obtained for the microvine lines for each of the three *L-IDH* isoforms. Both the V10 and V3 sequences for *L-IDH 3* and the *L-IDH 1* V10 sequences have only one allele of the two L1 parent alleles.

For *L-IDH 2*, a large, approximately 400 bp part of one intron was missing compared to the NCBI sequence for both the alleles from the L1. There were many Single Nucleotide Polymorphisms (SNPs) between both the NCBI sequences compared to both L1 sequences

and also between the two L1 alleles.

For *L-IDH 1* there were a few SNPs between the alleles and one allele matched the *L-IDH 1* sequence from NCBI. For *L-IDH 3* there were SNPs between the two alleles and one allele matched the NCBI sequence. *L-IDH 3* differed in size from the size of the NCBI L-IDH 3 gene sequence. A section of the third intron had not been sequenced and NCBI had predicted the size of the intron with Ns. For the L1 parent and the two homozygous lines sequencing of this intron was successful showing that it is 43 bp smaller than NCBI predicted. One of the L1 alleles also contained two deletions in introns four and five sized 16 bp and 19 bp respectively.

2.3.3 Expression, isolation and activity of the three L-idonate dehydrogenase isoforms

The three L-IDH cDNAs were cloned and placed into the pET52b (Novagen) vector for protein expression. The constructs were sequenced to demonstrate that there were no errors and that the whole sequence was in frame. The sequencing also encompassed the terminator and promoter sequence to determine if they were intact. After sequencing, the constructs were transformed into BL21(DE3) pLysS cells (Novagen). At first a basic protein expression method (see materials and methods section) was tried with 0.4 mM ITPG and an overnight induction at 16 ^{o}C. When this method yielded no protein detectable on a Coomassie blue stained SDS-PAGE gel, an induction for 3 hours at 37 ^{o}C was tried which also yielded no protein.

Due to the scarcity of L-idonate, the elution fractions were tested for activity with the reverse reaction with the substrate 5-keto-D-gluconic acid and NADH as a co-factor (see materials and methods for details). No change in absorbance was seen at 340 nm for any elution or wash fractions indicating that there was no active protein present.

The three L-IDH constructs were transformed into BL21(DE3) cells with no pLysS plasmid (Novagen) in case protein expression was repressed too much with the presence of the pLysS plasmid (Novagen 2011). The L-IDH proteins were induced and purified in the same method as above with induction at 37 ^{o}C for 3 hours. Again no protein expression was seen on an SDS-PAGE gel and there was no change in absorbance when the protein elutions were used in an activity assay.

The breakdown of ampicillin can occur in cultures from the production of β-lactamase

and the acidic pH created as the bacteria grow (Novagen 2011). This can lead to the loss of the plasmid during incubation and consequently, no protein expression. Ampicillin was replaced with the more stable carbenicillin in cultures and the protocol was changed slightly. The overnight culture was spun down to pellet cells, then re-suspended in LB before being used to inoculate the 100ml culture. This was then done again before induction. Each time, new antibiotics were added. Two higher concentrations of IPTG, 1 mM and 2 mM, were trialled to try and increase protein expression. After the addition of the IPTG the culture was incubated at 30 ^{o}C instead of 37 ^{o}C for 3 hours. There was no protein expression seen on an SDS-PAGE gel nor were any changes seen in absorbance during activity assays.

During the characterisation of the original L-IDH isoform there were no issues in protein expression and high concentrations of protein were purified (DeBolt 2006). This original construct had been kept in as a glycerol stock at – 80 ^{o}C and could be used as a positive control. The original construct was cloned into vector pET 14b so the three L-IDH isoforms were also cloned into the pET 14b vector (Novagen) to match the original construct. The original construct and the three L-IDH isoform constructs were then transformed into the same stock of BL21(DE3) pLysS cells.

After cloning L-IDH isoforms into the pET 14b vector they were sequenced to show that the genes were in frame and that the promoter, terminator and His tag were correct. With these new constructs in pET 14b protein expression was attempted with induction at 37 ^{o}C for 3 hours with 0.4mM IPTG (see materials and methods). The protein elutions were then run on an SDS-PAGE gel but no intense band of protein was distinguishable from the background bands on the gel for any of the isoforms, including the positive control.

However when 100 µl of the second protein elution was added to 5 keto-D-gluconic acid in a protein activity assay a change in absorbance was seen for both the positive control and L-IDH 3 although not for L-IDH 1. This meant that these proteins were being expressed but in a very low concentration and were unable to be seen on an SDS-PAGE gel. It also meant that L-IDH 1 could either be inactive or not present in a high enough concentration for activity to be seen with these assays.

Now knowing that protein concentration was low, focus was placed on increasing protein expression levels. Firstly, different induction temperatures and time lengths were trialled to determine if the protein was mostly insoluble. Proteins can become more soluble if longer inductions at lower temperatures are tried. This trial yielded no difference in protein expression. Insoluble fractions were run on an SDS-PAGE gel which showed no indication of large amounts of insoluble protein. Next, all the buffers and HisTalon resin were discarded and a HisTalon kit Gravity column purification kit (Clontech) was purchased, in case the old buffers or resin were contaminated with something that degraded proteins. But this kit did not increase purified protein concentrations after cells were induced at 37 ^{o}C for 3 hours with 0.4 mM IPTG. Then a higher IPTG concentration of 1mM and terrific broth instead of LB media were tried but this also yielded no change in protein expression levels.

Next, L-IDH constructs were re-transformed into BL21 (DE3) pLysS, since fresh colonies are understood to give better expression levels. To create overnight cultures, colonies were picked straight off the transformational plate. This did not appear to increase the concentration of protein.

In case there was an issue with our cell stock, a new stock of BL21 (DE3) pLysS cells was purchased (Novagen) and the constructs transformed into the new cells. An SDS-PAGE gel confirmed that this made no difference to protein expression.

At this time the reason there was no protein expression was unknown. Many options for increasing protein expression had been tried and all the buffers, the HisTalon resin and the cell stock had been replaced. Finally, a different protocol was trialled with larger cultures in a different lab. This protocol used a FPLC protein purification instead of the gravity columns, 1 L cultures instead of 100 ml and a cell disruptor for cell lysis. The lab provided us with LB, their incubator, and their equipment which meant new stocks of all reagents. Using this new method a larger concentration of the L-IDH positive control was purified and could be seen on a Coomassie stained SDS-PAGE gel (Figure 2.6). It was also active when tested in an activity assay with 5 keto-D gluconic acid. This method was then used to express and purify L-IDH 1 and L-IDH 3. L-IDH 1 produced a protein that could be seen on an SDS-PAGE gel (Figure 2.6) while the L-IDH 3 did not. For activity assays the L-IDH positive control, which closely matched L-IDH 3 as a control was used to compare to L-IDH 1. Future characterisation of L-IDH 2 was not attempted as by this time it had been established that L-IDH 2 was most likely a sorbitol dehydrogenase (Jia et al. 2015).

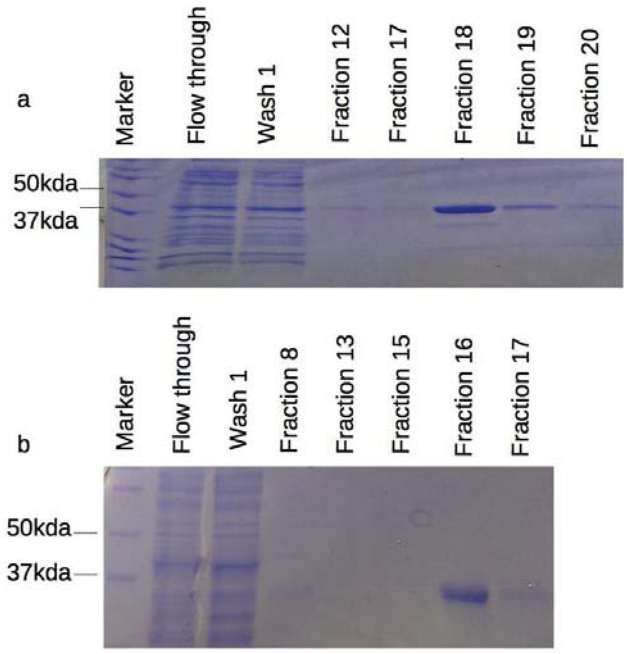


Figure 2.6: SDS-PAGE gels showing fractions from the FPLC purification of a) L-IDH original construct and b) L-IDH 1. a) Protein was estimated to be 40 kda. Most protein was eluted in fraction 18 with small amounts in fractions 19, 20 and also 12. b) Protein was also estimated to be 40 kda but proved to be a little smaller, most likely due to imperfections in the running of the gel. Protein was eluted in fraction 16 with a small amount in fraction 17.

Activity was seen with L-IDH 3 but not with L-IDH 1 for both the forward reaction with L-idonate and the reverse reaction with 5 keto-D gluconic acid (Figure 2.7 and 2.8). However the protein concentration was low, so the activity was low compared with the original tests (DeBolt 2006). The change in absorbance was very slight for L-IDH 3, with just a 0.01 change in absorbance for L-idonate and a 0.05 change for 5 keto-D gluconic acid. It is unknown whether L-IDH 1 would be active with a higher protein concentration. It was concluded that more experiments are needed with a higher concentration of L-IDH 1 to determine if it is active.

The band on the SDS-PAGE gel for L-IDH 1 (Figure 2.6 b) was slightly smaller than the estimated 40 kda that was expected. This may be due to the imperfect running of the gel. However it is also possible that L-IDH 1 was truncated and may explain why there was no activity. It was concluded that further work needs to be conducted to determine if L-IDH 1 is active or inactive.

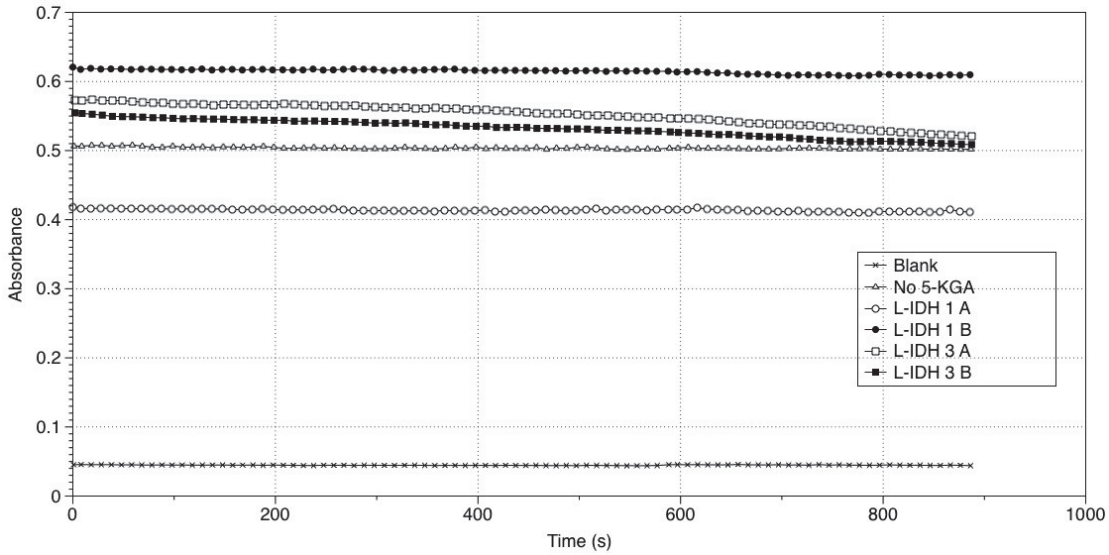


Figure 2.7: Activity of L-IDH 1 (replicates A and B) and L-IDH 3 (replicates A and B) (original construct) with 5-keto-D-gluconic acid (5-KGA). No 5-KGA control contained L-IDH 1 protein and NADH. Buffer was Tris pH 8.0. Both replicates of L-IDH 3 showed a decrease in absorbance as NADH was used up in the reaction.

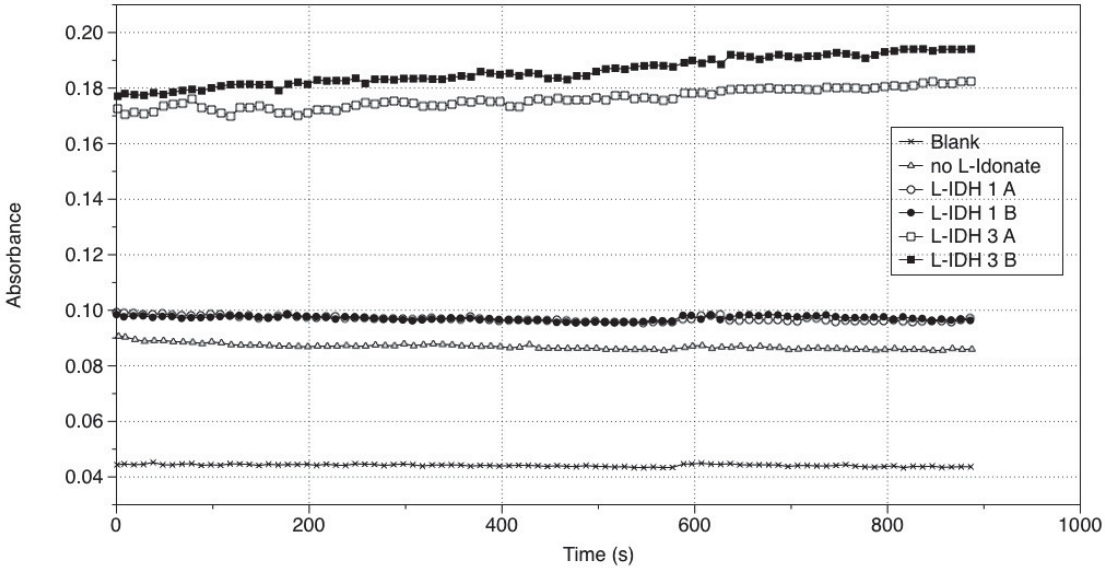


Figure 2.8: Activity of L-IDH 1 (replicates A and B) and L-IDH 3 (replicates A and B) (original construct) with L-idonate. No L-idonate control contained L-IDH 1 protein and NADH. Buffer was Tris pH 8.0. Both replicates of L-IDH 3 showed an increase in absorbance as NAD+ was converted to NADH in the reaction.

2.3.4 Over-expression of the three isoforms of L-idonate dehydrogenase in the microvine

cDNAs of the three L-IDH isoforms were transformed into the D-topo (Invitrogen) vector then into a Gateway compatible binary vector made for the expression of genes into the microvine (vector produced in the Thomas Lab). This vector (pBMTh2) had a constitutively active 35s promotor for gene expression in the microvine and GFP as a marker for transformation. The L-IDH-pBMTh2 constructs were transformed into Agrobacterium which was used for transfection of microvine callus (microvine line V6). Transfected callus was viewed under a UV microscope. Any callus that had been successfully transfected glowed green due to GFP expression (Figure 2.9). This transformed callus was then allowed to differentiate. When differentiation began, the callus was moved to a medium that promoted the creation of shoots, and when shooting occurred the shoots were moved to a medium that promoted the creation of roots. At the small plant stage they were moved to a magenta container to grow larger (Figure 2.9). It was at this stage that GFP expression of the small plants was checked by removing one leaf and studying it under a UV microscope (Figure 2.9). GFP can be seen in the veins of the leaf. Finally, plants were moved to soil and transferred to a growth room (Figure 2.9).

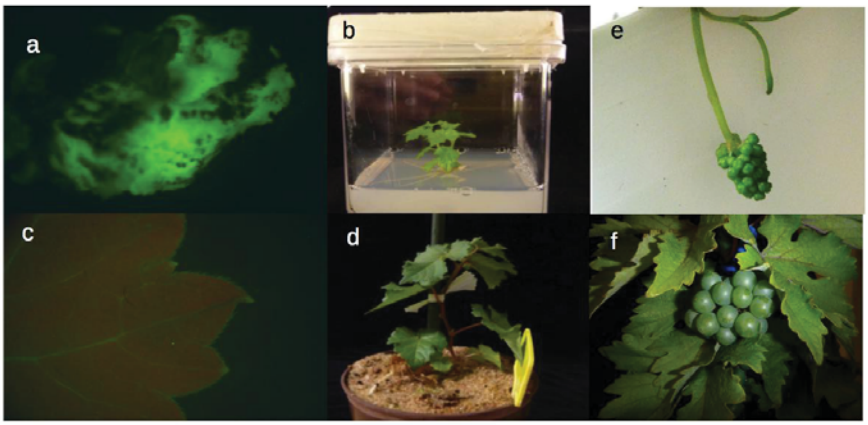


Figure 2.9: Creation of L-IDH over-expressing lines. a) Transgenic microvine callus showing GFP expression indicating the presence of the construct containing the L-IDH gene. b) L-IDH over-expressing plant in SM media in a magenta container to promote root and shoot growth before moving to soil. c) Leaf from an L-IDH over-expressing plant that shows that the transgene is present in leaves. The red filter is for chlorophyll. GFP is present in the tips and veins. d) Plants were moved to soil and left to grow in a growth chamber until they produced berries. e) Plants produced flowers and then f) fruit, from which the concentration of tartaric acid was measured.

The expression of the L-IDH transgenes was checked in each over-expressing plant with qRT-PCR on leaf cDNA. Four plants were identified expressing the GFP marker however, from this analysis, there was one plant that did not show any expression of the L-IDH 1 transgene and was not analysed further (Figure 2.10 a). Five plants containing the transgene were checked for the L-IDH 2 transgene (Figure 2.10 b) and four plants were checked for the L-IDH 3 transgene (Figure 2.10 c).

Samples of 5 to 8 berries were collected for tartaric acid concentration measurement from the L-IDH over-expressing plants at two time points: a pre-veraison time point and a point three to four weeks after veraison (mature berry time point). For the V6 untransformed control, 10 individual berry samples were collected from the two time points.

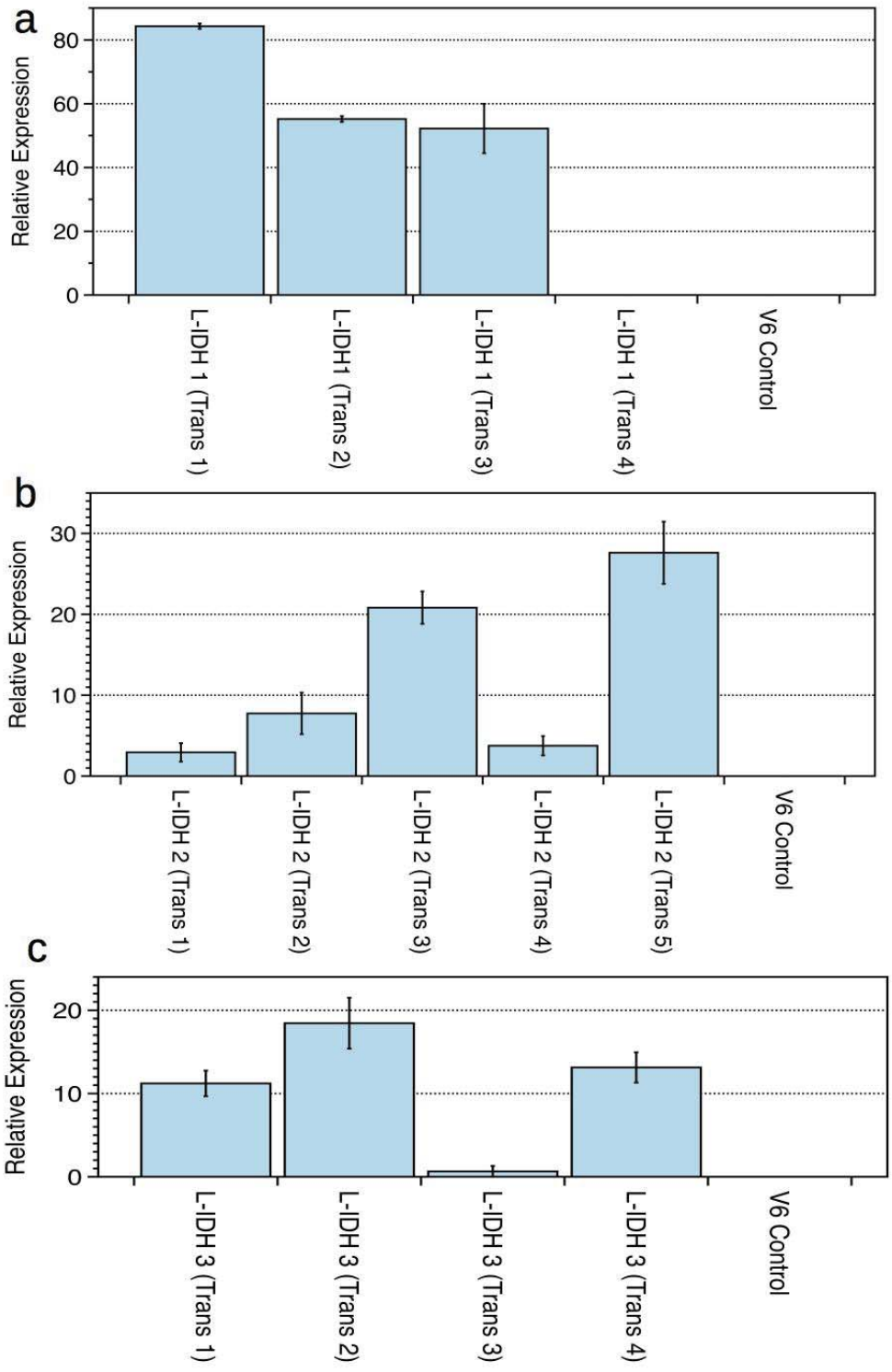


Figure 2.10: Expression of the L-IDH isoforms in leaf tissue from the transgenic L-IDH over-expressing microvines. a) The expression of transgenic L-IDH 1 in L-IDH 1 over-expressing microvines. b) The expression of transgenic L-IDH 2 in L-IDH 2 over-expressing microvines. c) The expression of transgenic L-IDH 3 in L-IDH 3 over-expressing microvines. V6 control is the untransformed negative control.

Leaves were also collected from the L-IDH over-expressing microvines and the V6 control for tartaric acid concentration measurement and qRT-PCR analysis.

For the pre-veraison time point there are two over-expressing plants, L-IDH 3 (trans 2) and L-IDH 3 (trans 4), which showed tartaric acid concentrations higher than the range of the V6 untransformed control's acid concentrations (Table 2.15 and 2.16). This indicated that there may be some tartaric acid accumulation in these plants that is due to the over-expression of the transgene. However, at the harvest time point, these plants had a tartaric acid concentration within the range of the controls.

Table 2.15: Tartaric acid concentrations (mg/g fresh weight) from 10 berries of the V6 untransformed control plant at the pre-veraison time point.

Sample name	Tartaric acid mg/g FW
V6 Pre-veraison 1	7.05
V6 Pre-veraison 2	5.83
V6 Pre-veraison 3	6.33
V6 Pre-veraison 4	8.48
V6 Pre-veraison 5	3.89
V6 Pre-veraison 6	7.06
V6 Pre-veraison 7	8.30
V6 Pre-veraison 8	6.84
V6 Pre-veraison 9	7.85
V6 Pre-veraison 10	5.96
Mean	6.76
Range	3.89-8.48

Table 2.16: Tartaric acid concentrations (mg/fresh weight) from pre-veraison time point berries from L-IDH 1, L-IDH 2 and L-IDH 3 over-expressing plants.

Sample name	Tartaric acid mg/g FW
L-IDH 1 (trans 1)	2.87
L-IDH 1 (trans 2)	1.14
L-IDH 2 (trans 1)	5.95
L-IDH 2 (trans 2)	3.91
L-IDH 2 (trans 3)	4.05
L-IDH 2 (trans 4)	2.97
L-IDH 2 (trans 5)	4.64
L-IDH 3 (trans 1)	4.64
L-IDH 3 (trans 2)	16.03
L-IDH 3 (trans 3)	3.26
L-IDH 3 (trans 4)	10.22

For the harvest berry time point L-IDH over-expressing plants had tartaric acid concentrations which were either within the range of the V6 untransformed controls or lower than the range of acid concentration seen in the V6 untransformed controls (Tables 2.17 and 2.18). The exception was L-IDH 1 (trans 3) plant which had very small berries (approximately 1 cm in diameter) that never produced seeds. The high tartaric acid concentrations seen in this plant may be caused by the small berries rather than by the transgene. Two of the other three L-IDH 1 over-expressing plants produced a low number of normal fertile berries with the rest of the bunch being small and infertile.

Table 2.17: Tartaric acid concentrations (mg/g fresh weight) from 10 berries of the V6 untransformed control plant at the mature berry time point.

Sample name	Tartaric acid mg/g FW
V6 mature berry 1	5.43
V6 mature berry 2	5.78
V6 mature berry 3	6.18
V6 mature berry 4	5.47
V6 mature berry 5	5.68
V6 mature berry 6	5.10
V6 mature berry 7	4.09
V6 mature berry 8	5.93
V6 mature berry 9	4.63
V6 mature berry 10	5.65
Mean	5.39
Range	4.08 – 6.18

Table 2.18: Tartaric acid concentrations (mg/g fresh weight) from mature berry time point berries from L-IDH 1, L-IDH 2 and L-IDH 3 over-expressing plants.

Sample name	Tartaric acid mg/g FW
L-IDH 1 (trans 1)	5.66
L-IDH 1 (trans 2)	5.40
L-IDH 1(trans 3)	11.38*
L-IDH 2 (trans 1)	2.43
L-IDH 2 (trans 2)	2.72
L-IDH 2 (trans 3)	2.48
L-IDH 2 (trans 4)	2.64
L-IDH 2 (trans 5)	3.47
L-IDH 3 (Trans 1)	2.82
L-IDH 3 (trans2)	2.84
L-IDH 3 (trans 3)	3.08
L-IDH 3 (trans 4)	2.35

* The plant only produced small berries with no seeds.

In leaves, the tartaric acid concentrations of the L-IDH over-expressing microvines lay within the range of concentrations seen in the untransformed V6 controls or were lower than the controls (Tables 2.19 and 2.20). Leaf acid seems to vary considerably between leaves of the same plant. The tartaric acid concentrations of the V6 untransformed controls had the largest range compared to tartaric acid concentrations from pre-veraison or harvest berries.

Table 2.19: Tartaric acid concentrations (mg/g fresh weight) from 10 leaves from the V6 untransformed control plant.

Sample name	Tartaric acid mg/g FW
V6 Leaf 1	3.27
V6 Leaf 2	14.12
V6 Leaf 3	16.54
V6 Leaf 4	11.32
V6 Leaf 5	11.23
V6 Leaf 6	14.31
V6 Leaf 7	6.99
V6 Leaf 8	17.46
V6 Leaf 9	9.26
V6 Leaf 10	15.68
Mean	12.02
Range	3.27-17.46

Table 2.20: Tartaric acid concentrations (mg/g fresh weight) from leaves of L-IDH 1, L-IDH 2 and L-IDH 3 over-expressing plants.

Sample name	Tartaric acid mg/g FW
L-IDH 1 (trans 1)	8.64
L-IDH 1 (trans 2)	5.16
L-IDH 2 (trans 2)	1.39
L-IDH 2 (trans 3)	1.76
L-IDH 2 (trans 4)	2.05
L-IDH 3 (trans 1)	3.91
L-IDH 3 (trans 3)	2.91
L-IDH3 (trans 4)	4.54

2.4 Discussion

When this research began, very little was known about the genes in the pathway of tartaric acid formation in the berry. To date, only two steps have candidate genes associated with them. 2-keto L-gulonate reductase converts 2-keto L-gulonic acid into L-idonate and has been shown to have strong activity with 2-keto gulonic acid (Burbidge 2011). The other characterised gene encodes an enzyme that catalyses the next step in the pathway, the conversion of L-idonate to 5-keto gluconic acid. The protein for this step has been shown to have activity converting L-idonate to 5-keto gluconic acid *in vitro* and is called L-idonate dehydrogenase (DeBolt et al. 2006).

L-IDH was discovered and characterised before the sequencing of the grapevine genome. By searching the draft grapevine genome, two more L-IDH-like genes were uncovered. By studying the annotation of the genome from NCBI it was determined that all three L-IDHs are in tandem on chromosome 16 in the order *L-IDH 1, L-IDH3 and L-IDH 2* (Figure 2.1). L-IDH 2 was also annotated in the opposite direction and is slightly longer than L-IDH 1 and L-IDH 3.

Alignments of the amino acid sequences from NCBI of these three L-IDH isoforms showed that L-IDH 1 and L-IDH 3 are indeed very similar to each other with only three amino acids different between the two (Figure 2.2), while L-IDH 2 is much more diverse in sequence compared to L-IDH 1 and L-IDH 3. The three amino acid changes between L-IDH 1 and L-IDH 3 are Serine to Glycine, which are both hydrophilic, Valine to Alanine, which are both hydrophobic, and Leucine to Phenylalanine, which are both hydrophobic. It is interesting to note that the hydrophobicity of each of the amino acids has not been

changed but the overall hydrophobicity of L-IDH 1 and L-IDH 3 are different. The Kyte and Doolittle method of determining the hydropathic character of a protein estimates a difference of 3 between the two (Kyte & Doolittle 1982). This method assigns a hydropathy number to each amino acid and these are then added together to get the proteins final hydropathic character. What these changes mean for protein activity is unclear. It is hoped that a structure of one of both of these isoforms of L-IDH can be resolved and this may add some insight into protein activity.

The expression pattern of L-IDH 2 over development is also different to the expression patterns of L-IDH 1 and L-IDH 3 (Figure 2.3). The expression patterns of L-IDH 1 and L-IDH 3 mirror the formation of tartaric acid over development: They begin high in early development and then drop sharply around veraison, when tartaric acid ceases to be made in the berry. By contrast, L-IDH 2 has an expression pattern that is high throughout development. The expression pattern of L-IDH 1 and L-IDH 3 look extremely similar and this may be due to the microarray probes cross hybridizing to both isoforms because of their similarity in sequence. Published studies have used qRT-PCR to determine relative expression of L-IDH 1 and L-IDH 3 but the expression values were for both of these isoforms together (Sweetman et al. 2012, Wen et al. 2014). RNAseq data shows that L-IDH 1 and L-IDH 3 are both expressed in early development but not in later development (Sweetman et al. 2012). Wen *et al.* (2014) used an L-IDH antibody to follow protein concentration across development and found that L-IDH was high early development and then had a stable lower expression for the rest of berry development. The authors do point out that this antibody may be able to bind to all three L-IDH isoforms. Martinez-Esteso *et al.* (2011) looked at proteins which varied across development and found that L-IDH protein was high in concentration in the early stages and then dropped over development.

In chapter 5 work is described in which qRT-PCR was attempted to determine relative expression of the three isoforms of L-IDH from different tissues of the grapevine including pre-veraison and mature berries. However, due to the similarity at the 3' end of the coding sequence the separation of L-IDH 1 and L-IDH 3 proved impossible by qRT-PCR and a pyrosequencing method was used to distinguish the expression of L-IDH 1 and L-IDH 3.

This analysis of the L-IDH isoforms presented in 2.3.1 posed several questions. Firstly, since L-IDH 2 is more diverse in sequence and has a different expression pattern that does not mirror tartaric acid synthesis in berries, is this isoform actually involved in tartaric acid synthesis? Secondly, since L-IDH 1 and L-IDH 3 are so similar, is it possible they were annotated incorrectly in the grapevine genome and that they are, in fact, alleles of the same gene? Thirdly, if L-IDH 1 and L-IDH 3 are separate loci, then do they have differences in activity or is one inactive? And finally, with the development of the microvine, it is possible to study the function of the L-IDH isoforms *in planta?*

Differences of L-IDH 2 – is it a sorbitol dehydrogenase?

A recent study (Jia et al. 2015b) used comparisons of sequences from other species to show that L-IDH 2 is in fact more closely related to sorbitol dehydrogenases than L-idonate dehydrogenase. This study also used modelling to show that these sorbitol dehydrogenases can be split into two categories, one containing L-IDH 2 and other sorbitol dehydrogenases from other species and the other containing L-IDH 1 and L-IDH 3 and other L-idonate dehydrogenases. The fact that L-IDH 2 is not an L-IDH would explain why its expression pattern does not mirror tartaric acid synthesis in the berry. It would be interesting to explore whether this protein has any activity with L-idonate. This had been tested once before but this study was a preliminary one and reaction conditions were not

optimised and there was no detectable activity with sorbitol (Burbidge 2011).

L-IDH 1 and L-IDH 3: alleles of the same gene?

The grapevine sequence found on NCBI comes from the sequencing of the near homozygous clone PN40024 (Jaillon et al. 2007). Wild-type grapevines are normally highly heterozygous. However, for the sequencing project, a near-homozygous plant was created. While this clone is considered to be homozygous, it is estimated to be 7% heterozygous. It may have been possible that one of these areas of heterozygosity was over the *L-IDH* genes, making *L-IDH 1* and *L-IDH 3* alleles of the same gene and not separate loci.

To answer this question, the three L-IDH isoforms were amplified from homozygous microvines. The answer was seen immediately when the PCR products of the amplification of *L-IDH 1* and *L-IDH 3* from genomic DNA from the two chosen homozygous lines were run on an agarose gel (Figure 2.5). *L-IDH 1* and *L-IDH 3* are not alleles of the same locus but two separate loci that have high similarity as both genes are seen in both homozygous lines and are different sizes. *L-IDH 2* did not amplify as well as *L-IDH 1* and *L-IDH 3* as brightness of bands of the agarose gel indicates that there is less L-IDH 2 present (Figure 2.5). The most likely reason for this difference in amplification efficiency is that the primers for *L-IDH 2* were not as efficient at the annealing temperature used in the PCR reaction, as all three isoforms were amplified in the same reaction and optimisation of individual primer sets was not attempted. If the PCR reaction was optimised for *L-IDH 2* primer set only there may have been a better result.

To definitively conclude that *L-IDH 1* and *L-IDH 3* are two separate genes the alleles from

the L1 parent and the two homozygous lines were sequenced. Because there were two different alleles in the L1 plant for both *L-IDH 1* and *L-IDH 3* it is impossible for them to be at the same locus and this supports our agarose gel results (Figure 2.5). Sequencing of these genomic DNA fragments showed that there are many differences in the sequence in the introns of *L-IDH1 and L-IDH 3*. In the L1 heterozygous parent there were two alleles of *L-IDH 1, L-IDH 2 and L-IDH 3*. The variations between the two L1 alleles for *L-IDH 3* was the most varied, with one allele containing two deletions not seen in the other allele or the sequence from NCBI.

The sequences for the three genes also have differences from the sequence of Pinot Noir that is deposited on NCBI including deletions of sequence. It is interesting that the sequences of *L-IDH 1 and L-IDH 3* are so varied, however the majority of these changes appear in the introns and are unlikely to make any difference to protein activity. There is only one change in the coding sequence of *L-IDH 1* that would lead to an amino acid substitution of Arg to Thr at position 282. This SNP appears in the two sequences of one of the L1 parent alleles. More sequencing of this position would need to be conducted to show that this substitution did exist. It is also unknown if L-IDH 1 is an active protein.

A section of intron three in the NCBI *L-IDH 3* sequence was not sequenced and its size predicted with Ns. This intron was sequenced for both L1 alleles and the two homozygous lines and provides new sequence information that can be added to correct the genome sequence.

Activity difference between L-IDH 1 and L-IDH 3

The activity of L-IDH 1 was unable to be properly tested due to issues with low protein

expression. Many different optimisation methods were attempted with no positive results. Finally a new method was trialled which was partially successful as it generated enough L-IDH 1 and control L-IDH 3 that could be seen on a Coomassie stained SDS-PAGE gel (Figure 2.6). However, this protein was not used fresh and the freeze thaw process may have reduced activity so that it could not be detected in activity assays (Figures 2.7 and 2.8) as activity of the L-IDH 3 control was also very low, much lower than reported previously (DeBolt et al. 2006). As a result it cannot be conclusively stated that it was active. However, the new method was not optimised fully and it may be that with appropriate optimisation of the new method, large amounts of protein could be produced. L-IDH 3 has been expressed previously (DeBolt et al. 2006, Wen et al. 2010, 2014) and both methods of expression were similar to the methods that were used to express the L-IDH proteins in this work. Wen *et al.* (2010, 2014) used a different vector from the same family as pET 14b called pET -30 A. It may be that the subtle differences in the make up of this vector allowed for better expression of L-IDH 1. While two vectors were tried as part of the optimisation of L-IDH expression, it may be that more vectors need to be trialled to find the vector that provides the highest expression of L-IDH proteins. It is interesting to speculate on whether L-IDH 1 does have activity. Grapevines do produce a lot of tartaric acid compared to other plant species and having two active copies of L-IDH may explain that.

Characterisation of L-IDH isoforms in planta

Transgenic grapevines for experimental purposes are time consuming to make, require lots of space to store and have a long generation time (Chaib et al. 2010). This makes grapevines difficult to use as an experimental system. However, the development of the microvine and its ability to be easily transformed, its ability to produce fruit year round in

a glass house and its dwarfism make it an ideal system to study genes *in planta* (Chaib et al. 2010). L-IDH 1, L-IDH 2 and L-IDH 3 could all be studied in this fashion. L-IDH 3 had only been characterised previously *in vitro* (DeBolt et al. 2006) and this *in vitro* study had not been translated into an *in planta* study. L-IDH 1 and L-IDH 2 had not been characterised at all.

L-IDH is thought to be the rate limiting step of the tartaric acid synthesis pathway (Saito & Kasai 1984) so it was hypothesised that an increase in tartaric acid may occur with the over-expression of this gene alone. However, the results were far from conclusive. Tartaric acid concentrations above the control concentrations were seen for two of the L-IDH 3, L-IDH 3 trans 2 and L-IDH 3 trans 4, over-expressing plants at the pre-veraison time point but not for any of the others (Tables 2.17 and 2.18). This increased tartaric acid concentration seen in the pre-veraison berries of these two plants is not seen in the harvest samples as concentrations at this time point are within the range of concentrations seen in the controls. Expression of the transgenes in leaves was tested with qRT-PCR in all the L-IDH over-expression plants (Figure 2.10). The L-IDH 3 trans 4 over-expressing microvine had the second highest expression levels of the L-IDH 3 over-expressing plants and the highest leaf tartaric acid concentrations. This suggests that there might be some link between expression of the transgene and increased tartaric acid concentration. Nothing is known about how tartaric acid synthesis is regulated in the berry. It is not known what genes are involved in the regulation of the steps of the pathway, how the pathway switches off at veraison or how much L-idonate is made in the plant. It may be that there is a limiting amount of L-idonate making it impossible to create more 5-keto-D-gluconic acid and therefore more tartaric acid. In future experiments, the intermediates of the pathway may be able to be studied by feeding leaves with C_{13}-labelled ascorbic acid (the precursor

of the tartaric acid synthesis pathway) and looking for where the majority of the C_{13} is found. At the present time L-idonate is very difficult to create meaning that the creation of a C_{13} standard for L-idonate for mass spectrometry is not feasible.

One L-IDH 1 over-expressing plant (L-IDH 1 trans 3) had concentrations of acid above the control concentrations for harvest time point (Tables 2.15 and 2.16). This individual had small berries that never produced seeds and stayed approximately 1 cm in diameter. All the L-IDH 1 over-expressing plants had issues related to berry fertility. Of the other three L-IDH 1 over-expressing plants, two plants produced a low number of normal fertile berries with the rest of the berries in the bunch being small and seedless and one plant never produced fruit at all with flowers dying after cap fall. Since these issues were also present in L-IDH 1 trans 4 which does not show expression of the transgene this suggests that the small berry size, high concentration of tartaric acid and the fertility issues may be due to the transformation itself and not effects of the over-expression of L-IDH 1 (Figure 2.10).

Leaf tartaric acid concentrations from the L-IDH over-expressing microvines were all within the range of the V6 untransformed control's concentrations. Leaf acid varies much more than berry acid but the reason for this is unknown (Tables 2.19 and 2.20) (see chapter 5 for more on leaf acid concentrations).

Tartaric acid addition in wineries to maintain low pH and stop microbial spoilage and oxidation during vinification is a large cost to winemakers. Yet, tartaric acid synthesis in the grapevine is poorly understood. Understanding more about the genes involved in tartaric acid synthesis may allow the use of these genes to create markers for marker

assisted selection which could enable the breeding of vines with increased concentrations of tartaric acid in berries at harvest, resulting in less tartaric acid being added in the winery and thereby reducing costs to winemakers. The research conducted in this chapter is a further step to understanding the isoforms of L-idonate dehydrogenase, by protein activity assays and using transgenic microvines to study the isoforms *in-planta.*

Chapter 3

Variation in malic and tartaric acid concentrations

3.1 Introduction

In grapevine berries, malic acid and tartaric acid each increase in concentration to a maximum just before veraison. Following veraison, malic acid breakdown occurs as sugar concentrations increase, while tartaric acid amounts remain relatively stable (Iland & Coombe 1988). Malic acid is synthesised and stored until veraison and is thought to participate in central metabolism after veraison as sugar concentrations increase (Sweetman et al. 2009). However, tartaric acid is not involved in central metabolism and the reason for the synthesis of tartaric acid is unknown. After veraison as sugar increases and malic acid decreases the fruit becomes much more edible and aids the dispersal of the now developed seed.

The concentrations of tartaric acid and malic acid are known to vary in many different cultivars (Kliewer et al. 1967, Soyer et al. 2003, Liu et al. 2006, Sabir et al. 2010, Munoz-Robredo et al. 2011, Pavlousek & Kumsta 2011, Preiner et al. 2013). A number of factors influence the variation in acid concentration, including environmental and genetic effects. There has been some research into the effects of climate on acid concentration, especially concerning malic acid, as it is known that malic acid breakdown increases when temperatures are high (Kliewer et al. 1967, Kliewer 1971, Ruffner et al. 1976, Sadras et al. 2013, Sweetman et al. 2014). Tartaric acid concentration does not seem to be affected by climatic conditions (van Leeuwen et al. 2004).

Potassium and pH in grapevine berries and juice

Potassium is the most abundant cation in grapevine berries (Davies et al. 2006). Potassium concentrations increase in growing tissues such as new leaves and berries as the growing season progresses. It is known that potassium is an essential nutrient for berry growth, but

the exact role it plays in berry development is unknown. Potassium may help in enzyme activation, cell membrane transport, maintenance of membrane potential and osmotic potential (Leigh & Wyn Jones 1984, Maathuis & Sanders 1996, Patrick et al. 2001, Mpelasoka et al. 2003). Potassium in the soil is taken up by plant roots and is then transported across cell membranes by ion channels and potassium transporters (Very & Sentenac 2003). The phloem is the major transporter of potassium to growing leaves and fruits and it can then be re-allocated to newer tissues as the plant grows (Mpelasoka et al. 2003). There are many factors which affect the concentration of potassium in grapevines including soil potassium concentration, cultivar, rootstock cultivar and rootstock/scion interactions, berry growth and seed number, canopy microclimate, canopy management, addition of fertilizers and irrigation (Mpelasoka et al. 2003).

In juice with pH of about 3.5 tartaric acid will contribute three times the number of protons as malic acid due to the pKas of the two acids (Ford 2012). Therefore tartaric acid is a larger contributor to the pH of wines than malic acid. In berries the pH decreases before veraison as tartaric and malic accumulate and then increases after veraison as malic acid is broken down (Johnson & Nagel 1976, Hrazdina et al. 1984). In wines, tartaric acid and potassium can form potassium bitartrate which is insoluble. If this occurs when the pH is below 3.56 the pH continues to drop because the equilibrium is pushed towards releasing more hydrogen ions. If the pH is above 3.65 when potassium bitartrate forms, hydrogen ions are lost from solution and pH then increases. In wines this can be further complicated by the alcohol concentration which can increase pH (Iland 2011).

Factors influencing tartaric and malic acid in berries

Genetic factors

The genetic factors influencing the variation in malic acid and tartaric acid seen in different cultivars are not as well studied. In a recent study, QTL analysis was used to look for genes involved in malic and tartaric acid metabolism (Chen et al. 2015). Several QTLs for malic acid were found and a list of genes within these areas was compiled. However, no QTLs for tartaric acid were found. In the most stable QTL for malic acid the study found ten genes which needed further investigation including four genes involved in transport, one involved in the TCA cycle and one involved in glycolysis. In melons, a QTL study found no co-localisation of QTLs for malic acid matched any of the major genes thought to play a role in malic acid metabolism (Cohen et al. 2012). A GWAS study in tomato found two SNPs linked to malic acid content which co-localized with an aluminum activated malate transporter like gene (Sauvage et al. 2014).

Factors that affect potassium concentration and pH

Factors affecting pH

Since tartaric acid and malic acid are the biggest contributors to pH in the berry, berry pH is affected by the same factors that influence these acids over berry development. pH is not affected by sun exposure (Crippen & Morrison 1986) but is affected by temperature (Sadras et al. 2013), with pH concentrations increasing in heated samples. This most likely due to the effect of high temperature on malic acid degradation. Van Leeuwen (2004) found that the two factors most likely to influence the percentage variance of pH values across different vines were cultivar and vintage, showing that like acid concentration, berry pH is influenced the most by environment and genotype.

Factors affecting potassium concentrations

Potassium and tartaric acid can form potassium bitartrate which is an insoluble salt and precipitates out of solution. In berries it is thought that potassium and tartrate are separate and any crystals formed are calcium oxalate monohydrate (Ruffner 1982, Webb et al. 1995, DeBolt et al. 2004). However, when grapes are crushed tartaric acid and potassium can mix and form potassium bitartrate which becomes increasingly insoluble as the concentration of alcohol increases during fermentation. Tartaric acid is added to wines in Australia to lower pH during winemaking. However, this is less useful if there is a high concentration of potassium in the juice as some of the added tartrate will form potassium bitartrate and precipitate (Mpelasoka et al. 2003). The interplay of acids and potassium suggests that a detailed understanding of berry potassium concentrations will be an important part of any future attempts to modify the levels of acid within grapes.

Variation of tartaric and malic acid in individual berries from the same vine

Very little is known about the variation in tartaric acid and malic acid in individual berries from the same vine. Many studies have been conducted into variability in a variety of berry factors including berry size, total soluble solids and ripening rate (Pagay & Cheng 2010, Dai et al. 2011, Gouthu et al. 2014). It is thought that since these factors are variable between individual berries, acid concentrations may vary as well. It has been found that the variation in ripening rate between individual berries in a bunch is highest around veraison and this variation becomes smaller towards the end of ripening (Gouthu et al. 2014). This may also affect the malic acid concentrations over development, such that berries may have more varied acid concentrations during ripening than at the end of ripening. By collecting sample berries and measuring the acid concentrations at intervals over development, the point where there is the least variation in acid concentrations could

be determined. It would be at this point that the whole berry sample would be harvested. For commercial wineries, which are looking for the most uniform fruit to make the best quality wines, this type of pre-harvest analysis might be useful.

Measurement of variability in tartaric and malic acid amounts in individual berries from the same vine and different vines

To measure variability in acid concentrations between individual berries from the same vine as well as between samples of berries from different vines, a high-throughput and accurate method to measure acid concentrations needs to be developed. This method needs to be able to accurately measure the acid concentration in batches of hundreds of samples, which means the samples need to be stably stored while the rest of the samples are prepared.

Tartaric acid and malic acid concentrations can be measured in a variety of ways. There are many different High Performance Liquid Chromatography (HPLC) techniques for measuring tartaric acid and malic acid such as reversed phase HPLC (Llorente et al. 1991, Melino et al. 2009, Zheng et al. 2009) and ion-exclusion HPLC (Frayne 1986, Schneider et al. 1987, Chinnici et al. 2005). These methods require multiple detectors (Frayne 1986, Chinnici et al. 2005), a double column (Frayne 1986), a pre-column separation step to separate acids from the sugars or phenolic compounds (Schneider et al. 1987, Chinnici et al. 2005), or a long run time of over 20 minutes per sample (Frayne 1986, Chinnici et al. 2005). Some methods require the direct injection of freshly squeezed grape juice, meaning that sample preparation must be performed immediately before running the samples; this is not practical when large batches are to be measured (Frayne 1986, Llorente et al. 1991, Zheng et al. 2009).

UHPLC-MS/MS is another way to measure the organic acids in juice and wines. This method does not require the separation of sugars and acid before the UHPLC-MS/MS run as the mass spectrophotometer detector is targeting malic and tartaric acid only. However, these methods have not been optimised for high-throughput applications. They use direct sample injection or samples mixed with water, which means that samples must be prepared immediately before detection. Furthermore, the run times are longer than 20 minutes per sample (Ehling & Cole 2011, Flores et al. 2012) which is unsuitable for high-throughput applications. A recently published method (Higginson et al. 2015) measures tartaric acid and malic acid in berry extracts with a 5 minute per sample run time and is suitable for the high-throughput measurement of hundreds of samples in one batch. The use of an extraction protocol in *ortho*-phosphoric acid allows for the freezing of samples while maintaining all tartaric acid in a measurable, soluble state because of the low pH. This method uses a 5 cm LC column for short sample run time and uses deuterated malic acid and tartaric acid internal standards for the accurate measurement of the two acids. It was used to measure tartaric acid and malic acid in single berries (as reported in section 3.3.1) and for the analysis of samples of multiple berries (see section 3.3.2). The method is detailed in the publication Higginson *et al.* (2015) which is included in the supplementary materials of this thesis.

Measurement of potassium and pH

When measuring the concentrations of malic and tartaric acid in grapevine berries it is also important to consider other factors that are closely related such as potassium concentration and pH. In this study, pH was measured using a portable pH meter and potassium was measured using Inductively Coupled Plasma Mass Spectrometry (ICP) analysis. pH and potassium concentration were both measured from the same ground powder used for the

acid concentration measurements in the four progeny populations over the 2012-2013 and 2013-2014 seasons.

3.2 Materials and methods

The UHPCL-MS/MS method was used to measure malic acid and tartaric acid in individual berry samples from the one vine to help answer the question “Is there variation in individual berries and, if so, how does this affect the sampling of subsets of berries for acid measurement?”. Ten individual berry samples from three bunches from two vines were collected and the acids were measured (collected randomly from across the bunch). This data was then used to estimate the standard error in tartaric acid and malic acid measurements when collecting a random sample of *n* berries from a vine. This information is useful for designing studies, such as developmental studies, where subsets of berries need to be collected from vines.

Berry sampling

Berry samples for tartaric acid and malic acid measurement were collected from individual vines belonging to ten populations of vines. In the 2011-2012 season, samples were taken from a subset of vines from each of the populations to determine the range of variation in each population at two time points, a pre-veraison point and a mature berry point. From this data four populations were chosen and were sampled in the two subsequent seasons, the 2012-2013 and 2013-2014 seasons. In Chapter 4, this data from progeny populations was used in a GWAS and QTL analysis.

Grape berries sampled in this study were collected from vines growing in the SARDI Nuriootpa Agricultural Research Centre in South Australia. All berries came from progeny populations grown as part of a breeding program. All populations had the same female parent and a different male parent.

Berries for bunch variation analysis

Berries were collected from two vines on the 13th of February 2013 at 14 weeks post anthesis. Three bunches of berries from each vine were randomly chosen for sampling and ten berries were randomly picked from each bunch. These two vines were individuals collected from progeny population 10. All berries were frozen at -40 ^{o}C after picking.

Berries from progeny populations

Sampling time points

Pre-veraison time point: hard berries were picked from bunches where a few berries had begun to soften.

Mature berry time point: All samples from one population were collected on the same day to decrease any variations in malic acid caused by temperature changes. Berries from each vine in the population were collected when the majority of individuals in that population had Brix measurements of 22^{o} or higher.

In the 2011-2012 season, five berries from each vine were collected at both the pre-veraison and mature berry time point. Berries were collected from a subset of between 12 and 18 vines (plus parents) from ten populations. The pre-veraison samples were collected on either the 19th of January 2012 or the 25th of January, 9 weeks or 10 weeks from anthesis respectively. The mature berry samples were collected either on the 2nd of March 2012, the 7th of March 2012, the 8th of March 2012 or the 15th of March 2012 with these dates representing 16 weeks, 17 weeks or 18 weeks after anthesis. The four populations chosen to be sampled in subsequent years were populations 3, 6, 9 and 10.

In the 2012-2013 season, ten berries were collected from individuals of the four chosen

populations at the pre-veraison and mature berry time points. Ten berries were picked instead of five because it proved difficult to grind just five berries to a fine powder in the available IKA mill due to the small amount of space they took up in the grinder cup. Pre-veraison samples were collected between the 16th of January 2013 and the 30th of January 2013 (10 to 12 weeks after anthesis). The mature berry samples were collected between the 12th of February 2013 and the 27th of February 2013 (14 to 16 weeks after anthesis).

In the 2013-2014 season, a 30 berry sample was collected from each vine of the same four populations collected at the two time points. The pre-veraison samples were collected between the 23rd of January 2014 and the 4th of February 2014 (10 to 12 weeks after anthesis). The mature berry samples were collected on the 13th of February and the 20th of February (13 to 14 weeks after anthesis). The sample of 30 berries was chosen based on the single berry data presented in section 3.3.2.

Berry sampling for pH and potassium analysis

See above section "Berries from progeny populations". The same prepared ground, frozen samples were used for potassium and pH measurement. Only samples from population 10 at the mature berry time point in the 2013-2014 season were used for potassium concentration measurement.

Preparation of berry samples

Acid extraction protocol

Berries for both the bunch variation experiments and the measurement of tartaric and malic acid from progeny populations were ground and extracted in the same way. Please see Higginson *et al*. (2015) for details of the extraction protocol (this paper is included in

the thesis supplementary materials). Individual berries or samples of berries were ground with an IKA A11 basic mill (Staufen, Germany) in liquid nitrogen. Grinding to a fine powder was done with 2-3 five second pulses in the mill.

In the 2013-2014 season, when 30 berries were collected, an upscaled version of the acid extraction was used, due to the increase the total amount of powder from each sample. Instead of 0.3g, 3 g of powder was added to 12 ml of 0.5 M *ortho*-phosphoric acid and incubated with shaking at room temperature for one hour. This extraction was centrifuged and 1 ml of the supernatant was filtered through a 45 µm PVDF membrane. All extracts were stored at -20 ºC.

Extraction method validation

To test if there was any acid left in the pellet after the extraction process, the pellets of four previously extracted, pre-veraison samples were chosen randomly for a secondary *ortho*-phosphoric exaction. Samples were extracted following the method listed in the above section and the pellet was retained and subjected to a second extraction process. Malic and tartaric acid were measured in these second extractions. Malic acid ranged between 0.01 and 0.02 mg/ml and tartaric ranged from 0.07 to 0.10 mg/ml. This concentration was between 120 and 220 times smaller than the original extraction for malic acid concentration and 25 to 40 times smaller for the original tartaric acid extraction concentration. This showed that the extraction process could remove most of the tartaric and malic acid from the starting tissue. Please see Higginson *et al.* (2015) for more information.

Tests were also conducted to determine the reproducibility of sampling from a larger amount of powder to create the extraction. Five extractions were created from a pre-

veraison sample and a mature berry sample and malic and tartaric acid were measured using the UHPLC-MS/MS method. The standard deviation of the five pre-veraison extractions was found to be 0.26 mg/berry for malic acid and 0.10 mg/berry for tartaric acid. For the mature berry extractions the standard deviation was found to be 0.02 mg/berry for malic acid and 0.07 mg/berry for tartaric acid. The standard deviations are all below 5 % of the mean acid sample and indicate that the extraction method was reproducible. Please see Higginson *et al.* (2015) for more details.

HPLC method for determining organic acids in berry samples

Two methods for measuring the malic acid and tartaric acid in berry samples by HPLC were trialled. Each method used a different HPLC column for the separation of the acids.

Method one:

Column: Phenomenex Kinetex C18 (2.6 u, 150 mm x 4.6 mm) maintained at 40 ^{o}C.

A sample of 20 µl was injected onto the column. A mobile phase 10 mM KH_2PO_4 with a flow rate of 0.5 ml/min for the first 5 minutes and 1.6 ml/min for the final 15 minutes. Acids were detected with a photodiode array detector at 210 nm.

Method two:

Column: Phenomenex Synergi fusion column (150 mm X 4.6 mm) maintained at 40 ^{o}C. A sample of 5 µl was injected onto the column.

A gradient was used with a mobile phase of 20 mM KH_2PO_4 (solvent A) and methanol (solvent B). The gradient used was as follows: 0-10 minutes, up to 10% B; 10-14 minutes, down to 0% B; 14-20 minutes, wash 100 % A.

Separation of sugars and organic acids with SPE columns

Sugars can co-elute with malic and tartaric acid during HPLC (Frayne 1986). Several methods can be employed to separate these sugars from the acids including pre-treatment using disposable anion exchange columns. Strata X-AW weak anion exchange columns (Phenomenex) and Strata X-A strong anion exchange columns (Phenomenex) were trialled. These columns add a step to the process of the HPLC method as the samples are run through these columns first and acids and sugars can be eluted off the column separately by changing the conditions of the elution step, effectively separating organic acids from sugars. Both columns used the same method the only difference was the amount of acid that could be bound to the column. The pH of the acid extraction needed to be between 6.3 and 7.2. This was achieved by adding 1M sodium hydroxide and checking the pH regularly with pH strips. The column was conditioned with 0.5 ml of methanol then equilibrated with 100 mM of KH_2PO_4 at a pH of between 6.3 and 7.2. Sample was loaded and column was washed with 100mM of KH_2PO_4 at a pH of between 6.3 and 7.2 and then 0.5 ml of methanol. The column was dried with a vacuum and the acids eluted with 0.5 ml of 5 % formic acid in methanol or 2 % acetic acid in methanol. As methanol was incompatible with the 10 mM KH_2PO_4 mobile phase, the elution was then vacuum dried and re-suspended in 0.5 ml of 0.5M *ortho*-phosphoric acid. No acids eluted off these columns so Strata X-A strong anion exchange columns (Phenomenex) were trialled with the same method listed for Strata X-AW columns.

Dilution of samples for UHPLC-MS/MS run

Internal standard

d_3-2,3,3-Malic acid was purchased from Sigma-Aldrich (Milwaukee, WI). d_2-2,3-Tartaric acid was purchased from CDN isotopes (Pointe-Claire, Quebec, Canada). These two deuterated acids were mixed together in equal amounts to create an internal standard

which was added to each sample to a final concentration of 1 mg/L.

Variation in individual berries

The individual berry samples were thawed and diluted 1:100 in sterile water. A 1 ml aliquot of this dilution was put into a HPLC vial and the internal standard mix was added to a final concentration of 1 mg/L.

Samples from progeny populations

For the measurement of acid in progeny populations pre-veraison samples were diluted 1:200 and mature berry samples were diluted 1:100 in sterile water. A 1 ml aliquot of the dilutions was put into a HPLC vial and the deuterated standard was added to a final concentration of 1 mg/L. In the 2012-2013 and 2013-2014 seasons a pipetting robot was employed to dilute all samples and add the internal standard. Pre-veraison samples were diluted 1:200 and mature berry samples 1:100 but into a much smaller volume of 100 µl in a 96 well plate; 20 µl of the deuterated internal standard was then added to a final concentration of 1 mg/L.

UHPLC-MS/MS

The development of the UHPLC-MS/MS method for measuring malic acid and tartaric acid was done in collaboration with Dr. Natoyia Lloyd of the Australian Wine Research Institute. See Higginson *et al*. (2015) for details on method for LC-MS/MS (this paper is included in the thesis supplementary materials).

UHPLC-MS/MS provided the concentration of tartaric acid and malic acid in g/L. For the purpose of comparison between vines these concentrations were converted to mg/berry.

This was achieved by converting the concentration in g/L to mg/g of ground powder and then multiplying the mg/g by the weight of a single berry.

Calculations of optimum sample size

The statistical software package GenStat V.15 (Payne et al. 2013) was used to estimate the variance components in each acid individually.

The total variation (σ^2) in tartaric or malic acid within a single vine consists of the variation between individual berries, berries within the same bunch and the variation among bunches (Eq. 1):

(1)

$$\sigma^2 = \sigma^2 berries + \sigma^2 bunches$$

The standard error of the sample mean in such sample is calculated as in Eq. 2. This assumes that a balanced cluster sample of *n* berries was taken from a single vine by choosing at random *b* bunches on the vine and then choosing at random *c* berries in each bunch.

(2)

$$SE(mean) = \sqrt{\frac{1}{n}\left[\frac{(\sigma berries^2 + c\,\sigma bunches^2)(b-1) + \sigma berries^2 b(c-1)}{cb-1}\right]}$$

Our goal in estimating the variance components was to recommend the best combination of *b* and *c* to achieve a desirable accuracy of the estimation of the mean.

pH measurements

A small sample of powder (0.5 g) was transferred to a 1.5 ml tube. The samples were

thawed and spun at 12,000 X g for 5 mins in a centrifuge to pellet the solid particles. The liquid was used for pH analysis by pipetting approximately 100 - 200 µl onto the Horiba 212 pH meter (Horiba, Japan) which was calibrated before use with standard buffers of pH 7 and pH 4 following the manufacturer's instructions.

Potassium analysis

The method of sample preparation was derived from Zacinas *et al.* (1987) and Davies *et al*. (2006). 100 mg of ground powder was added to a glass test tube. 2 ml of concentrated nitric acid was added and incubated overnight at room temperature in a fume hood. Samples were then incubated for two hours at 60 ^{o}C, then at 70 ^{o}C for one hour and finally at 125 ^{o}C for four hours to digest all solid particles in the samples. The samples were left to cool overnight. The samples were then diluted 1:10 to a total of 20 ml with milli-Q water and filtered through a no. 42 Whatman paper (GE healthcare) to remove any remaining solids. The potassium concentration was measured via ICP analysis which was performed by CSIRO Analytical Services Unit (Adelaide).

3.3 Results

3.3.1 Variation in malic and tartaric acid in individual berries

Ten single berry samples were collected from three bunches from two vines from the same progeny populations at 14 weeks after anthesis. These berries were ground into a fine powder individually and extracted in *ortho*-phosphoric acid. Malic and tartaric acid in these extracts were measured using an UHPLC-MS/MS method developed by the AWRI (Higginson et al. 2015).

Variation was seen in both tartaric and malic acid between berries from the same bunch (Figure 3.1). The smallest range in tartaric acid was between 5.88 mg/berry and 10.20 mg/berry for vine 1 bunch 2 and the largest was between 8.29 mg/berry and 17.70 mg/berry for vine 2 bunch 3. For malic acid the smallest range for a bunch was between 0.28 mg/berry and 1.67 mg/berry for vine 1 bunch 2 and the largest was between 0.82 mg/berry and 3.92 mg/berry for vine 2 bunch 1.

As well as variation between the berries within one bunch there was also variation between the mean bunch acid amounts of the three bunches for each vine (Figure 3.1). These range from 0.76 to 1.77 mg/berry and 2.27 to 2.56 mg/berry for malic acid for vine 1 and vine 2 respectively. Tartaric acid amount ranged from 7.25 to 7.63 mg/berry for vine 1 and from 10.25 to 16.30 mg/berry for vine 2. Vine 1 also showed a much smaller range of variation in mean bunch acid amount than vine 2.

A comparison of the mean of the total berry sample (30 berries) showed that there was a difference in the means for both vines. Vine 1 had a tartaric acid amount of 7.47 mg/berry

and a malic acid concentration of 1.78 mg/berry. Vine 2 had a tartaric acid amount of 13.27 mg/berry and a malic acid amount of 2.42 mg/berry.

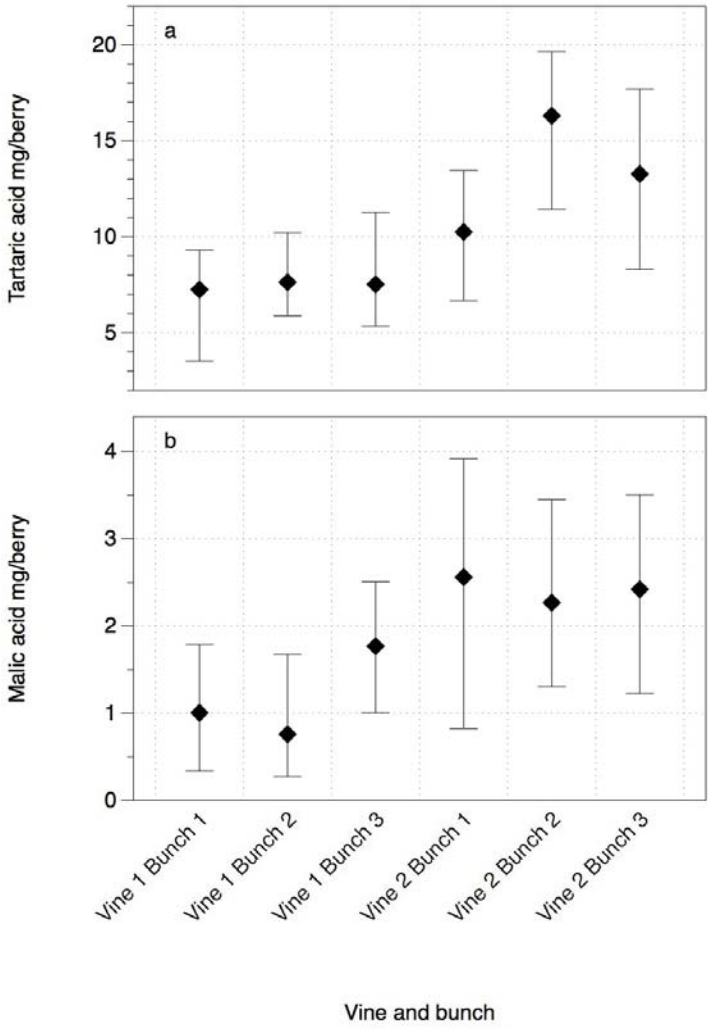


Figure 3.1: a) Tartaric acid and b) malic acid amounts from individual berries from the same vine. Ten berries were collected from three bunches from two vines. Central diamond indicates the mean acid amount for the ten berries from that bunch and whiskers indicate the range of acid for the bunch.

The variance components and the sources of variation were calculated for the data above (Table 3.1). There were three sources of variation: the individual berries, the bunches from each vine and the overall vine itself. The vines contributed the most to variation but this was expected as it is well documented that different cultivars have different amounts of acid, and these two vines are siblings from a progeny population. For tartaric acid, individual berries contributed 40 % of the mean acid concentration and bunches contributed 48 % of mean acid concentration to variation. For malic acid, individual berries contributed about 27 % of total mean acid to variation and bunches 6 % of total mean acid variation.

Table 3.1: Estimated sources of variation in a random sample of berries. The variance components and standard error (SE) in mg/berry were estimated from the concentration of acids in single berries. Ten individual berry samples were taken from three bunches from two vines (different genotypes).

Source of variation	Variance components and (SE) of tartaric acid	Variance components and (SE), of malic acid
Vines	15.3 (23.88)†	0.72 (1.09)
Bunches within vines	4.10 (3.25) [40%]‡	0.10 (0.11) [6%]
Berries within a bunch	5.00 (0.96) [48%]	0.48 (0.09) [27%]

† standard error. ‡ Percentage of the variance compared to the overall mean (Overall tartaric acid mean 10.37 mg/berry and overall malic acid mean 1.80 mg/berry).

An estimation of the percentage standard error (SE) of the overall mean acid arising from different samples sizes was made (Figure 3.2) using the equations listed in the materials and methods. As expected, as the sample size increases the estimation of the true mean of acid amount becomes more accurate. For a sample of 15 berries the SE is 10.6 % of the overall mean of malic acid and 6.9% of the overall mean of tartaric acid. For a sample of 50 berries the SE was 5.8% of the overall mean for malic acid and for tartaric acid the SE

is 3.8% of the overall mean (Figure 3.2).

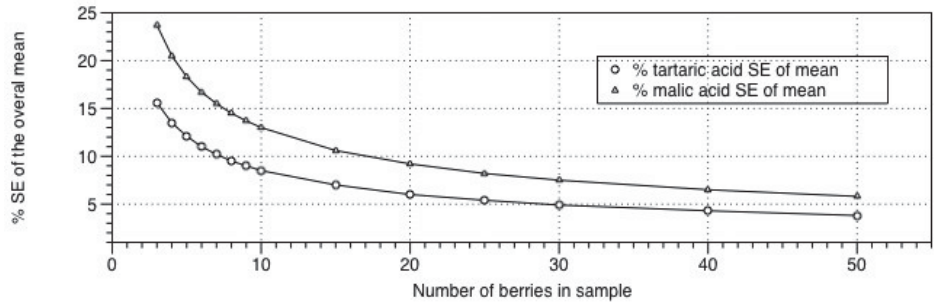


Figure 3.2: Estimated percentage standard error (SE) of measurement of tartaric acid (circle) and malic acid (triangle) from a sample of n berries collected randomly from a vine.

3.3.2 Variation of malic and tartaric acid from vines belonging to progeny populations

Progeny populations are populations of vines created from crossing two parental vines and then collecting and propagating the seeds that result from the cross into progeny vines. At the SARDI Nuriootpa research station ten progeny populations consisting of approximately 1200 vines have been planted for use in genetic studies. The populations contained between 50 and 100 individuals depending on the number of germinated seeds from each cross and number of planted vines that survived to produce fruit.

In the 2011-2012 season, berries from a subset of 12 to 18 vines were sampled from each population. These berries were used to create acid extracts from which the malic and tartaric acids were measured by the UHPLC-MS/MS method presented in Higginson *et al.* (2015). *A*ll the populations exhibited a wide range of acid amounts. Malic acid and tartaric acid vary considerably both within a population and between populations (Tables 3.2 and 3.3). The highest pre-veraison malic acid amount was measured in an individual from population 6 at 23.99 mg/berry. The lowest pre-veraison malic acid concentration was measured in an individual from population 4 at 2.79 mg/berry. The highest mature berry malic acid concentration was measured in an individual of population 10 at 8.14 mg/berry. The lowest mature berry malic acid concentration was measured in an individual from population 8 with a concentration of 0.10 mg/berry. The highest pre-veraison tartaric acid concentration was measured in an individual from population 1 with 27.86 mg/berry and the lowest pre-veraison tartaric acid concentration was in an individual of population 10 at 4.73 mg/berry. The highest mature berry tartaric acid concentration was measured in an individual of population 9 at 20.14 mg/berry. The lowest mature berry tartaric acid concentration was measured in an individual from population 8 at 3.19 mg/berry.

The population with the highest mean for pre-veraison malic acid was population 6 with 14.23 mg/berry and the population with the lowest mean was population 8 with 7.66 mg/berry. The population with the highest mean for mature berry malic acid was population 9 with 2.45 mg/berry and the population with the lowest mean was population 8 with 1.27 mg/berry. The population with the highest pre-veraison tartaric acid concentration was population 3 with 11.13 mg/berry and the lowest mean was measured in population 7 with 8.10 mg/berry. The population with the highest mean of mature berry tartaric acid concentration was population 9 with 12.19 mg/berry and the lowest was population 5 with 7.91 mg/berry.

Table 3.2: Pre-veraison berry and mature berry malic acid concentration of subset of individuals from ten populations for the 2011-2012 season. Female parent malic acid concentration in pre-veraison berries was 10.14 mg/berry and in mature berries was 2.26 mg/berry.

	Pre-veraison				Mature		
cross	Range mg/berry	Mean mg/berry	Male parent mg/berry		Range mg/berry	Mean mg/berry	Male parent mg/berry
1	3.32-15.34	8.67	8.16		0.66-3.45	1.76	5.58
2	6.38-19.24	10.97	8.76		0.69-3.32	1.88	1.42
3	5.17-19.81	12.05	8.56		0.53-6.03	1.58	4.60
4	2.79-15.99	9.49	-		0.68-3.21	1.54	3.43
5	4.59-16.68	10.27	12.69		0.49-2.93	1.49	6.51
6	8.96-23.99	14.28	20.14		1.31-3.23	2.06	5.89
7	6.61-19.84	11.58	11.15		0.70-3.35	1.82	1.57
8	4.04-14.19	7.66	5.56		0.10-3.95	1.27	5.87
9	4.08-14.25	9.63	21.45		0.26-7.81	2.45	14.56
10	4.70-17.65	8.43	26.27		0.27-8.14	2.03	9.14

Table 3.3: Pre-veraison berry and mature berry tartaric acid concentration of subset of individuals from ten populations for the 2011-2012 season. Female parent tartaric acid concentration in pre-veraison berries was 7.24 mg/berry and in mature berries was 8.29 mg/berry.

	Pre-veraison				Mature		
cross	Range mg/berry	Mean mg/berry	Male parent mg/berry		Range mg/berry	Mean mg/berry	Male parent mg/berry
1	5.64-27.86	10.81	4.41		5.84-16.67	10.12	9.50
2	7.65-14.62	10.15	10.34		4.66-12.81	9.46	12.67
3	5.36-17.39	11.13	9.86		5.23-15.72	9.40	10.79
4	6.78-14.09	9.81	-		5.56-13.87	9.31	13.68
5	6.34-19.24	10.45	10.83		5.31-11.87	7.91	12.16
6	7.12-14.39	10.38	11.16		7.29-11.87	9.00	8.17
7	5.31-11.13	8.10	7.11		4.63-13.84	8.02	6.39
8	5.90-11.06	8.17	4.91		3.19-18.61	8.67	8.99
9	6.84-15.07	10.07	10.69		6.23-20.14	12.19	27.57
10	4.73-14.19	8.15	10.88		4.82-18.42	10.60	13.03

From the 2011-2012 season data, four populations were chosen to sample in full (every individual) at the two time points. These populations were chosen based on the malic and tartaric acid amounts and the number of individuals in the populations. The selected populations had wide variation in acid amounts across all the individuals. They also had relatively large numbers of individuals. Population 3 was the smallest with 53 individuals and population 10 was the largest with 94 individuals. Population 6 contained 71 individuals and population 9 contained 63 individuals.

The five berries collected at each time point in the 2011-2012 season proved too few to grind easily into a fine power in liquid nitrogen so in the 2012-2013 season it was decided that ten berries should be collected from each time point. Again malic acid and tartaric acid were measured with UHPLC-MS/MS. In this year each individual from the four

populations was measured to get a clearer picture of the variation in acid concentration in these populations (Tables 3.4 and 3.5). The highest pre-veraison malic acid concentration was measured in an individual from population 3 at 14.59 mg/berry. The lowest pre-veraison malic acid concentration was measured in an individual from population 10 at 0.73 mg/berry. The highest mature berry malic acid concentration was measured in an individual of population 10 at 8.82 mg/berry. The lowest mature berry malic acid concentration was measured in an individual from population 3 with a concentration of 0.04 mg/berry. The highest pre-veraison tartaric acid concentration was measured in an individual from population 9 with 17.58 mg/berry and the lowest pre-veraison tartaric acid concentration was in an individual of population 10 at 2.38 mg/berry. The highest mature berry tartaric acid concentration was measured in an individual of population 9 at 9.61 mg/berry. The lowest mature berry tartaric acid concentration was measured in an individual from population 6 at 2.49 mg/berry.

The population with the highest mean for pre-veraison malic acid was population 9 with 5.72 mg/berry and the population with the lowest mean was population 3 with 3.99 mg/berry. The population with the highest mean for mature berry malic acid was population 10 with 2.07 mg/berry and the population with the lowest mean was population 3 with 0.63 mg/berry. The population with the highest pre-veraison tartaric acid concentration was population 9 with 8.59 mg/berry and the lowest mean was measured in population 6 with 5.60 mg/berry. The population with the highest mean of mature berry tartaric acid concentration was population 9 with 9.61 mg/berry and the lowest was population 6 with 2.94 mg/berry.

Table 3.4: Malic acid concentration from the four chosen populations for the 2012-2013 season at the pre-veraison and mature berry time points. Female parent pre-veraison malic acid concentration was 4.38 mg/berry and mature berry malic acid was 1.51 mg/berry.

	Pre-veraison				Mature berries		
Cross	Range mg/berry	Mean mg/berry	Male parent mg/berry		Range mg/berry	Mean mg/berry	Male parent mg/berry
3	0.89-14.59	3.99	1.45		0.04-2.92	0.62	0.04
6	1.66-9.82	4.75	3.54		0.12-3.34	0.96	0.32
9	2.01-14.09	5.72	7.77		0.13-5.29	1.56	3.04
10	0.73-11.99	5.32	5.02		0.25-8.82	2.07	3.47

Table 3.5: Tartaric acid concentration from the four chosen populations for the 2012-2013 season at the pre-veraison and mature berry time points. Female parent pre-veraison tartaric acid concentration was 5.80 mg/berry and mature berry malic acid was 6.12 mg/berry.

	Pre-veraison				Mature berries		
Cross	Range mg/berry	Mean mg/berry	Male parent mg/berry		Range mg/berry	Mean mg/berry	Male parent mg/berry
3	2.44-10.47	6.14	4.44		3.11-11.09	6.68	5.18
6	2.76-12.32	5.60	7.47		2.94-9.63	6.19	5.66
9	4.25-17.58	8.59	12.83		5.04-16.08	9.61	14.51
10	2.38-15.92	6.67	5.15		3.46-16.83	7.94	9.82

In the 2013-2014 season, 30 berries per vine at the two time points were collected to reduce the standard error in the measurement of acids of a sub-set of berries from one vine (see section 3.3.1 for details) (Table 3.6 and 3.7). The highest pre-veraison malic acid concentration was measured in an individual from population 6 at 20.38 mg/berry. The lowest pre-veraison malic acid concentration was measured in an individual from population 10 at 1.29 mg/berry. The highest mature berry malic acid concentration was measured in an individual of population 9 at 6.11 mg/berry. The lowest mature berry malic acid concentration was measured in an individual from population 9 with a concentration

of 0.13 mg/berry. The highest pre-veraison tartaric acid concentration was measured in an individual from population 6 with 19.82 mg/berry and the lowest pre-veraison tartaric acid concentration was in an individual from population 10 at 3.72 mg/berry. The highest mature berry tartaric acid concentration was measured in an individual from population 9 at 20.29 mg/berry. The lowest mature berry tartaric acid concentration was measured in an individual from population 6 at 3.59 mg/berry.

The population with the highest mean for pre-veraison malic acid was population 9 with 9.13 mg/berry and the population with the lowest mean was population 10 with 5.72 mg/berry. The population with the highest mean for mature berry malic acid was population 6 with 2.66 mg/berry and the population with the lowest mean was population 10 with 1.00 mg/berry. The population with the highest pre-veraison tartaric acid concentration was population 9 with 9.60 mg/berry and the lowest mean was measured in population 10 with 7.60 mg/berry. The population with the highest mean of mature berry tartaric acid concentration was population 9 with 11.72 mg/berry and the lowest was population 10 with 8.55 mg/berry.

Table 3.6: Malic acid concentration from the four chosen populations for the 2013-2014 season at the pre-veraison and mature berry time points. Female parent pre-veraison malic acid concentration was 6.07 mg/berry and mature berry malic acid was 1.04 mg/berry.

	Pre-veraison				Mature berries		
Cross	Range mg/berry	Mean mg/berry	Male parent mg/berry		Range mg/berry	Mean mg/berry	Male parent mg/berry
3	2.28-18.67	7.05	6.86		0.18-5.42	1.39	0.39
6	3.92-20.38	9.04	12.74		0.69-4.99	2.06	1.35
9	1.48-15.55	9.13	6.20		0.13-6.11	1.48	6.05
10	1.29-17.44	5.72	17.44		0.15-3.53	1.00	3.53

Table 3.7: Tartaric acid concentration from the four chosen populations for the 2013-2014 season at the pre-veraison and mature berry time points. Female parent pre-veraison tartaric acid concentration was 5.99 mg/berry and mature berry malic acid was 7.14 mg/berry.

	Pre-veraison				Mature berries		
Cross	Range mg/berry	Mean mg/berry	Male parent mg/berry		Range mg/berry	Mean mg/berry	Male parent mg/berry
3	5.35-11.82	8.39	10.89		6.61-15.43	9.86	8.83
6	3.97-19.82	8.18	9.56		3.59-18.32	8.93	9.30
9	5.54-13.89	9.60	8.06		7.30-20.29	11.72	20.29
10	3.72-14.37	7.60	13.31		5.19-17.54	8.55	13.37

3.3.3 Variation in pH and potassium concentration in berry samples from progeny populations

pH values varied within the four progeny populations, with pre-veraison samples having a smaller range in pH than the mature berry samples (Tables 3.8 to 3.11). At the pre-veraison time point, the mean pH value was 2.6 for population 3, 2.6 for population 6, 2.5 for population 9 and 2.9 for population 10 in the 2012-2013 season. In the 2013-2014 season for the pre-veraison time point, the pH value was 2.5 for population 3, 2.6 for population 6, 2.7 for population 9 and 2.7 for population 10. At the mature berry time point the mean pH value was 3.6 for population 3, 3.5 for population 6, 3.6 for population 9 and 3.5 for population 10 in the 2012-2013 season. For the 2013-2014 season mean pH was 3.4 for population 3, 3.6 for population 6, 3.6 for population 9 and 3.7 for population 10. pH was lower in pre-veraison samples than in mature berry samples as was expected.

As malic acid amounts decrease over development, while tartaric acid (mg/berry) is stable, it is assumed that loss of malic acid contributes to the increase in pH seen in the mature berry samples. pH also differed between the years in each population (Tables 3.8 to 3.11).

Potassium concentrations were measured by ICP analysis from the same ground powder as pH values and acid concentration measurements. Only the 2013-2014 mature berry samples from population 10 were measured. There was a spread of potassium concentrations across individuals of population 10 (Figure 3.4), with a mean of 2.22 mg/g of fresh weight and a range of 1.32 to 3.23 mg/g of fresh weight.

Table 3.8: pH values from individuals of four progeny populations for the 2012-2013 season at the pre-veraison time point.

pH	No. of individuals for each population			
	Population 3	Population 6	Population 9	Population 10
2.3	0	1	2	0
2.4	0	0	3	0
2.5	4	0	37	0
2.6	16	3	8	10
2.7	4	5	2	9
2.8	7	15	6	14
2.9	0	23	0	19
3.0	0	2	0	4
3.1	0	20	0	19
3.2	0	1	0	2
3.3	0	0	0	7

Table 3.9: pH values from individuals of four progeny populations for the 2013-2014 season at the pre-veraison time point.

pH	No. of individuals for each population			
	Population 3	Population 6	Population 9	Population 10
2.3	1	0	0	1
2.4	5	0	0	4
2.5	21	13	3	20
2.6	1	27	12	14
2.7	2	19	18	23
2.8	0	6	20	11
2.9	0	4	1	4
3.0	0	0	1	4
3.1	0	0	0	2
3.2	0	0	0	0
3.3	0	0	0	0
3.4	0	0	1	0

Table 3.10: pH values from individuals of four progeny populations for the 2012-2013 season at the mature berry time point.

pH	No. of individuals for each population			
	Population 3	Population 6	Population 9	Population 10
2.6	0	0	0	1
2.7	0	0	0	1
2.8	0	0	0	1
2.9	0	2	0	0
3.0	0	8	1	9
3.1	0	1	2	7
3.2	4	10	4	12
3.3	1	8	4	12
3.4	3	6	4	6
3.5	4	15	9	10
3.6	1	0	3	3
3.7	7	12	10	4
3.8	0	3	1	5
3.9	4	4	8	3
4.0	6	3	1	10
4.1	1	1	0	0
4.2	1	1	1	2
4.3	0	0	1	2
4.4	0	1	0	0

Table 3.11: pH values from individuals of four progeny populations for the 2013-2014 season at the mature berry time point.

pH	No. of individuals for each population			
	Population 3	Population 6	Population 9	Population 10
2.9	0	0	1	0
3.0	2	2	1	1
3.1	0	1	1	0
3.2	7	2	1	2
3.3	2	9	3	5
3.4	5	0	7	7
3.5	5	21	3	5
3.6	2	1	23	21
3.7	6	17	2	8
3.8	0	4	14	11
3.9	3	5	0	4
4.0	0	1	2	6
4.1	0	0	0	1
4.2	0	1	0	6
4.3	0	1	0	1
4.4	0	0	0	3

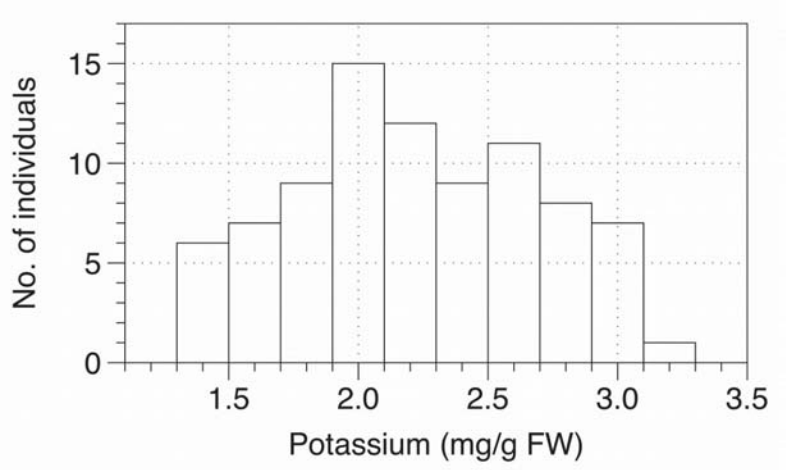


Figure 3.4: Potassium concentrations of population 10 mature berry samples in the 2013-2014 season. FW = Fresh weight.

3.4 Discussion

Measuring malic and tartaric acid in individual berries and samples of multiple berries

At the start of this study HPLC was considered as a way of measuring tartaric acid and malic acid in grape berry extracts. HPLC is the most common way to measure malic acid and tartaric acid in grape berries, wine and juice, and the equipment was readily available. When developing a method for the measurement of tartaric acid and malic acid, several factors were considered. Firstly, over the course of this project tartaric acid and malic acid would need to be measured in hundreds of samples, requiring a quick high-throughput method. Secondly, the co-elution of sugars with malic acid and tartaric acid in the mature berry samples is known to affect the accurate measurement of malic and tartaric acids in the samples.

A method using Kinetex C18 column (Phenomenex) with a KH_2PO_4 buffer was trialled to measure malic acid and tartaric acid with detection at 210 nm but there were unknown co-elutants in the peaks for mature berry samples (data not shown). Another method using a Synergi fusion column (Phenomenex) was also briefly trialled. The method for this column had a 20 minute run time and offered no chance to minimise run times (data not shown). So was considered not suitable for a high-throughput method.

It was thought that the unknown co-elutants could have been sugars so the removal of sugars was attempted before the HPLC run with a SPE column. The use of these columns was a lengthy process, and without the capability in our laboratory for large scale batch processing, would have taken many hours of processing. This was not ideal for batch sizes of over 300 samples. Furthermore, the methanol in which the acids were eluted off the SPE column was incompatible with the KH_2PO_4 buffer requiring all the methanol to be

evaporated and the samples re-suspended in KH_2PO_4 before the HPLC run, adding to the sample preparation time. The elutions from these columns also produced very poor recoveries of both tartaric and malic acid and often no acid was eluted off the column at all. Other published methods have successfully used similar columns to separate sugars from organic acids before HPLC. However, they used juice from freshly squeezed samples or commercial juices or wines with small batch sizes (Castellari et al. 2000, Chinnici et al. 2005). In our case berries were frozen after sampling and therefore the extraction (in *ortho*-phosphoic acid) process was needed to stop any tartaric acid from forming an insoluble precipitate with potassium during thawing. The use of a water extraction or fresh juice injection method was not chosen as a water extraction does not buffer the pH and so cannot be frozen without insoluble tartrate salts forming and direct injection of fresh juice could not be done because berries were stored frozen before analysis due to the large number of samples. In the future, a better method for collecting large numbers of berry samples which allowed for the same or next day use of the SPE columns, removing the need to freeze berries and the acid extraction, could be developed. Furthermore the use of these columns was time consuming, the method would also benefit from the development of a high-throughput method for streamlining the processes possible with the use of a robot with a vacuum pump capability.

To overcome the above problems a high-throughput UHPLC-MS/MS method was developed in conjunction with the AWRI to measure tartaric acid and malic acid in grape berry extracts and the publication detailing this method is attached in the supplementary materials (Higginson et al. 2015). This method had a 5 cm column for fast separation of malic and tartaric acid and deuterated internal malic and tartaric acid standards for accurate quantification. There was also the opportunity to develop a system for pipetting

robots to do most of the dilution of samples into a 96 well plate that led to faster preparation time and more accurate pipetting. Extraction of the acids in *ortho*-phosphoric acid was performed which provides a low pH and keeps all forms of tartaric acid in solution even when starting from frozen material. These extracts could also be frozen themselves while the rest of the batch was processed.

In the 2013-2014 season the extraction volumes were upscaled due to the larger amounts of ground berry material generated from 30 berries samples opposed to 5 to 10 berry samples. In this upscaling, the ratios of berry powder to *ortho*-phosphoric acid was kept similar and all other parts of the extraction method were the same for both volumes.

Causes of variability in potassium, pH and acid concentrations

In section 3.3.1, variations were seen in the tartaric acid and malic acid from the bunches (mean of the ten berry samples from that bunch), between the means of the 30 berries for each of the vines and between the individual berry measurements themselves. Variation was observed between the individuals in the four populations as well as between the two populations. There was also variation in the acid amounts between the two years. Several published studies have also measured acid concentration in populations of vines. These studies found differences between the years (Liu et al. 2007, Duchene et al. 2014, Chen et al. 2015), and between different populations (Liu et al. 2007), which are consistent with our findings.

pH varied within each of the four populations tested (Tables 3.8 to 3.11). The pre-veraison time point had a smaller range of pH concentrations than the mature berry time point. Published studies only focus on mature berry juice pH as this is most important for wine

making. Soyer *et al.* (2003) measured pH concentration in several varieties of grapevine and found that pH ranged from 2.6 to 3.9 at a mature berry time point. Kliewer *et al.* (1967) measured pH in several varieties and found it ranged from 3.06 to 4.22 collected at an early and a late mature berry stage. In this study similar results were obtained with mature berry pH in the four populations ranging from 2.6 to 4.4 in the 2012-2013 season (Table 3.10) and from 2.9 to 4.4 in the 2013-2014 season (Table 3.11). The slightly higher pH seen in this study may be due to the hot climate in which the grapes were grown which promotes the break down of malic acid and increases the pH.

The pH also differed between the two years for each of the populations (Tables 3.8 to 3.11). This is consistent with published literature. Pavlousek and Kumsta (2011) found that the means of the pH over the three years measured were significantly different at the 95 % confidence interval. Van Leeuwen *et al.* (2004) found that variance in berries was influenced more by climate and cultivar than by soil over the years studied.

Potassium concentrations were measured for individuals of population 10 at the mature berry time point in the 2013-2014 season (Figure 3.4). There was a large spread of values, ranging from 1.32 to 3.23 mg/g of fresh weight. Williams and Biscay (1991) found potassium concentrations in ripe berries berries to be 8.5 mg/g fresh weight, while Davies *et al.* (2006) found the concentration to be approximately 2.5 mg/g fresh weigh and Iland and Coombe (1988) found the concentration to be approximately 10 mg/g fresh weight. The concentrations measured from population 10 were around the values presented by Davies *et al.* (2006). There are several possible reasons that the potassium concentration was lower than the other two studies. The age of the berries that were measured, the use of ground powder in the ICP analysis and the number of berries used in the study.

The cause of the variation seen in acid in individual berries from the same vine and in potassium, pH and acid berries from vines from the progeny populations is unknown. Ripening rate and environment may influence organic acid amounts both between berries from the same vine and the amounts between different vines and potassium concentration between different vines. Added to this, variation between different cultivars, or individuals of a progeny population may also have genetic factors at play influencing acid and potassium.

It is known that ripening rates of individual berries vary, and it has been reported that berries in a bunch go through a 'syncing' of ripening where berries that are behind in ripening at veraison have a higher ripening rate after veraison so that at 50 days post mid veraison all berries have similar Brix values and colour indexes (Gouthu et al. 2014). Between veraison and 50 days post mid veraison there is much more variability in ripening rate across berries from the same bunch and this could have consequences for the malic acid concentration, potassium concentration and pH during this period.

Ripening rates across the progeny populations also differed. Veraison occurred over 11 days in one population alone. Mature berry Brix measurements, an indication of ripeness, also varied considerably between 14 and 25 for one population indicating a difference in ripening rate. As malic acid breakdown occurs over the ripening period, differences in the ripeness of berries may lead to variations in malic acid in berries of different vines, berries of the same vine and between the two years. Also potassium concentrations in berries increase from veraison until harvest (Iland & Coombe 1988, Williams & Biscay 1991, Davies et al. 2006) so differences in the ripeness of berries may lead to variations in the potassium concentrations seen between the individuals of the progeny populations.

There may also be environmental aspects to this variation. The environment is known to play a large role in the difference in acid concentrations between vines (van Leeuwen et al. 2004). Temperature has a well documented role in the breakdown of malic acid during ripening, with higher temperatures leading to decreased malic acid (Kliewer 1968, Sadras et al. 2013). It may be that the temperature of individual berries differs and leads to a difference in malic acid and pH. Berries that are exposed to the sun for longer are hotter and so malic acid breakdown is increased in these berries. Vines may also have a sunny and a shady side depending on position in the vineyard, with the result that bunches in sunny areas may have lower malic acid and pH (Kliewer & Lider 1968, Crippen & Morrison 1986, Bergqvist et al. 2001).

Tartaric acid is thought be be relatively stable with respect to the environment. Temperature is not thought to affect tartaric acid concentration or cause differences in tartaric acid concentration between fruits of different temperatures on one vine (Kliewer & Lider 1968). Studies have shown tartaric acid concentrations to be higher in sun exposed berries (DeBolt et al. 2008, Melino et al. 2011), lower in sun exposed berries (Pereira et al. 2006) or that there is no difference in tartaric acid concentrations between sun exposed berries and shaded berries (Crippen & Morrison 1986). It may be that light intensity does play a role in the difference in tartaric acid concentrations seen between berries from the same vine but more research needs to be conducted in this area before a definitive answer can be reached.

Potassium uptake in grapevines may be affected by root temperature. Zelleke and Kliewer (1985) found that there was a small difference in potassium concentration in petioles and

roots with higher concentrations at 25 ºC compared to 12 ºC. However, these differences were not always significant. Light intensity did not appear to affect the concentration of potassium in berries (Crippen & Morrison 1986).

The 2012-2013 and 2013-2014 seasons experienced quite different weather conditions and this undoubtedly affected the concentration of malic acid in berries as differences can be seen in each individual between the two years. 2014 had the warmest January of all the seasons with a mean maximum temperature of 32 ºC compared to a 30 ºC average (Australian Bureau of Meteorology 2015). However, the mean minimum temperature was very similar to the average at 14.7 ºC compared to the average of 14.5 ºC. This increased heat during the day and smaller increase at night may have led to an increased loss of malic acid during ripening. Sweetman *et al.* (2014) found that content of malate in berries seemed to decrease more when day temperatures were increased to a higher degree than night temperatures.

There is thought to be a genetic component to the difference in acid amounts from vines of different cultivars and also between individuals of progeny populations. This will be explored more in chapter 4.

Measurement of tartaric and malic acids by UHPLC-MS/MS

The study presented here was a small preliminary study into individual berry variability. Due to the time constraints of berry collection and processing for UHPLC-MS/MS and the cost involved in the measurement, only 60 berries were sampled. A wider study, encompassing more vines, needs to be conducted to show if the large berry variability seen here holds true in other cultivars, and at other time points across development. This future

study, could also look at individual berries from several vines of the same cultivar from the one vineyard. However studying berries across development is difficult as the preparation time is greatly increased with larger numbers of samples, unless a way of non-destructively measuring berry acid can be developed. Non-destructive testing of berries for acid concentration measurement would be advantageous and would remove the effects of variation when measuring acid in berries across development. However, this is an ongoing field of research. Nuclear Magnetic Resonance (NMR) techniques have been explored to non-destructively measure organic acids in grape berries (Coombe & Jones 1983, Gonzalez-Caballero et al. 2011). Coombe (1992) sampled 2 µl of juice from berries using a micro-capillary inserted into the berry and sealing the wound each time. However the effects of removing juice from a berry across development are not understood. Near infra-red spectrometry has also been used as a way to non-destructively measure the malic and tartaric acid concentration in whole grape berries (Gonzalez-Caballero et al. 2011, Barnaba et al. 2014). Mid infra-red spectrometry has been used to measure tartaric and malic acid concentrations in grapes but non-destructive measurement methods have not been produced (Swanepoel et al. 2007).

Sample size analysis

From the data presented here it is clear that it is advantageous to collect the largest sample size possible. The larger the sample sizes the smaller the standard error. 30 berries was decided on as an ideal number of berries to collect as the percentage standard error relative to the overall mean was under 10 % for both malic and tartaric acids (Figure 3.2). But experimental design constraints means that such a large number of berries may not always be available, especially for developmental studies where many samples from the one vine need to be collected. Removing large numbers of berries at one time point may produce

physiological changes that affect acid amounts later on in development. This information could be used to plan future experiments by providing an estimate of the standard error of malic acid and tartaric acid measurement if sample sizes are in the range of 3 to 50 berries.

This study also showed the importance of randomly sampling across the whole vine as mean bunch acid amounts also varied (Figure 3.1). Collecting all berries from one bunch may also bias acid measurements. Sato *et al*. (2000) measured titratable acidity, soluble solid concentration and berry weight and found that the variability between individual berries from the same bunch was more than the variation between bunches of berries for the three traits. This is consistent with our study for malic acid but not for tartaric acid where the variance of the berries within a vine and the berries within a bunch were similar (Table 3.1).

A non-peer reviewed technical brief by Wolpert and Vilas (1992) explained a two-step method for estimating the yield of vineyards by collecting data from a subset of vines based on the level of standard error that is required. This method uses similar principles to the methods used here to determine the standard error if collecting subsets of berries of different sizes.

Choice of collection time

Ripening rates across each of the populations varied considerably and anthesis dates were spread across several weeks. This means that not all individuals in the population were at the same point when sampling occurred. For the pre-veraison time point, sampling occurred over a three week period. Samples were collected from each vine when a few berries had begun to soften.

For the mature berry time point, all the samples of one population were collected on the same day as malic acid breakdown is linked to temperature, and the higher the temperature the more malic acid breakdown occurs (Kliewer 1968, Sadras et al. 2013). Therefore, if samples were collected at the same Brix (an indication of maturity) on different days, any changes in acid concentration could be linked to either temperature fluctuations or genetic differences. A value of 22 ºBrix was chosen as the ideal collection time point. When the majority of individuals in the population were at 22 ºBrix, samples were collected from each vine in the population. This collection method is also has limitations as the ripening stage may have an effect on malic acid concentration. By plotting the Brix of each vine against the mature berry malic acid for each population it was hoped that any trends linking the two could be seen (data not shown). But no real trend linking the Brix and mature berry malic acid amount was found. However it is important to note that the brix measurements were not taken from the same berries as the acid measurements. A suggested improvement to the berry preparation methods could include a brix measurement before acid extraction.

Difference in pre-veraison and mature berry tartaric acid

Tartaric acid is synthesised in the berry up to veraison. After veraison there is no breakdown of tartaric acid and amounts (mg/berry) should be relatively stable. However, in our measurements there was a difference in pre-veraison and mature berry tartaric acid when measured in mg/berry. There may be several reasons why this is true. Firstly, it is very possible that tartaric acid had not quite finished accumulating when the berries were sampled at the pre-veraison time point. Since the berries were not collected from a series of time points across development for each individual, the exact point of highest tartaric

acid was unknown. Secondly, a subset of berries was sampled at each time point and may have introduced errors as acid amounts vary between individual berries from the same vine. In each year a different sized subset of berries was collected. The 2013-2014 season's acid measurements would be the most accurate based on the estimations of standard error seen in section 3.3.1. Finally, There are always small random and systematic errors introduced due to the extraction process, dilution process and the UHPLC-MS/MS measurement of acid. These errors are detailed in the publication which details the method (Higginson et al. 2015) which is provided in the supplementary materials.

Variations in tartaric and malic acids were found between individual berries from a single vine and between individual vines from progeny populations. While there are environmental factors influencing the variation in these two acids in grape berries both from the same vine and between individuals of populations, there is also a component of the variation in the two acids that can be contributed to genetics. This variation can be used for GWAS and QTL studies to discover regions of the genome that influence this variation and what genes may be involved in this variation. This will be explored further in the work described in Chapter 4.

Chapter 4

A genome wide association study and validation with QTL analysis to determine areas of the genome of *Vitis vinifera* involved in the metabolism of tartaric and malic acids

4.1 Introduction

Tartaric acid and malic acid accumulation in grape berries are known to be heritable traits. Duchene *et al.* (2014) determined that there was significant phenotypic variability in tartaric acid and malic acid in a Riesling and Gewurztraminer cross and this is thought to apply to other crosses as well. Here we use the variation seen in tartaric and malic acid data at the two time points detailed in Chapter 3 in conjunction with the filtered SNP marker set developed from the GBS of the individuals in the four populations to determine areas of the genome involved in variation of malic acid and tartaric acid using GWAS and QTL analysis.

With the development of Next Generation Sequencing (NGS) methods both GWAS and QTL analysis have become much easier to conduct in plants, especially GWAS which requires large numbers of markers to be useful (Deschamps et al. 2012). GBS is a method of NGS where whole genomes of individuals are sequenced at low sequencing depth and it can be used to find SNPs which can be employed as markers in GWAS and QTL.

GBS methods have become feasible in the past few years with the decrease in the cost of NGS (Elshire et al. 2011). GBS involves the low coverage sequencing of many individuals across the whole genome and then alignment to the reference genome and to each other to find SNPs which can then be used for a variety of applications such as GWAS, QTL mapping and genomic selection. GBS is performed by digesting the DNA of individuals of a population with restriction enzymes and then ligating barcode sequence onto the ends of each of the fragments and these sequences are pooled for sequencing (Elshire et al. 2011, He et al. 2014). The barcodes then help to identify the sequences from each individual at

the time of sequence alignment and genome re-construction. From these sequences, the SNPs are then formed into a SNPs set that can be used in downstream applications.

The use of GBS allows the creation of SNP sets containing thousands of SNPs markers and this is useful as the more markers used in the GWAS study the greater the resolution (He et al. 2014). In this case progeny populations were sequenced using GBS methods to create a set of SNP markers that could be used to detect statistical associations between the SNP markers and the heritable trait, the berry accumulation of tartaric and malic acid. Since the actual polymorphisms that may be causing the trait are unknown, statistics are used to infer the location of these polymorphisms by determining genetic markers that co-segregate with the causative trait polymorphisms (Oraguzie & Wilcox 2007).

Both QTL analysis and GWAS have different limitations in their methods. For QTL analysis, large numbers of markers may not necessarily create a better resolution of QTLs as the discovery of QTLs depends on recombination frequency of individuals. This is unlike GWAS, which relies on the phenotypic variance within a population and so a larger number of markers will increase the resolution (Deschamps et al. 2012, Korte & Farlow 2013). QTL analysis requires between 10^2 and 10^3 markers for genome coverage while GWAS requires between 10^5 and 10^9 (Oraguzie & Wilcox 2007). This means that populations in GWAS can be much smaller than those used for fine mapping QTL studies. Since GWAS relies on phenotypic variance, and phenotypic variance is dependent on how much difference in phenotypic effect exists between the two variants and their frequencies in the population, rare variants and traits with small effect sizes are sometimes not picked up (Korte & Farlow 2013). However, these problems can be alleviated with a QTL

analysis, if the right cross is chosen. GWAS is a relatively new method in grapevine genetics and has only been performed in simulations (Fodor et al. 2014). Association studies have been performed using grapevines but they have looked at the association with markers and a single gene (Emanuelli et al. 2010, Vargas et al. 2013). Emanuelli *et al.* (2010) performed a GWAS study to find SNPs in the *VvDXS* gene that associated with muscat flavor. However they did not use GBS and instead only re-sequenced their gene of interest which had already been identified from a previous QTL study (Battilana et al. 2009). Vargas *et al.* (2013) used association analysis to link SNPs in the *VvGA11* gene to quality traits in grapevine berries. Because GBS and GWAS are new techniques in grapevine a QTL analysis was performed here as well to validate the GWAS results. QTL and GWAS are complementary methods and using them together will hopefully remove some of the limitations of each method separately (Korte & Farlow 2013).

The GBS method employed by researchers at the Buckler Laboratory at Cornell University (Elshire et al. 2011) was used to sequence the individuals of the four populations. The method was developed to reduce genome complexity, decrease sample preparation time and to make it easier to sequence species with high genetic diversity. This makes it possible to sequence large numbers of individuals with a large genome in species with high genetic diversity such as the grapevine (Elshire et al. 2011). This method involved the restriction digest of samples with ApeKI followed by the ligation of barcodes with matching sticky ends. This restriction digest and ligation can be done in the same well plate as comparable buffers are used, reducing sample preparation time and complexity (Elshire et al. 2011).

The sequencing reads obtained from this method for the four progeny populations described in Chapter 3 were aligned to the grapevine genome to find SNPs. These SNPs were then analysed and the best and most useful SNPs were used to form a marker set for GWAS and QTL analysis. Many QTL studies have been performed in grapevine for a variety of traits including disease resistance (Barba et al. 2014), bunch architecture (Correa et al. 2014) and seedlessness and berry weight (Doligez et al. 2002). GWAS can find traits with relatively small populations. Atwell *et al.* (2010) looked for associations with 107 traits in 199 *Arabidopsis thaliana* inbred lines and these lines had between 100 and 200 individuals. This method is ideal for use in grapevine because it is not easy to produce large grapevine populations as they are slow growing and take up a large amount of space. GBS has only recently become feasible due to reductions in the cost of NGS and the creation of suitable methods for sequencing large and genetically diverse genomes (Elshire et al. 2011). The method presented in Elshire *et al.* (2011) uses restriction digests of DNA to create the libraries for GBS and discovers enough SNP markers to get good coverage across the whole genome for GWAS and QTL analysis, unlike previous methods which required a large investment in discovering SNPs and then genotyping them in populations (Baird et al. 2008).

4.2 Materials and Methods

Leaf collection and DNA extraction

The first few leaves below the growing tip were collected for DNA extraction from each individual in the four progeny populations that were used for the determination of tartaric acid and malic acid concentration as described in Chapter 3. Multiple DNA extractions were required to collect the amount of DNA needed for the GBS method. DNA extractions were performed by the AGRF (Adelaide) in a 96-well plate format and eluted in Tris-EDTA. Two elutions (100 µl and a 50 µl) were performed in separate plates.

Preparation of DNA for Genotyping-by-Sequencing

100 ng of DNA was run on a 1 % (w/v) agarose gel to determine the quality of the DNA extracted. Gel was made from 1 % (w/v) agarose gel with SYBR Safe DNA gel stain (Life Technologies) added to the gel for DNA visualisation. 1X TBE was used as the running buffer. The DNA ladder GeneRuler 1 kb DNA ladder (ThermoScientific) was used for base pair length analysis. Gels were run for 30-45 minutes at 100 Volts. DNA was visualised under blue light.

RNase digest was performed on any samples that were found to still contain RNA; 0.5 µl of RNase cocktail (Ambion) was added to the DNA preparation and incubated for three hours at room temperature. DNA was then re-run on a gel to that check all RNA was digested.

Ten percent of the 315 samples were digested with restriction enzyme HindIII to determine the size of DNA fragments to confirm the suitability of the DNA for the GBS process.

Restriction digest contained 1 X buffer 2 (NEB), 100 ng of DNA, 1 µl of HindIII enzyme and autoclaved milli-Q water to 20 µl. Reactions were incubated for three hours at 37 ^{o}C and then run on a 1 % (W/V) agarose gel (as above) to check that the DNA pattern was a smear indicating successful digest by HindIII. See supplementary materials for the agarose gel pictures of the DNA and restriction digests.

DNA (1.5 µg) samples were transferred to a 96-well plate and dried down using a vacuum centrifuge. DNA was sent for GBS to the Cornell University Institute of Biotechnology as per their specifications (Elshire et al. 2011).

Next generation sequencing and alignment to the genome to find SNPs (performed by the Cornell University Institute of Biotechnology)

Due to the low sequencing depth of some individuals observed in the first sequencing run all samples were sequenced a second time. All alignments and SNP calling were performed at Cornell using their Tassel-GBS pipeline (Glaubitz et al. 2014). The Tassel pipeline used by Cornell for the GBS analysis used Tassel Version 3.0.160 (August 22nd, 2013). The genome sequence used for the alignment was originally published in Jaillion *et al.* (2007). Alignment was generated with bwa aligner version 0.7.5a-r405. Unfiltered SNPs generated from the alignment were 218,365. VCF tools version 0.1.10 was used to determine depth and missingnesss of the unfiltered SNPs (Table 4.1). The unfiltered data sets were then filtered to 83,284 SNPs of good quality.

Table 4.1: Depth of sequencing and missingness in SNPs found in the four progeny populations sequenced by the genotyping-by-sequencing method. Missingness is the number of sites that have not not been sequenced.

	mean	median	standard_deviation
individual depth	23.631	23.26	8.381
site depth	23.631	18.027	22.335
individual missingness	0.133	0.119	0.1
site missingness	0.133	0.013	0.28

Selection of programs for GWAS and QTL analysis

For GWAS the program Tassel (Bradbury et al. 2007) was chosen as it was designed especially to handle the large amounts of data generated by NGS and was developed in the Buckler laboratory at Cornell University. For QTL analysis the R package Onemap (Margarido et al. 2007) was chosen to create the genetic maps and the R package qtl (Broman et al. 2003) was chosen for the QTL analysis.

Filtering of SNPs to create a good quality set for GWAS and QTL analysis

Initial GBS filtering found 83,284 SNPs present in the genomes of the 315 individuals from the four progeny populations. These SNPs were further filtered using a mixture of the developed Tassel GBS pipeline, the Tassel 4 GUI (version 4.3.2 and newer version 4.3.6) and FileMaker 12 database scripts to filter the SNP files (Figure 4.1). The grapevine has 19 chromosomes but the genome assembly also has 13 smaller fragments unmapped to a chromosome. These were renamed in sequential order for ease of use (Table 4.2).

For merging GBS data from the two sequencing runs the following procedure was done. First, any errors in the naming of individuals were fixed so that the individuals with duplicated sequence data could be merged properly. Then the separate runs of all

individuals were merged using the Tassel-GBS plugin 'MergeIdenticalTaxa'. Next, each population was separated from the others using the Tassel-GBS plugin 'includeTaxaInFile'. The file for each population was then loaded separately into the Tassel 4 GUI in the correct chromosome order (so that the resulting file was in the correct order) and all files were merged into one genome file for each population using the join function. As the calling of SNPs was done with all individuals and not separated into populations, there were a number of loci that were homozygous in an individual population. These sites were removed using the filter function with a minor allele frequency between 0.1 and 0.9.

Table 4.2: The names of the 13 chromosome fragments that are unmapped to a chromosome. Names were changed to a sequential order starting at 20 for ease of use.

Original Chromosome name	New Chromosome name
C100	C20
C300	C21
C400	C22
C500	C23
C700	C24
C900	C25
C999	C26
C1000	C27
C1100	C28
C1200	C29
C1300	C30
C1600	C31
C1700	C32
C1800	C33

The files were exported from Tassel as a hapmap file. These files were then imported into FileMaker 12. The files were filtered on a minor allele frequency of between 0.4 and 0.6 to remove any with a high error rate; all SNPs with missing data were filtered out and then the file was filtered into two files. One file contained SNPs that were heterozygous in the female parent and homozygous in the male parent (called the female file); the other contained SNPs that were heterozygous in the male parent and homozygous in the female parent (called the male file). Several individuals had sequencing errors in some loci causing an impossible genotype when compared to the parents. These wrong genotypes were replaced by "NA" and were not analysed.

Genome wide association study of pre-veraison and mature berry tartaric and malic acids.
Analysis was performed using Tassel 4 version 4.3.6 GUI. The two filtered hapmap files for each population were imported into Tassel. These files were joined to create one file and then the parents were removed. A Principal Component Analysis (PCA) was performed and plotted to determine if any individuals differed significantly from the rest of the population.

The phenotype data was also imported into Tassel and merged with the SNP data for each individual. Phenotype data consisted of acid pre-veraison and mature berry malic and tartaric acid amounts in mg/berry (see Chapter 3). This data and the PCA were used for the GWAS using a General Linear Model (GLM).

Adjustment of P values

The GLM results were exported from Tassel and prepared for P-value adjustment. This was done with the R package Multtest (Pollard et al. 2005). A P-value below 0.05 was considered significant using the method described in Benjamini and Hochberg (1995).

Validation of the results from GWAS with mapping and QTL analysis

Creation of maps

Maps were created using the filtered SNP data that was used for the GWAS (with incorrect SNPs removed) and the R package Onemap 2.0-3 (Margarido et al. 2007). Populations 3 and 9 had their two SNP files (male heterozygous and female heterozygous) joined for analysis while the files for populations 6 and 10 were kept separate due to their size. Onemap creates a two point pairwise analysis between all pairs of markers with a LOD of 3 and maximum recombination frequency of 0.5 based on the method in Wu et al. (2002). To create the linkage groups a LOD of 6 and a maximum recombination frequency of 0.25 were used. The exception was the population 6 male SNP file which would not separate into the correct number of linkage groups; in this instance, parameters were changed to a LOD of 8 and a maximum recombination frequency of 0.25. To create the map the multipoint likelihood of all orders was compared to find the best order of the markers.

QTL analysis

Once the map was created it was transferred along with the phenotype data into the R package qtl (Broman et al. 2003). To create the map, first, the probabilities of the underlying genotypes were calculated using hidden Markov model technology. Then a genome scan with one QTL model was performed using a normal model that assumed a normal distribution, and a maximum likelihood method using the EM algorithm (Dempster et al. 1977, Lander & Botstein 1989).

Creation of Manhattan plots for the GWAS and QTL analysis

Manhattan plots were created using the R package Qqman (Turner 2014).

Genes in QTL regions

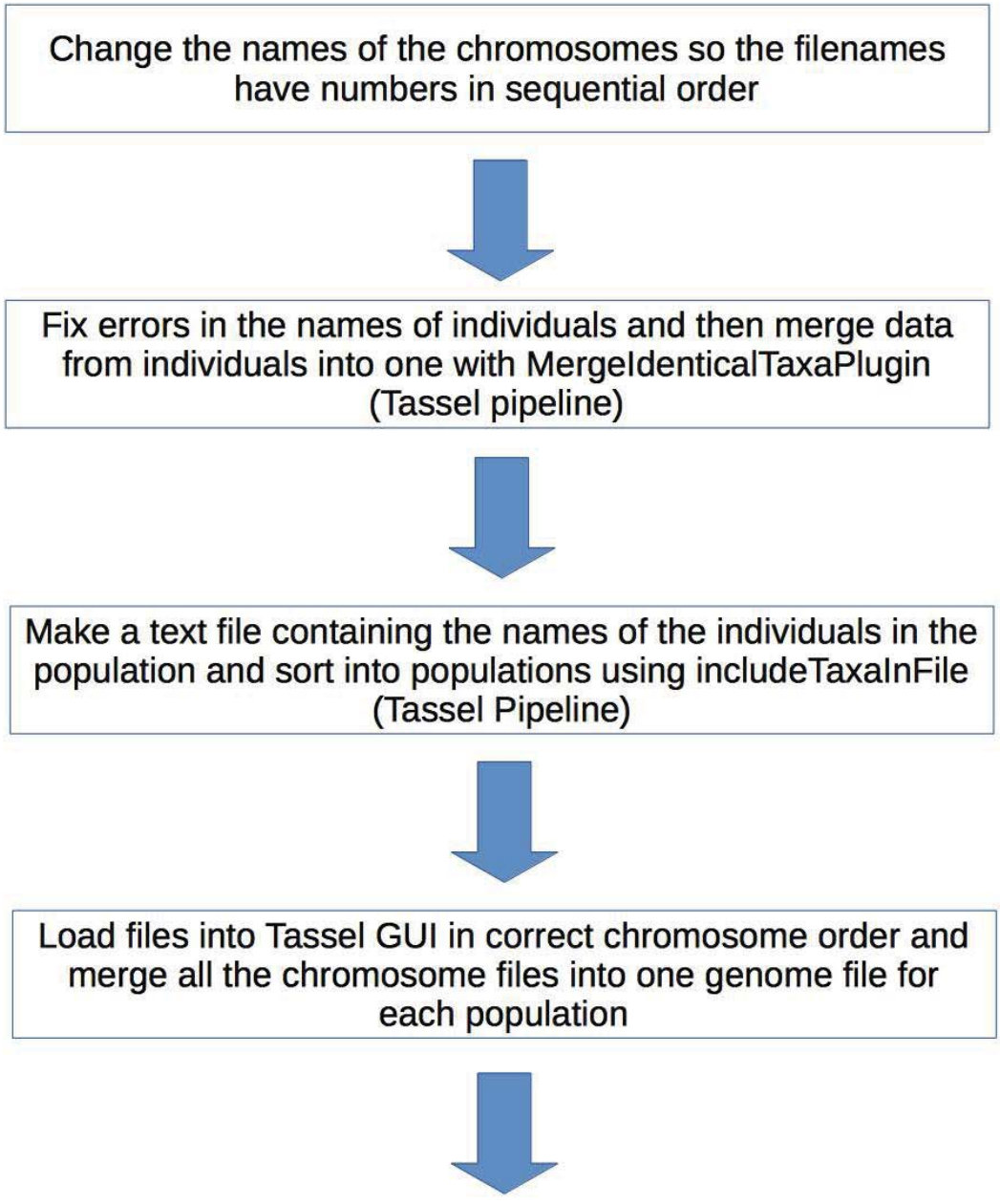
It was decided that the area between SNPs S5_487,880 and S5_2,743,133 was one QTL. This area had some of the highest P-values in both the GWAS and QTL analysis. While this region did vary between QTL analysis and the GWAS analysis, and between pre-veraison and mature berry tartaric acid these SNPs seemed to be consistently of a significant P-value. The DNA sequence in this area from the assembly used by Cornell was blasted against the NCBI assembly of the genome to determine if their regions differed. It was discovered that they did not. A list of the genes in this region was then compiled using NCBI data and analysed.

4.3 Results

DNA was extracted from leaves from each individual in the four progeny populations and quantified and evaluated for purity prior to sending to the Cornell University Institute of Biotechnology for GBS in a method they had developed as part of the VitisGen project which was adapted from Barley (Elshire et al. 2011). Approximately 80,000 SNPs were found from the four populations as a whole.

It was found that many of these SNP loci had errors and missing data and so had to be removed. The most useful SNPs were those that were homozygous in one parent and heterozygous in the other parent. These SNPs were filtered and used in later analysis. To do this, a pipeline was developed that used some of the elements of the Tassel pipeline (Glaubitz et al. 2014) and filtering with database scripts. The pipeline used for the filtering of the data is detailed in Figure 4.1.

The Grapevine genome has 19 chromosomes and 13 smaller fragments unplaced on the genome. Sequencing reads were aligned to these 19 chromosomes and 13 fragments. For ease of progress through the pipeline these chromosome fragments were named 20 to 32 instead of the names originally assigned. Errors in labeling of individuals from the progeny populations were fixed so that the separate sequencing runs could be merged correctly. Names were changed and the identical taxa merged using a plugin that is part of the Tassel-GBS pipeline (Glaubitz et al. 2014). Files were then split into the four populations. As alignment to the genome and SNP discovery had been performed on the four populations as a whole, some SNPs were completely homozygous in one population and these SNPs were removed for each population individually.


Change the names of the chromosomes so the filenames have numbers in sequential order
Fix errors in the names of individuals and then merge data from individuals into one with MergeIdenticalTaxaPlugin (Tassel pipeline)
Make a text file containing the names of the individuals in the population and sort into populations using includeTaxaInFile (Tassel Pipeline)
Load files into Tassel GUI in correct chromosome order and merge all the chromosome files into one genome file for each population

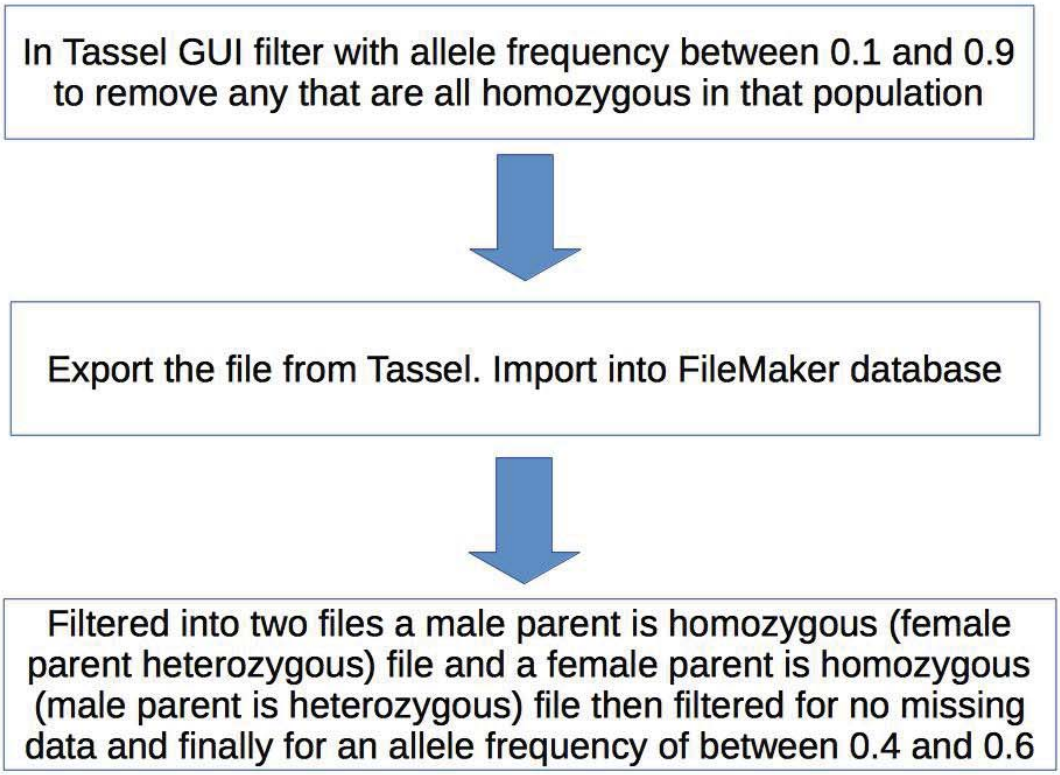


Figure 4.1: Flow chart of the process used to filter SNPs into smaller sets used for GLM and QTL analysis. Details on Tassel and Tassel pipeline can be found in publications Glaubitz et al. (2014) and Bradbury et al. (2007).

This created a unique set of SNPs for each population. To remove these sites, each population was filtered with a minor allele frequency of between 0.1 and 0.9 in the Tassel GUI. These files were exported from Tassel and loaded into FileMaker 12 database software. Using this database SNPs were sorted on a minor allele frequency of between 0.4 and 0.6 and any missing data were also filtered out. Finally these files were filtered into two output files, one with SNPs heterozygous in the female parent and homozygous in the male parent and one with SNPs heterozygous in the male parent and homozygous in the female parent. Following this filtering population 3 had 1,474 SNPs across the two files, populations 6 had 8399 SNPs, population 9 had 5244 SNPs and population 10 had 11,650 SNPs.

A PCA was performed in Tassel using these filtered files and the results were plotted to determine if there were any individuals who had vastly different genotypes compared to the rest of the population (Figure 4.2). Errors in labeling of plants during the vineyard planting, the collection of the DNA or in the placing of the DNA in the plates to send to Cornell may have led to incorrect individuals being sampled or sequenced. Most of the individuals in a population form a cluster. For population 3, a few individuals were on the edges of this cluster but they had at least one other individual very close by (Figure 4.2 a). For population 6 they are more spread out (Figure 4.2 b). For population 9, individual 50 was far from the others along the X axis but still within the range on the Y axis (Figure 4.2 c). For population 10 there was a good cluster (Figure 4.2 d). In conclusion no individuals were considered to be assigned to the wrong population.

To perform the GWAS, the cleaned up genotype data was used along with the phenotype data collected from the pre-veraison and mature berry stages for the four populations (see

Chapter 3) and the PCA. The resultant P-values were then used in resampling-based multiple hypothesis tests to account for any false positives. The corrected P-values from the four populations for pre-veraison and mature berry tartaric and malic acids over the two years for both the GWAS and QTL analysis were then ordered using Manhattan plots (Figures 4.3-4.30).

For the GWAS only population 9 showed any significant SNPs for the two years (Figures 4.3 to 4.12). There was only one significant SNP with a P-value below 0.05 for the GWAS in the 2012-2013 season and this was for mature berry tartaric acid. This SNP was located on chromosome 19 and had a P-value of 0.016 (Figures 4.3 and 4.4). In the 2013-2014 season significant SNPs were seen in pre-veraison and mature berry tartaric acid and none were seen for malic acid at both time points. Pre-veraison tartaric acid had approximately forty SNPs located across chromosome 5 with P-values ranging from 0.012 to 0.048 (Figure 4.5 and 4.6). There were also three SNPs on chromosome 7 with P-values between 0.012 and 0.044 and one on chromosome 19 (which does not match the one SNP seen in the 2012-2013 season) (Figures 4.5, 4.7 and 4.8). For mature berry tartaric acid there were the same forty or so SNPs on chromosome 5 with P-values between 0.007 and 0.037 seen in the pre-veraison analysis, one SNP on chromosome 11 with a P-value of 0.040 and two on chromosome 19 with P-values of 0.007 and 0.025 (Figures 4.9 to 4.12).

The same cleaned up SNP data that was used for the GWAS was used for the QTL analysis. Data was entered into the R package Onemap to make genetic maps of the four populations. Due to the large size of populations 6 and 10, these two were not merged into a single file and instead kept as a male parent heterozygous SNP data file and a female parent heterozygous SNP data file; two maps were created for these two populations as

there was no ability in the Onemap or qtl packages to combine maps. To make the maps, SNPs were organised into linkage groups. As all the SNPs had been mapped to the grapevine genome and their physical locations were known, the grouping of the SNPs could be compared to their physical position. The populations split into approximately 30 linkage groups with the 19 largest corresponding to the 19 *V. vinifera* chromosomes. The majority of SNPs in a linkage group all aligned to the same chromosome with a few SNPs from other chromosomes. These SNPs from other chromosomes were left in their linkage groups as they may have been incorrectly mapped on the physical map during the original alignment process.

These maps were then transferred to the R package qtl along with phenotype data to perform the QTL analysis. Many more significant SNPs were found in the QTL analysis than were found in the GWAS for the two years. For the 2012-2013 season significant SNPs with a LOD of above 3 were found in population 10 male parent, population 10 female parent population 3 and population 6 male parent (Figures 4.13 to 4.19). In the mature berry tartaric acid data analysis thirteen significant SNPs were found on chromosome 18 in population 10 male parent with LODs of 3.45 to 3.02 and fourteen significant SNPs were found on chromosome 12 in the populations 6 female parent with LODs of 3.76 to 3.02 (Figures 4.14 and 4.15). For population 10 female parent three significant SNPs were found on chromosome 3 with LODs of 3.30 to 3.18 and four SNPs on chromosome 1 with LODs of 3.06 to 3.02 in the mature berry malic acid data analysis (Figures 4.17 and 4.18). For population 3 for the mature berry malic acid data analysis two significant SNPs were found on chromosome 8 with a LOD of 3.43 (Figures 4.19). For population 6 male parent for mature berry malic acid fourteen significant SNPs were found on chromosome 1 with LODs of between 3.99-3.06 (Figure 4.18).

For the 2013-2014 season significant SNPs with a LOD of above 3 were found in population 10 male parent, population 10 female parent and population 9 (Figure 4.20-4.30). For population 10 male parent pre-veraison malic acid seven significant SNPs were found on chromosome 9 with LODs of 3.44 to 3.06 (Figure 4.22). For population 10 female parent, one significant SNP on chromosome 1 was found with an LOD of 3.02 for mature berry malic acid and two SNPs on chromosome 5 with LODs of 3.89 to 3.10 for mature berry tartaric acid (Figures 4.24 and 4.29). For population 9 two significant SNPs were found on chromosome 1 both with an LOD of 3.17 for pre-veraison malic acid (Figure 4.22). The same forty SNPs on chromosome 5 for pre-veraison tartaric acid that were significant in the GWAS were also significant in the QTL analysis with LODs of 3.93 to 3.02 (Figure 4.26). The same two SNPs for pre-veraison tartaric acid significant in the GWAS on chromosome 7 were also significant in the QTL analysis both with an LOD of 3.61 (Figure 4.27). For population 9 mature berry tartaric acid, the same forty SNPs seen in the pre-veraison tartaric acid and the pre-veraison and mature berry tartaric acid in the GWAS were also significant in the QTL analysis, however this significant area was expanded to include five more SNPs (Figure 4.29). One significant SNP was also seen on chromosome 11 which was significant in the GWAS as well (Figure 4.30).

There was a set of significant markers on chromosome 5 for population 9 pre-veraison and mature berry tartaric acid for both the GWAS and QTL analysis. The most significant of these span a region between approximately 400,000 bp and 3,000,000 bp. The genes within this region were studied to determine if any could be identified as possible candidates in tartaric acid metabolism. To achieve this the areas between the markers on the genome alignment was aligned against the genome assembly used by NCBI to determine gene annotations. Areas in both assemblies matched in position. A list of all

genes in that area was composed using the NCBI annotations. As this region is very large, it contained a large number of genes. Of the genes, many of them were uncharacterised and many were receptors, DNA binding proteins and two were ABC transporters.

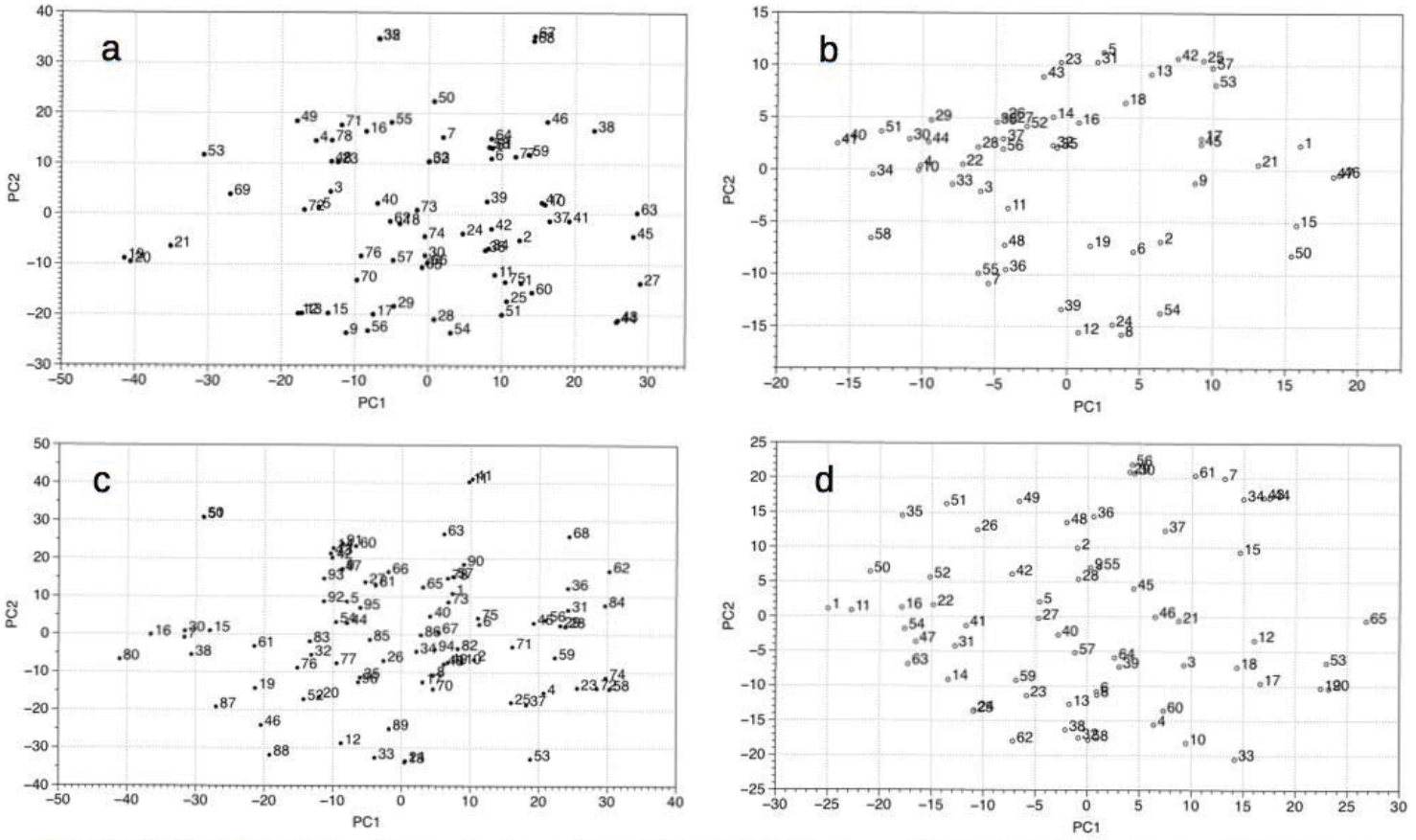


Figure 4.2: PCA plots of populations from filtered SNP data. These PCAs were used for GLM analysis. a) Population 6. b) Population 3. c) Population 10. d) Population 9.

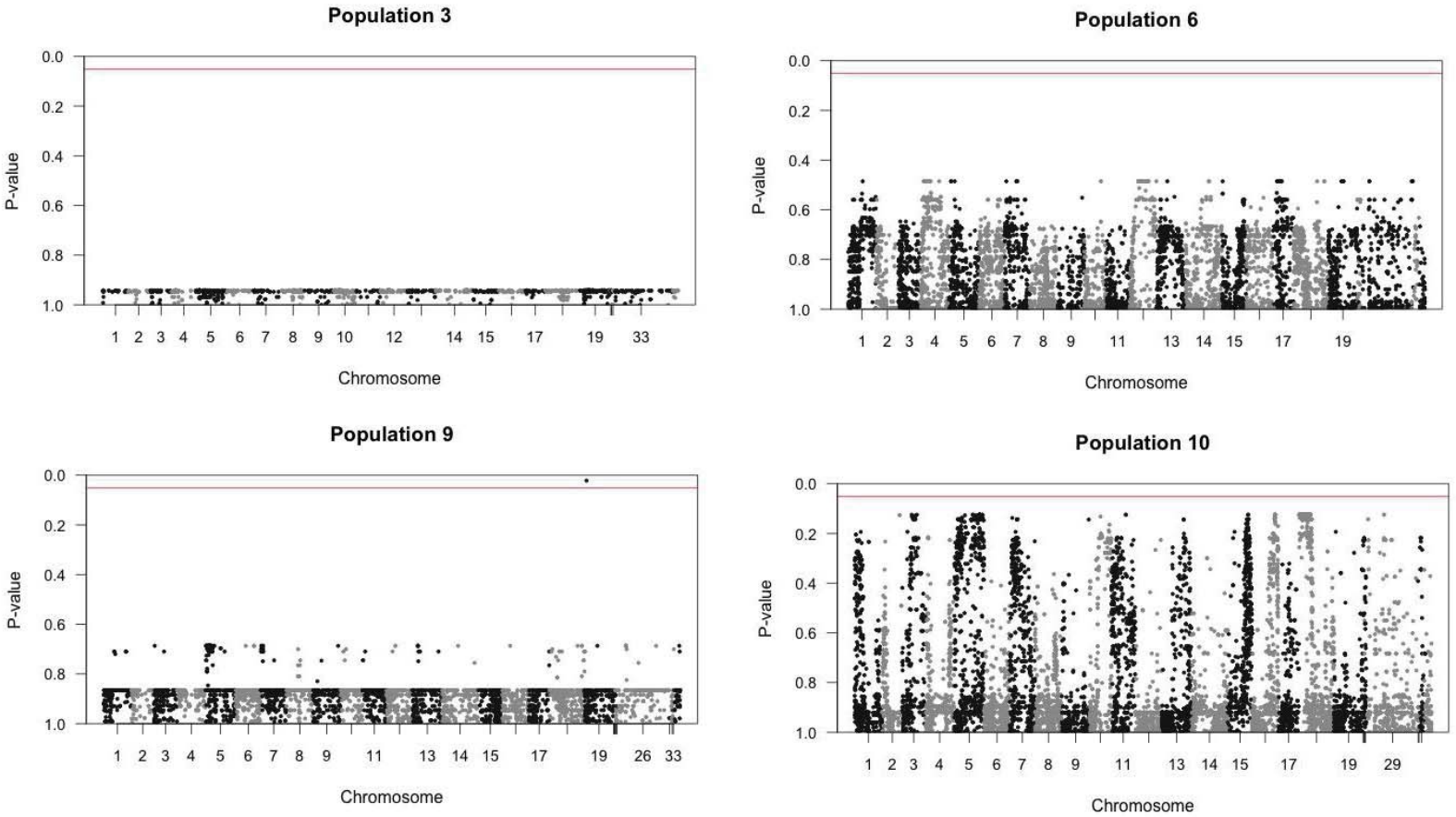


Figure 4.3: GWAS results for the four populations in the 2012-2013 season for mature berry tartaric acid. There is one significant SNP with a P-value under 0.05 in population 9 on chromosome 19.

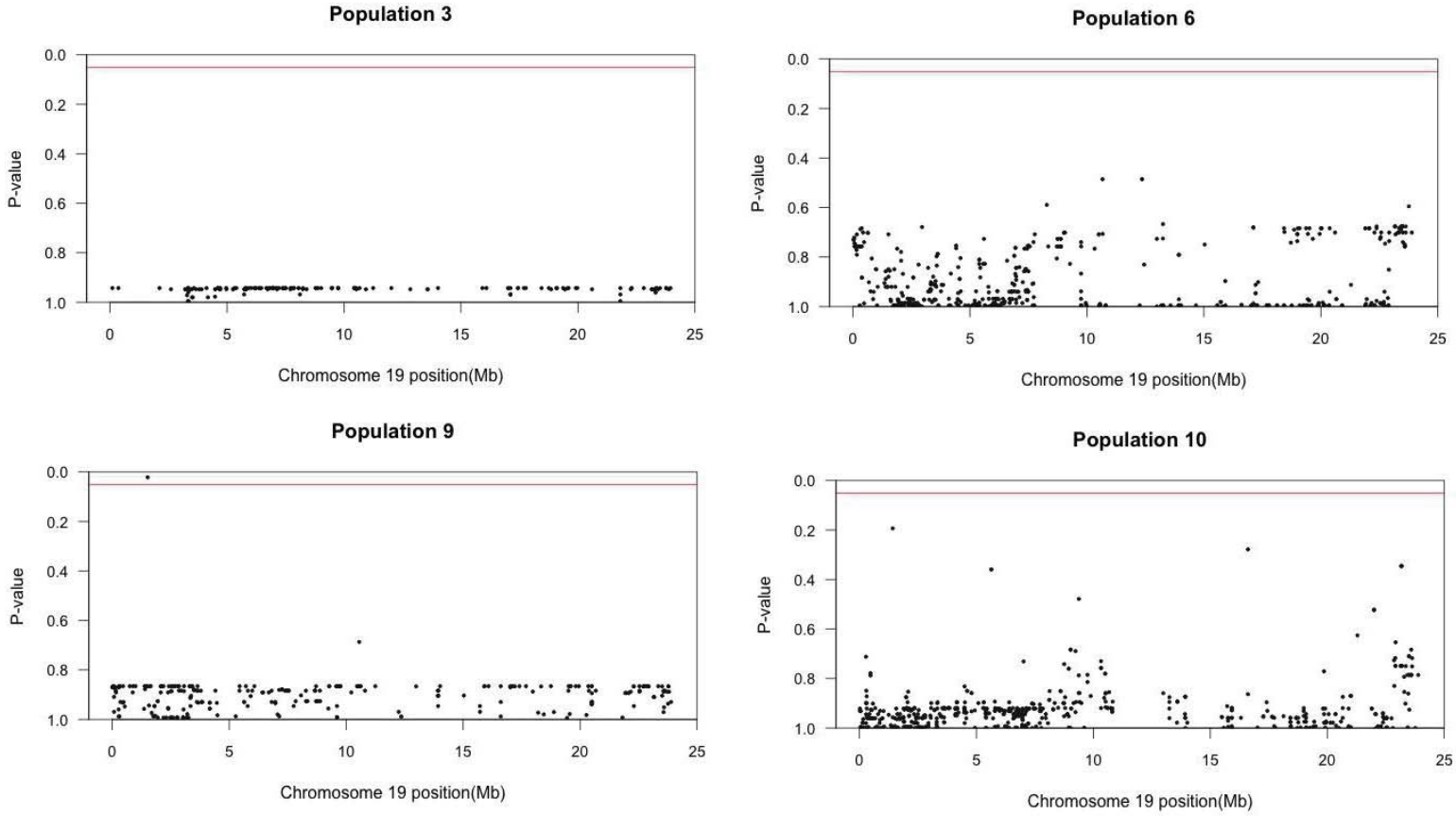


Figure 4.4: GWAS results for four populations, chromosome 19, in the 2012-2013 season for mature berry tartaric acid. There is one significant SNP with a P-value under 0.05 in population 9 on chromosome 19.

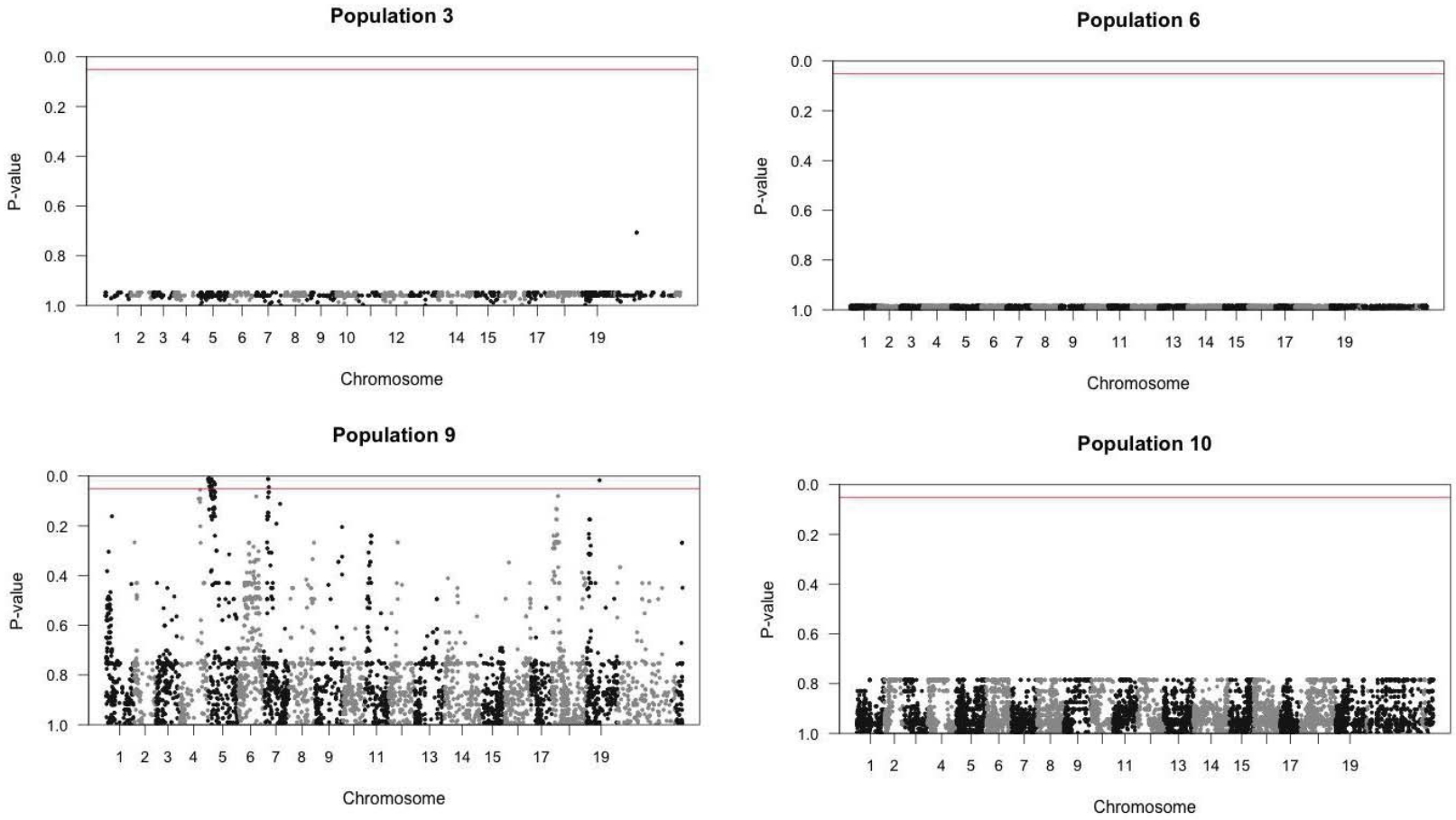


Figure 4.5: GWAS results for four populations in the 2013-2014 season for pre-veraison tartaric acid. There are only significant SNPs with a p-value under 0.05 for population 9.

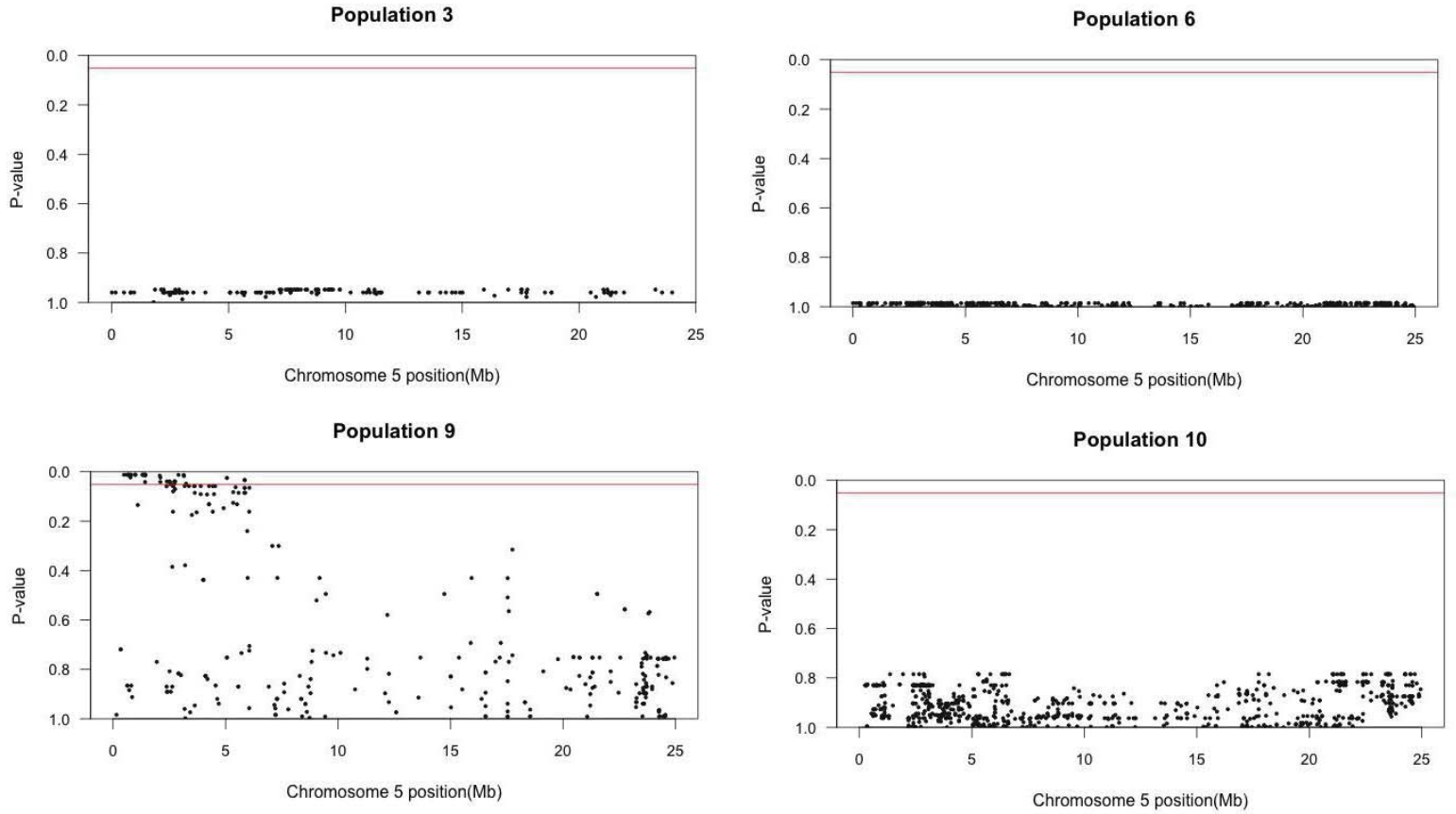


Figure 4.6: GWAS results for four populations, chromosome 5, in the 2013-2014 season for pre-veraison tartaric acid. There are only significant SNPs with a p-value under 0.05 for population 9.

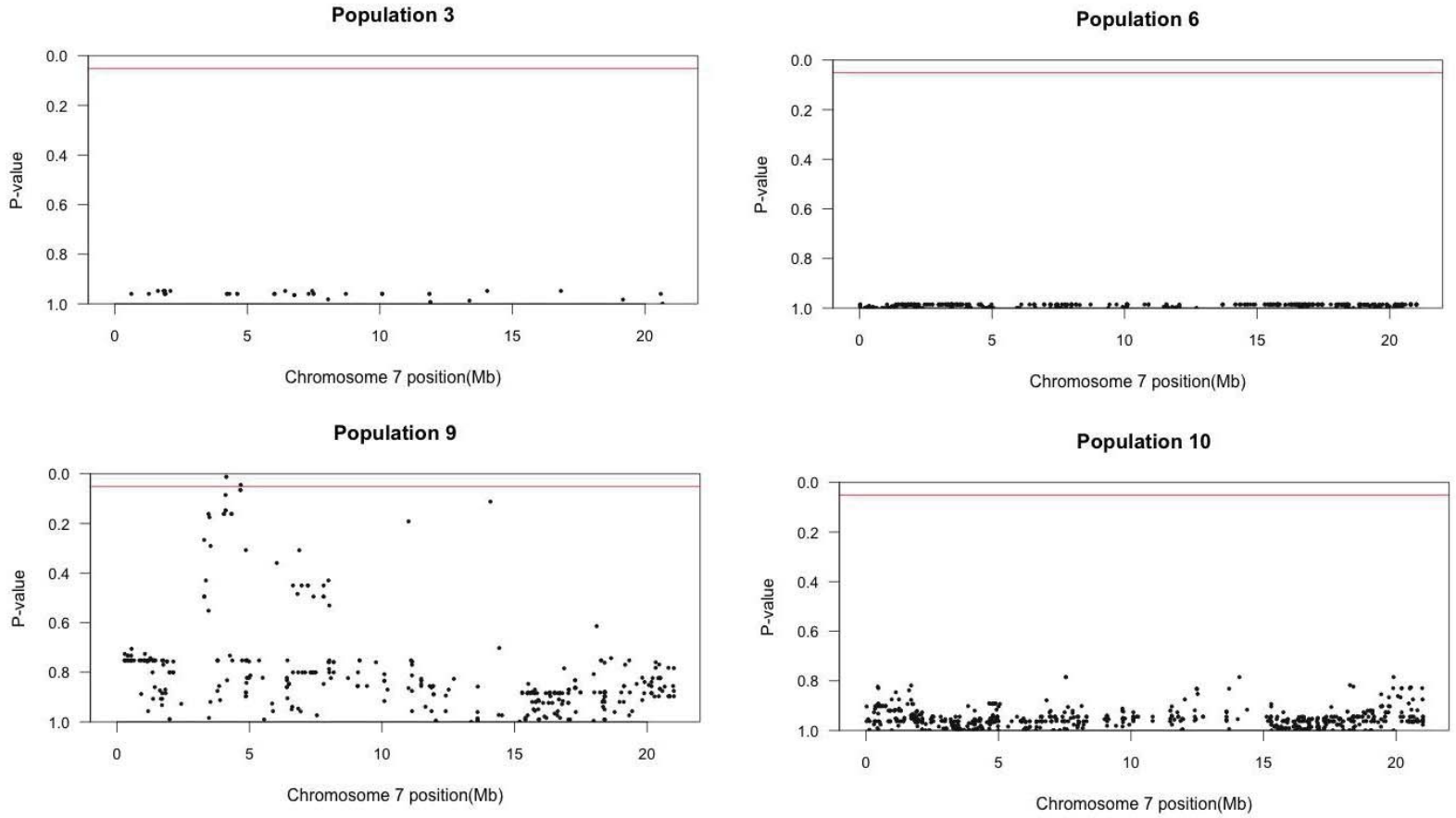


Figure 4.7: GWAS results for four populations, chromosome 7, in the 2013-2014 season for pre-veraison tartaric acid. There are only significant SNPs with a p-value under 0.05 for population 9.

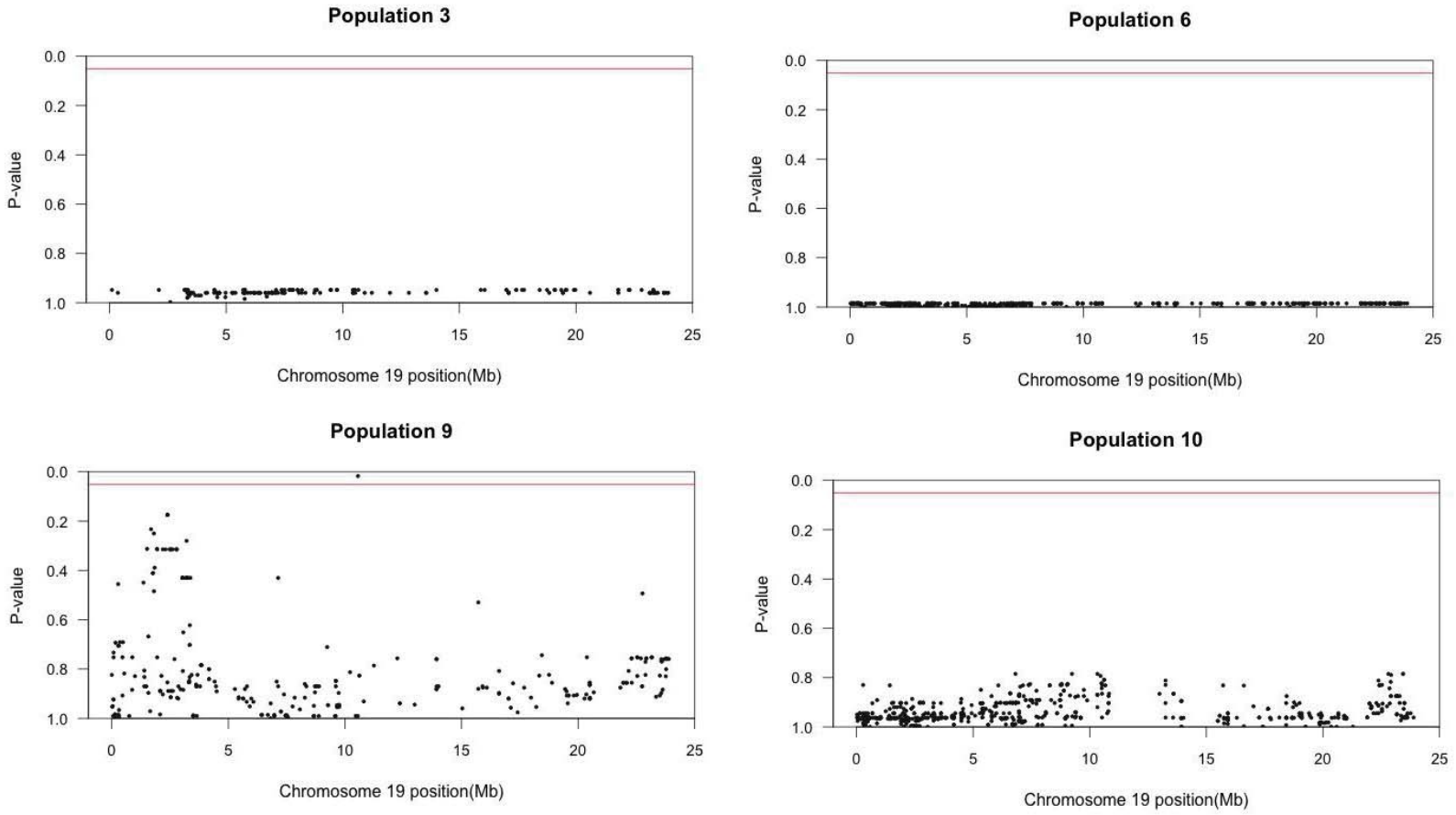


Figure 4.8: GWAS results for four populations, chromosome 19, in the 2013-2014 season for pre-veraison tartaric acid. There are only significant SNPs with a p-value under 0.05 for population 9.

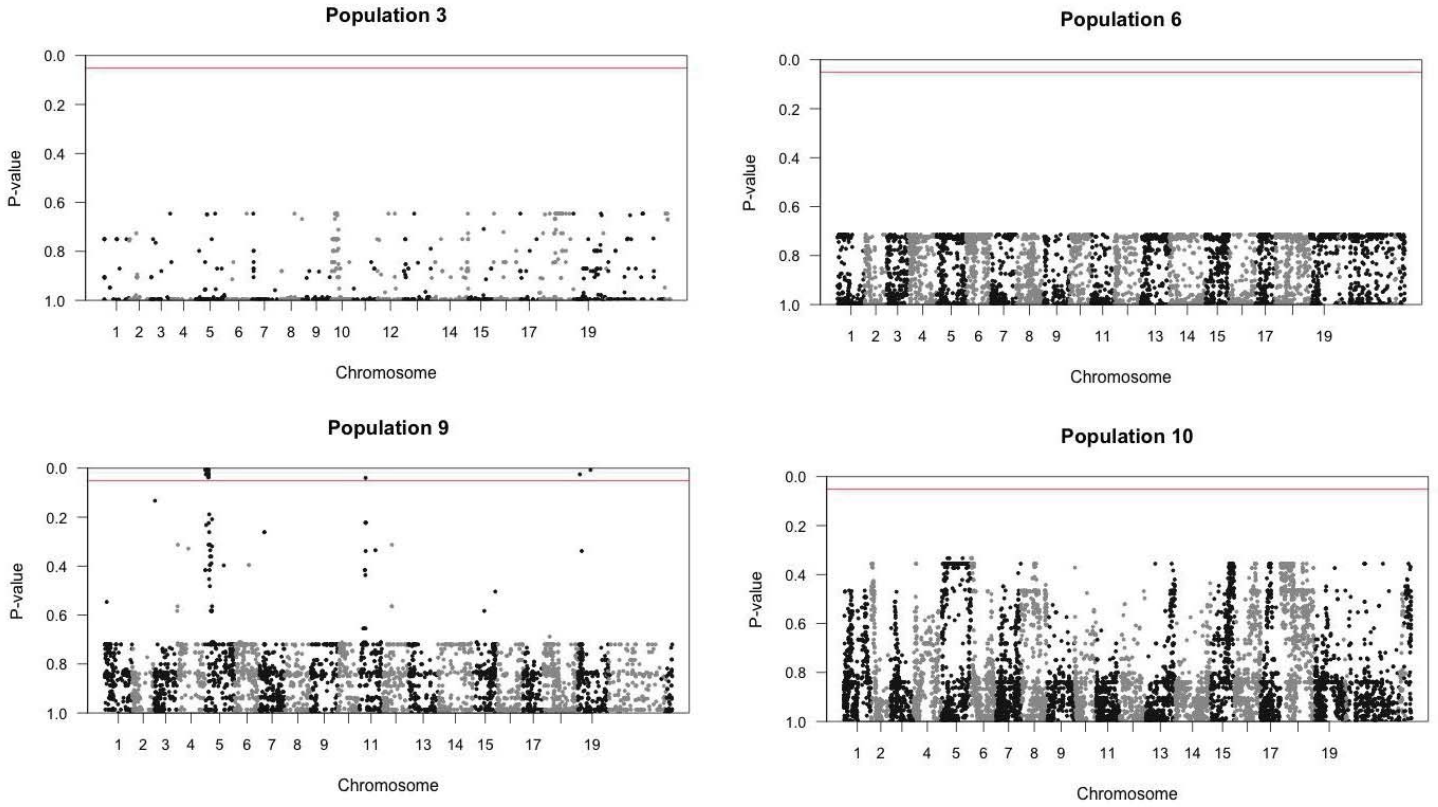


Figure 4.9: GWAS results for four populations in the 2013-2014 season for mature berry tartaric acid. There are only significant SNPs with a p-value under 0.05 for population 9.

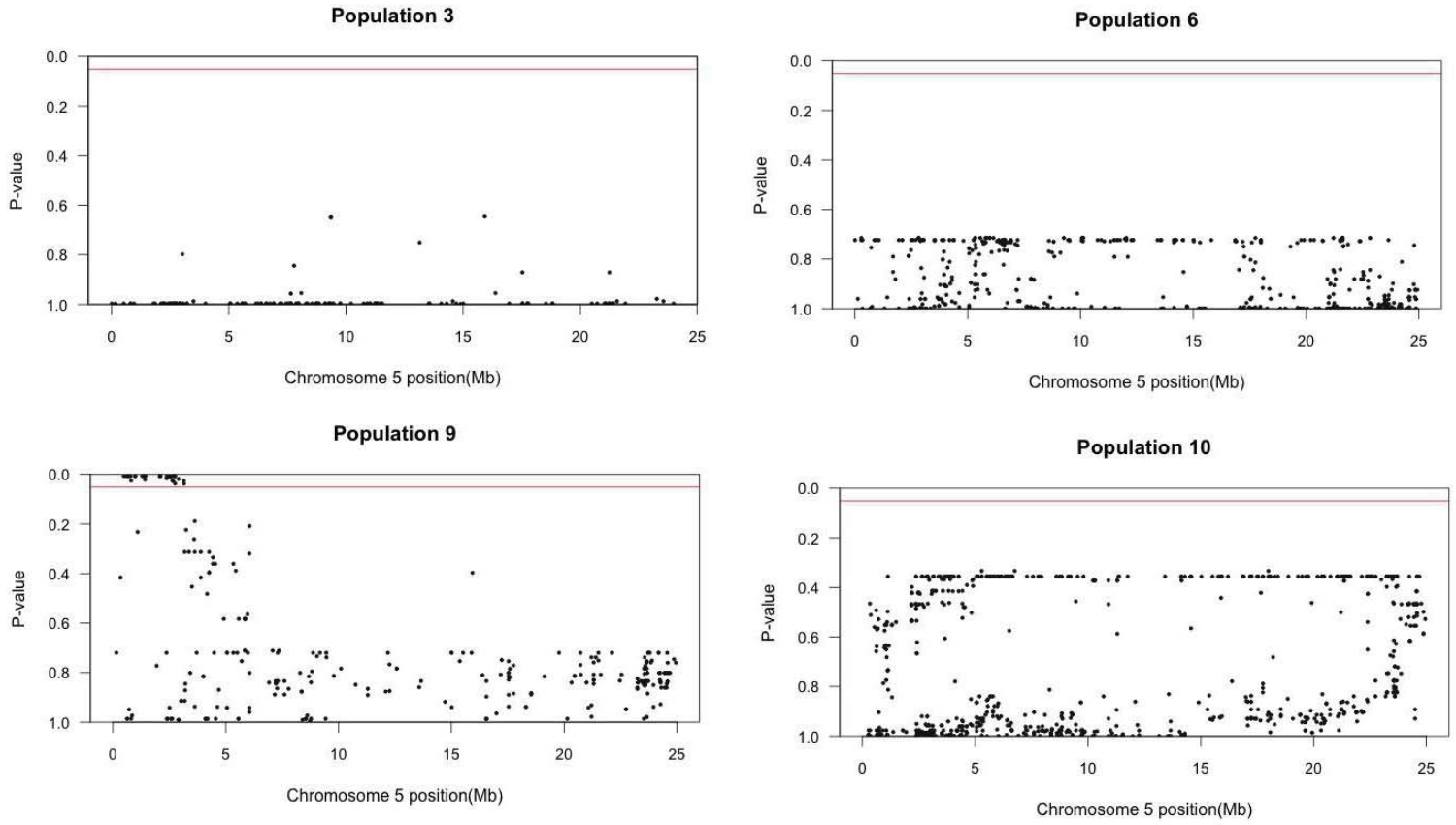


Figure 4.10: GWAS results for four populations, chromosome 5, in the 2013-2014 season for mature berry tartaric acid. There are only significant SNPs with a p-value under 0.05 for population 9.

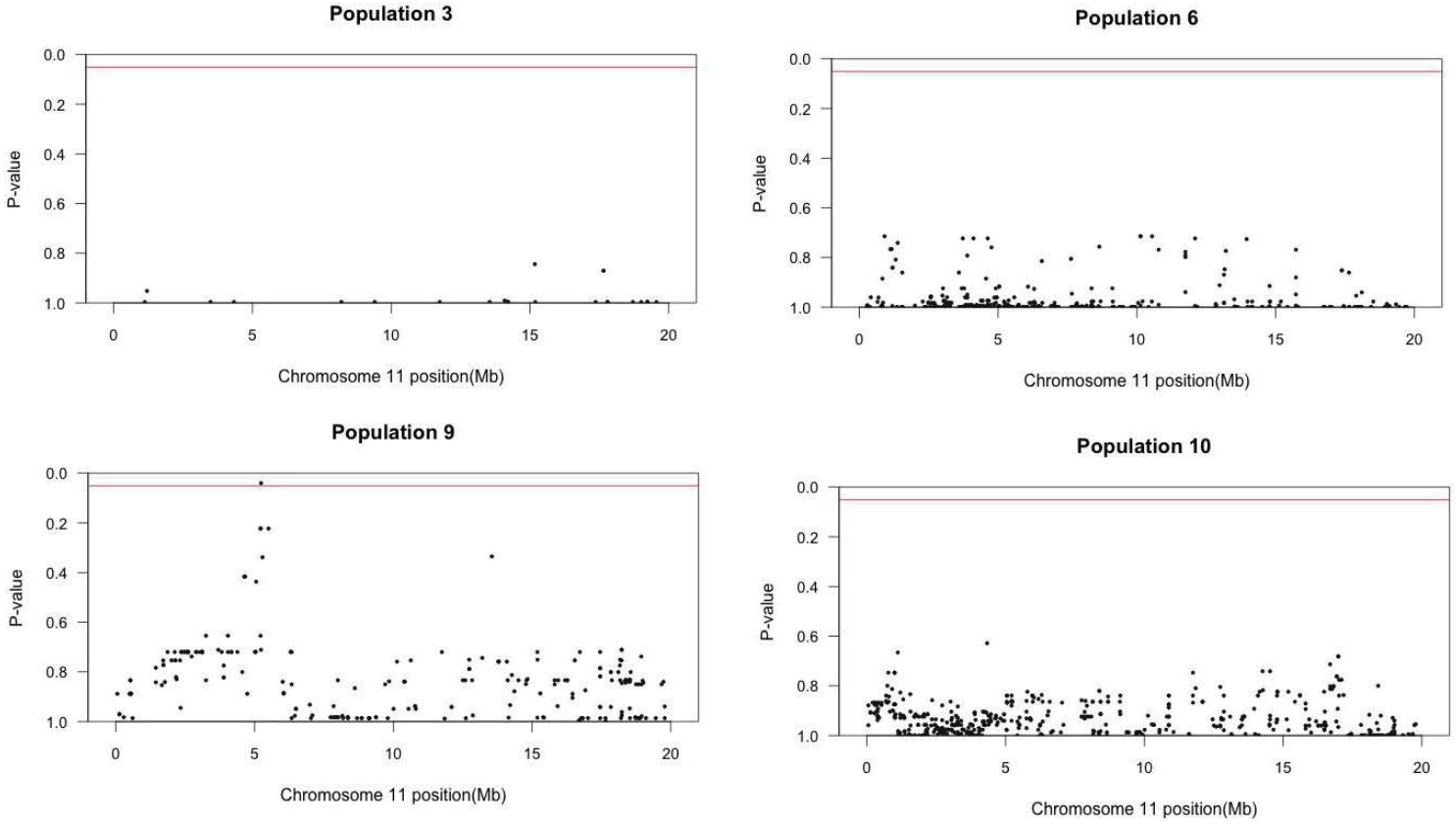


Figure 4.11: GWAS results for four populations, chromosome 11, in the 2013-2014 season for mature berry tartaric acid. There are only significant SNPs with a p-value under 0.05 for population 9.

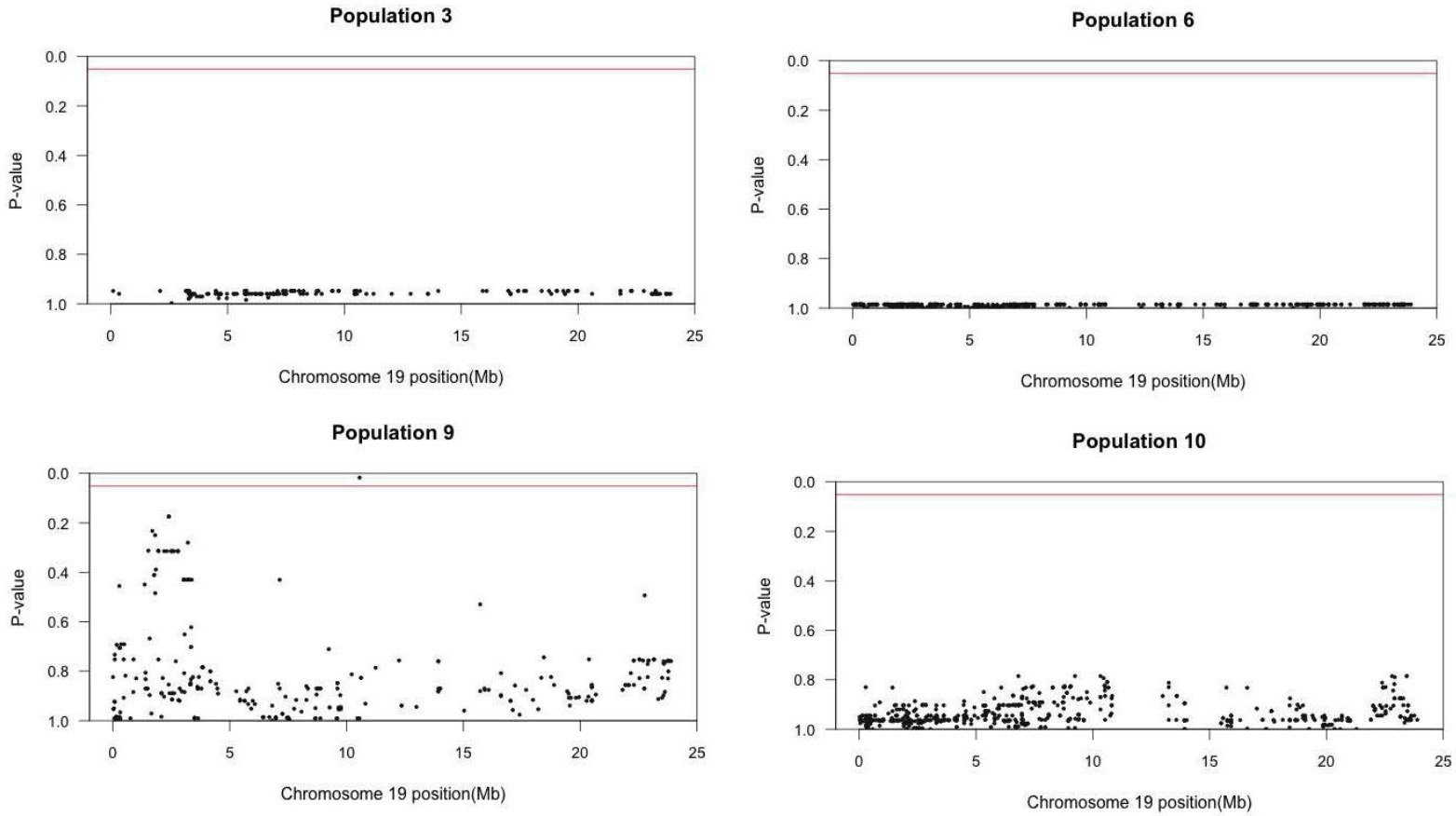


Figure 4.12: GWAS results for four populations, chromosome 19, in the 2013-2014 season for mature berry tartaric acid. There are only significant SNPs with a p-value under 0.05 for population 9.

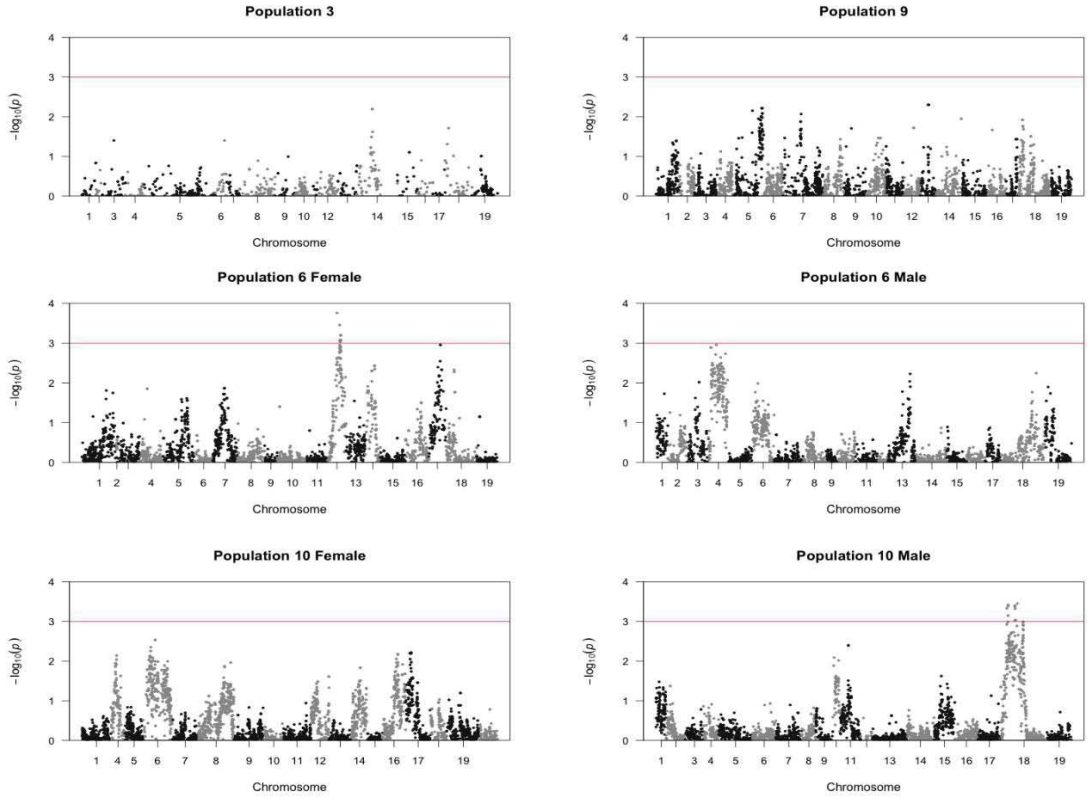


Figure 4.13: QTL analysis results for four populations in the 2012-2013 season for mature berry tartaric acid. There are significant SNPs with a LOD over three for population 6 female and population 10 male.

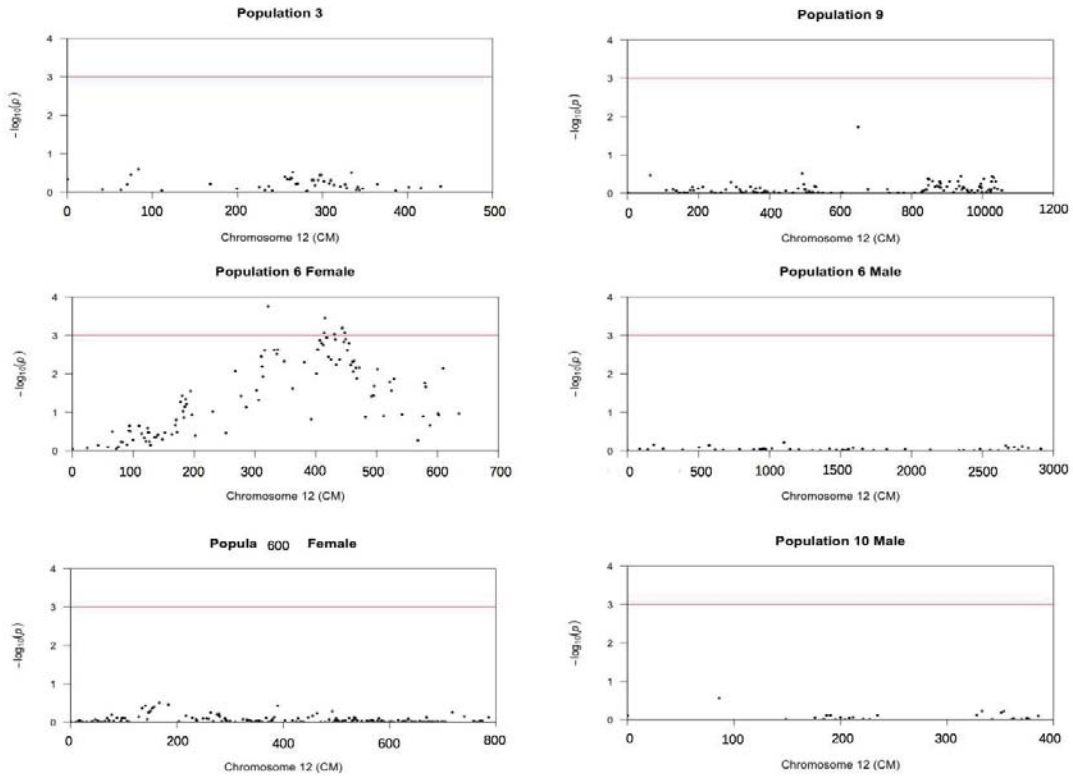


Figure 4.14: QTL analysis results for four populations, chromosome 12, in the 2012-2013 season for mature berry tartaric acid. There are significant SNPs with a LOD over three for population 6 female.

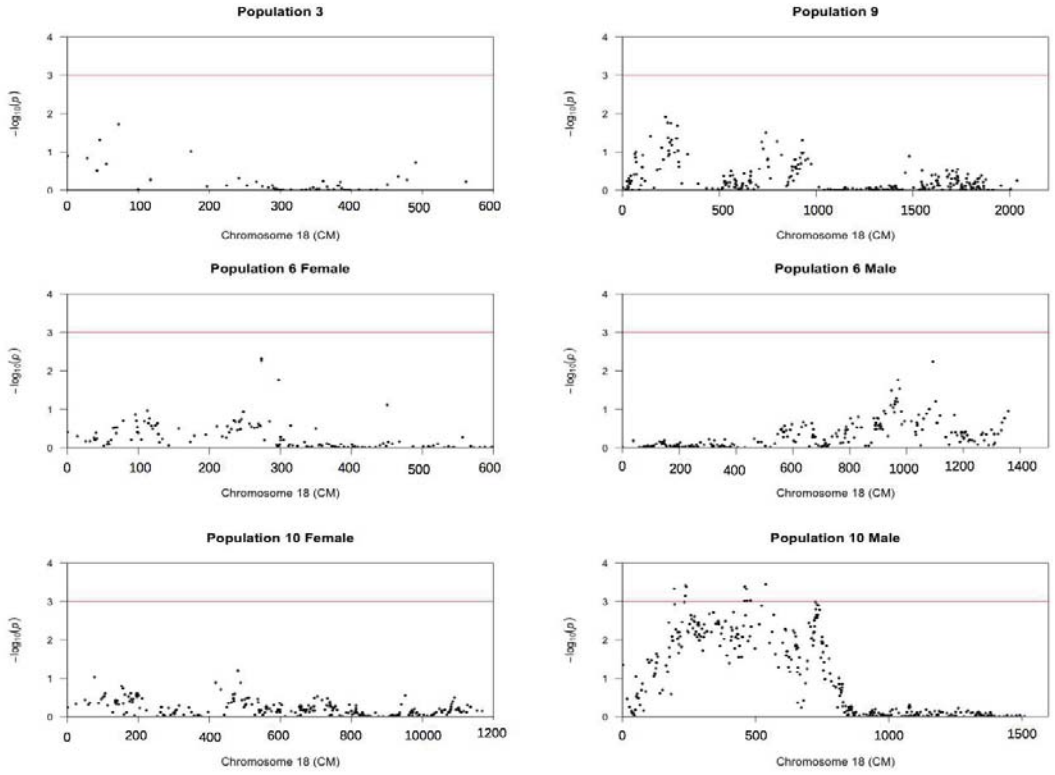


Figure 4.15: QTL analysis results for four populations, chromosome 18, in the 2012-2013 season for mature berry tartaric acid. There are significant SNPs with a LOD over three for population 10 male.

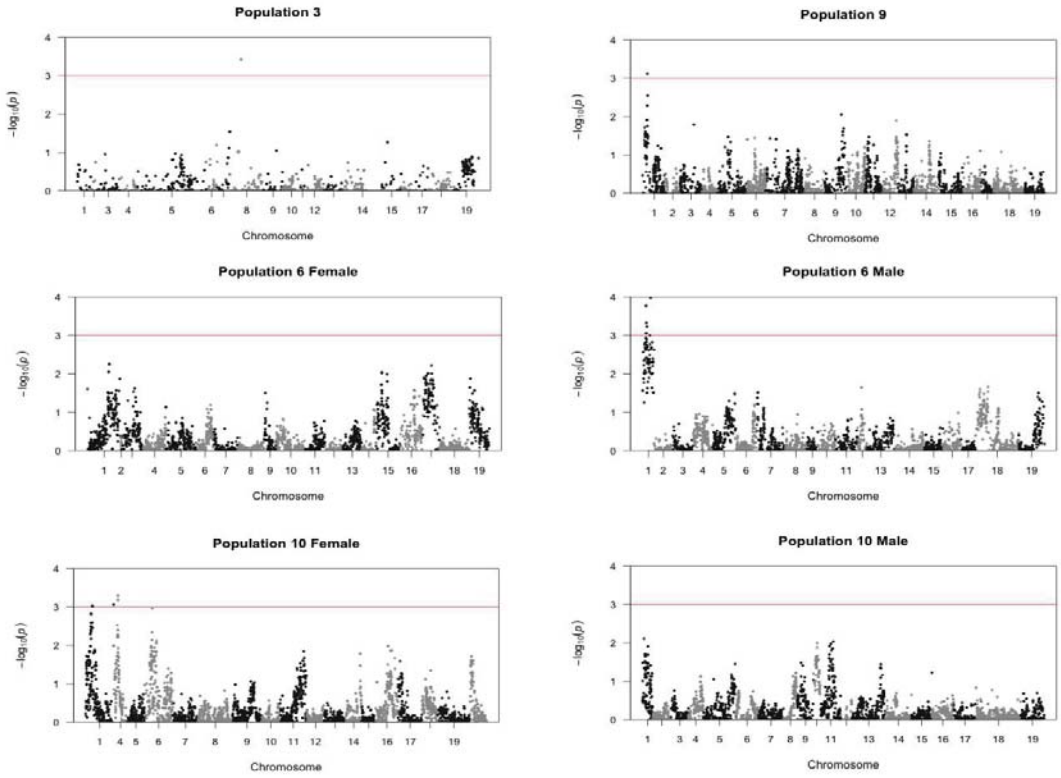


Figure 4.16: QTL analysis results for four populations in the 2012-2013 season for mature berry malic acid. There are significant SNPs with a LOD over three for population 3, population 9, population 6 male and population 10 female.

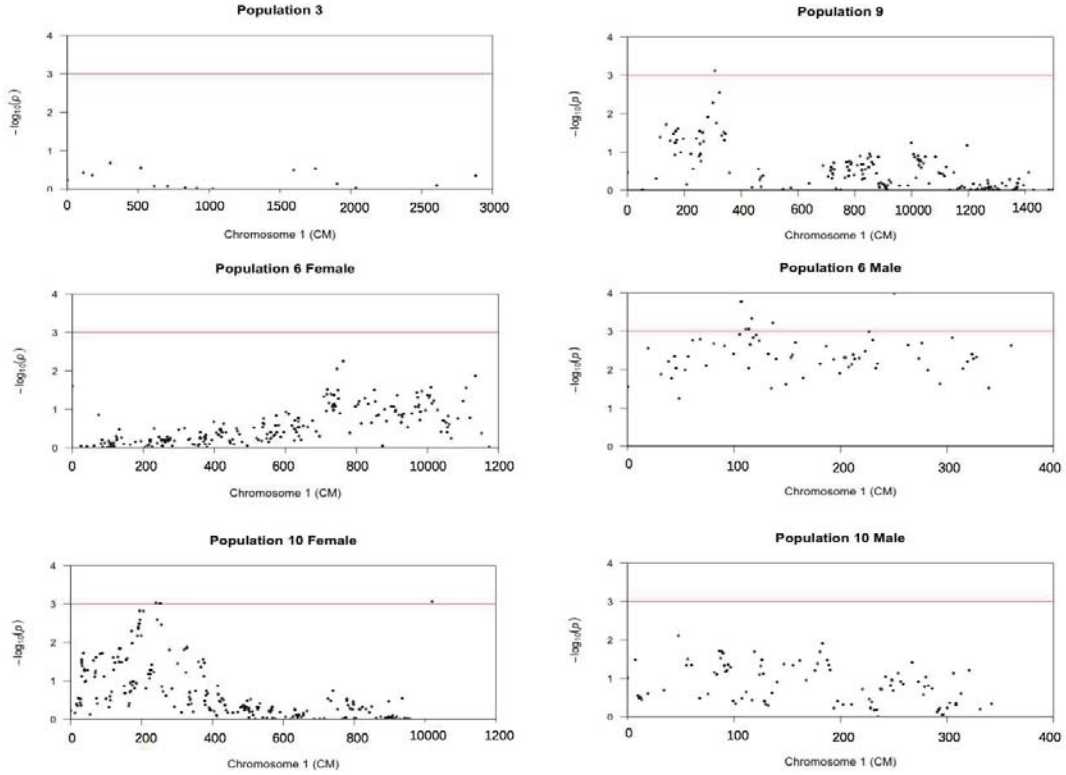


Figure 4.17: QTL analysis results for four populations, chromosome 1 in the 2012-2013 season for mature berry malic acid. There are significant SNPs with a LOD over three for population 9, population 6 male and population 10 female.

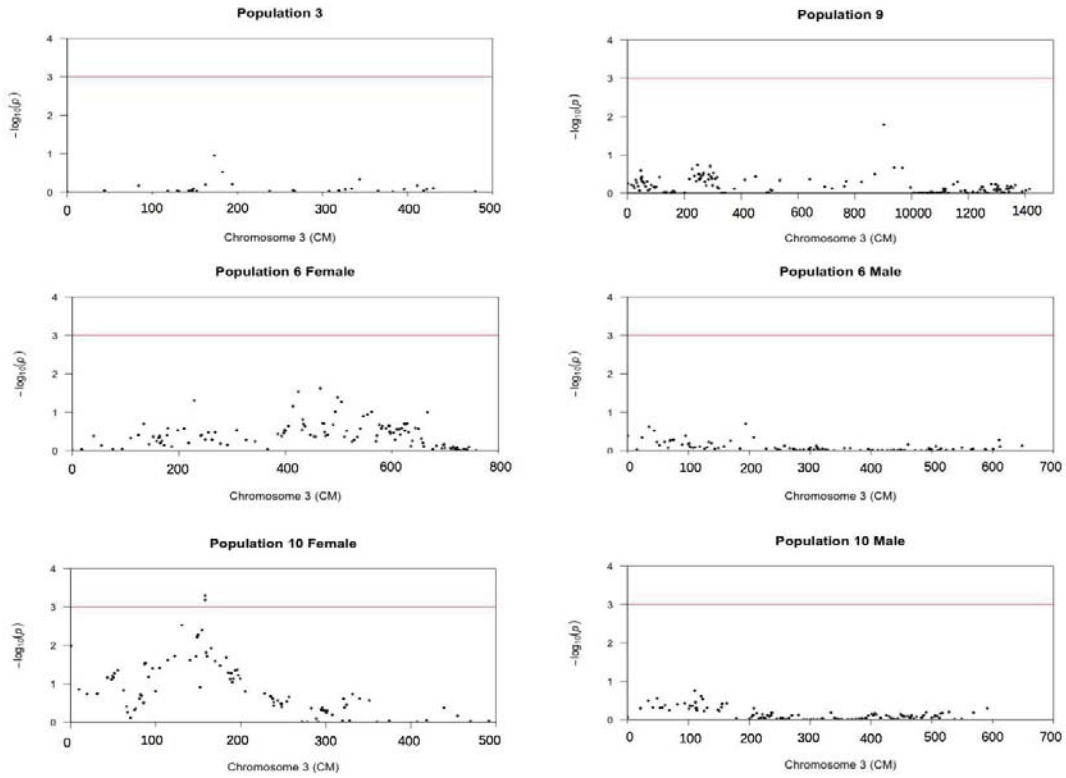


Figure 4.18: QTL analysis results for four populations, chromosome 3 in the 2012-2013 season for mature berry malic acid. There are significant SNPs with a LOD over three for population 10 female.

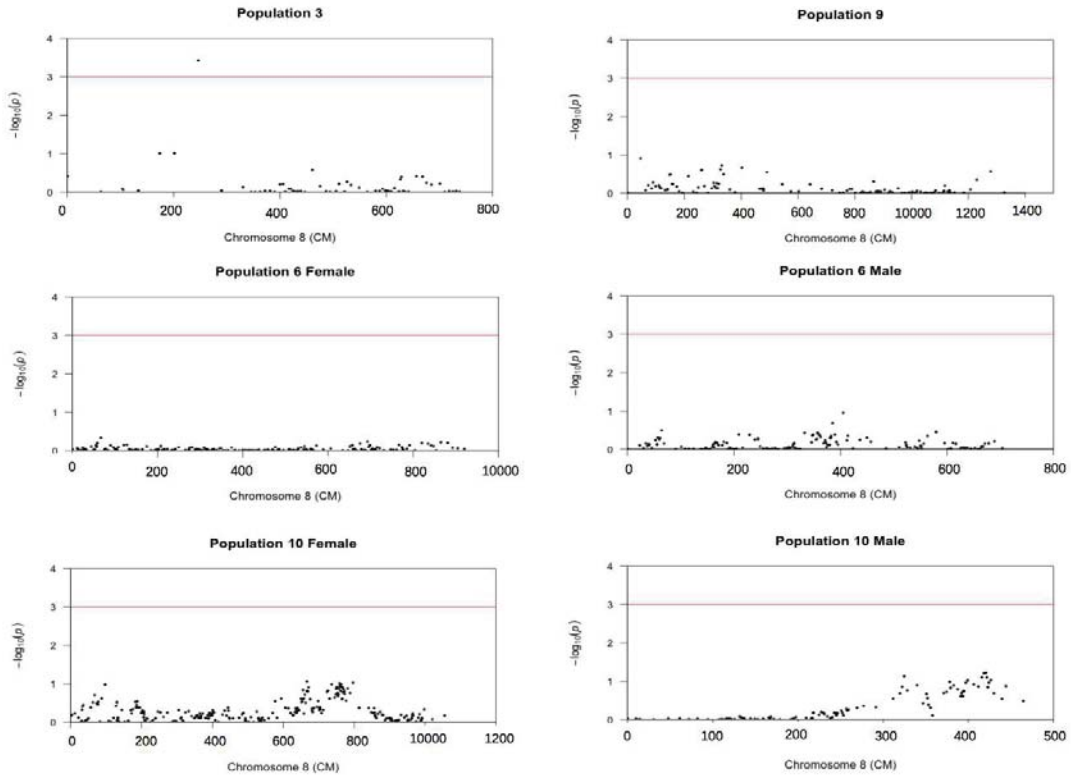


Figure 4.19: QTL analysis results for four populations, chromosome 8 in the 2012-2013 season for mature berry malic acid. There are significant SNPs with a LOD over three for population 3.

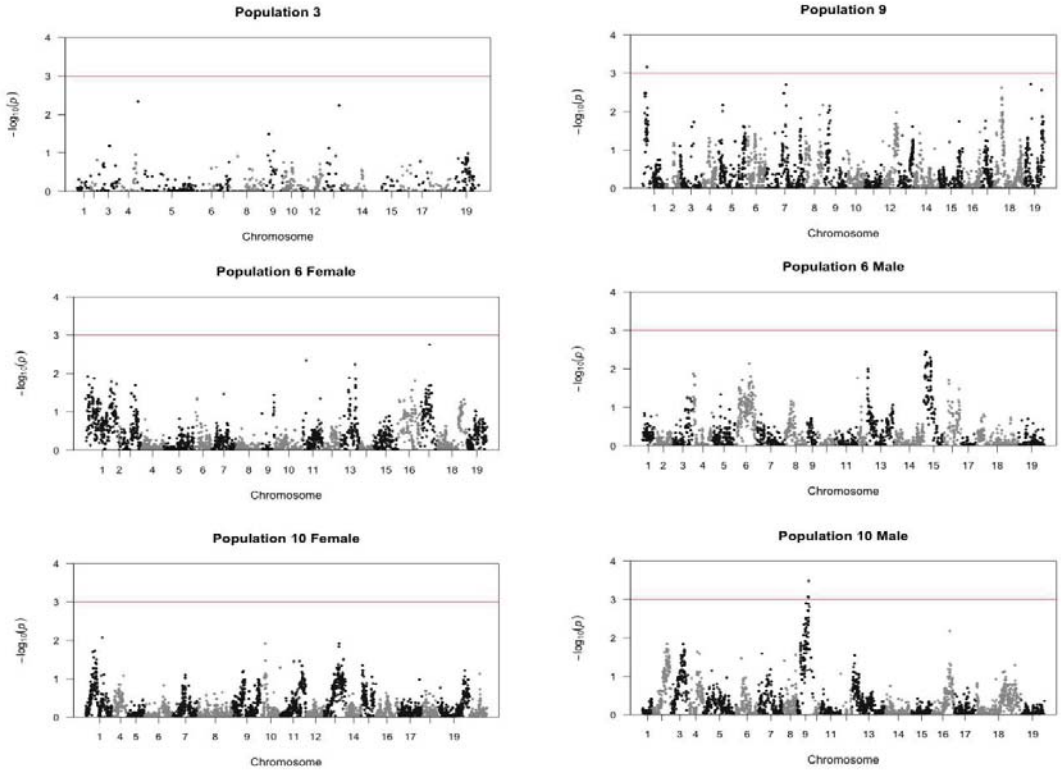


Figure 4.20: QTL analysis results for four populations in the 2013-2014 season for pre-veraison malic acid. There are significant SNPs with a LOD over three for population 9 and population 6 male.

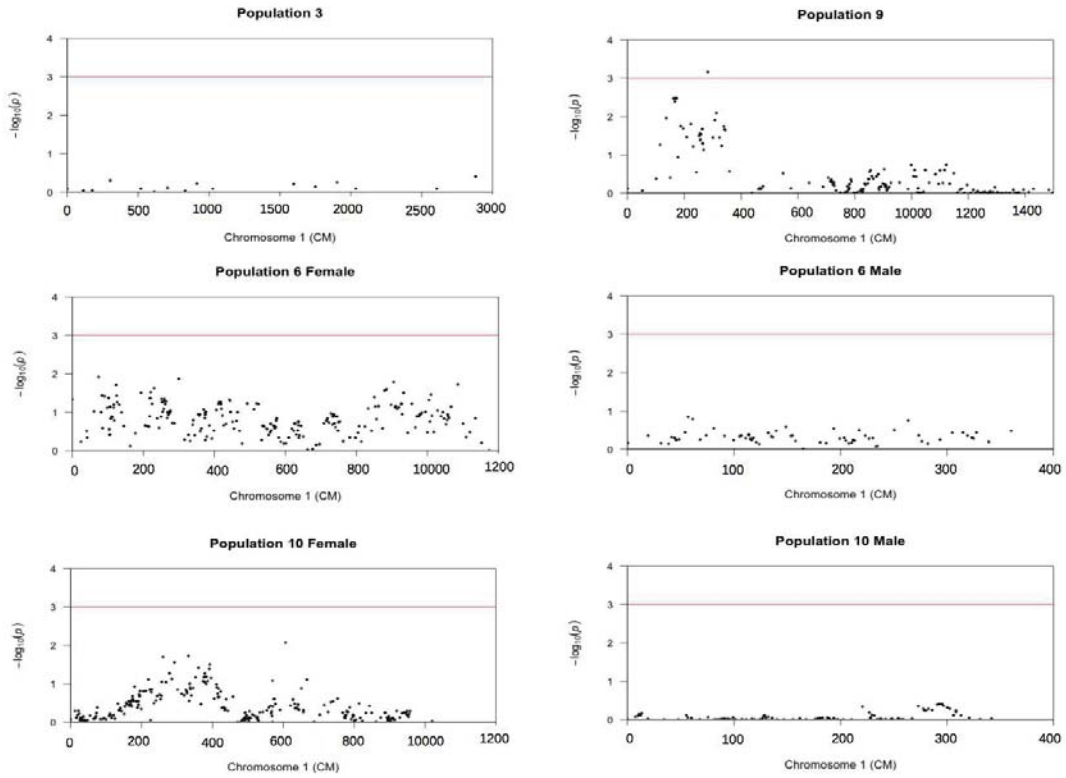


Figure 4.21: QTL analysis results for four populations, chromosome 1, in the 2013-2014 season for pre-veraison malic acid. There are significant SNPs with a LOD over three for population 9.

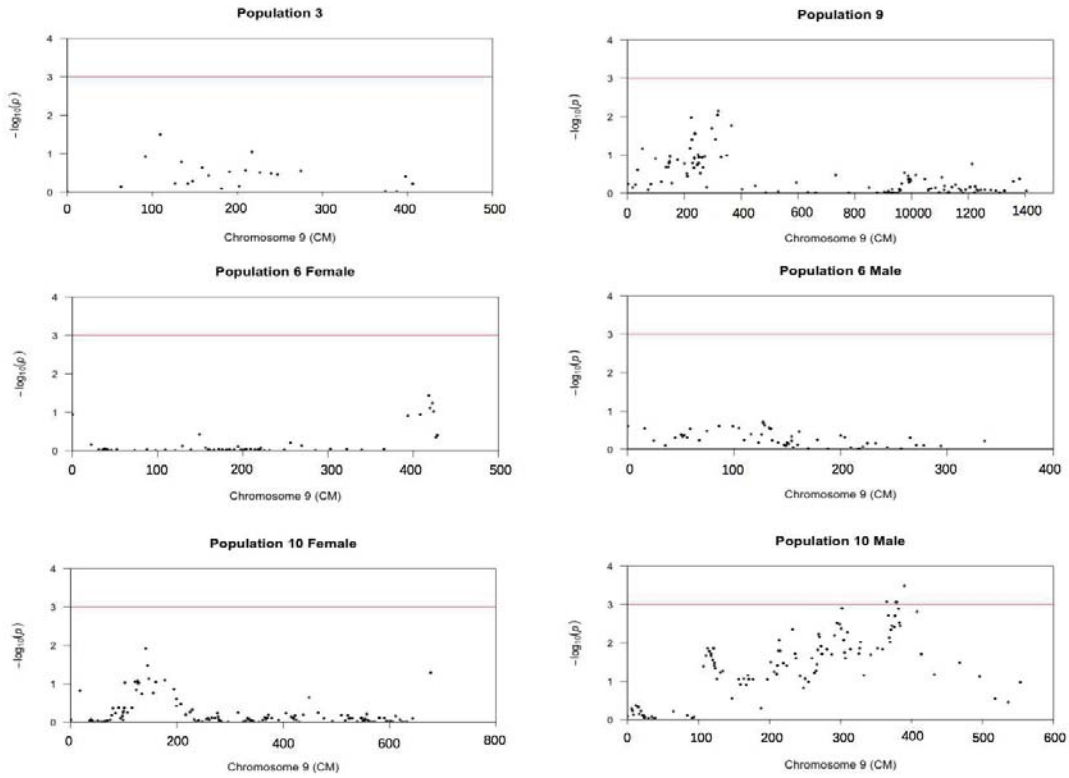


Figure 4.22: QTL analysis results for four populations, chromosome 9, in the 2013-2014 season for pre-veraison malic acid. There are significant SNPs with a LOD over three for population 6 male.

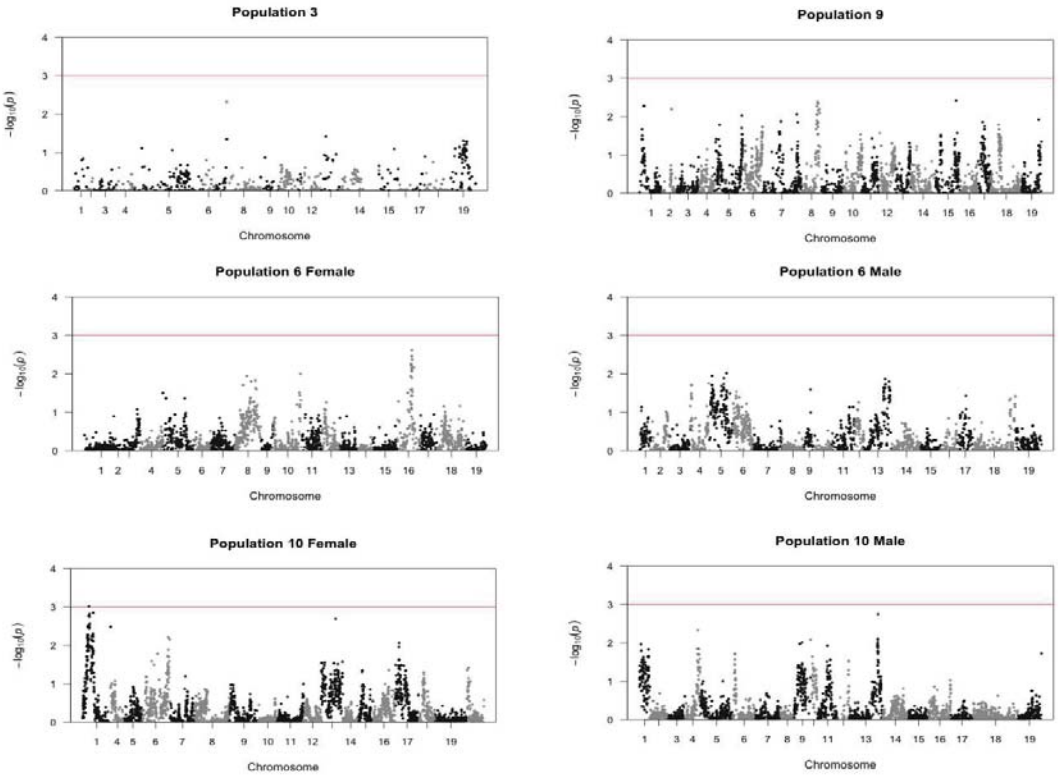


Figure 4.23: QTL analysis results for four populations in the 2013-2014 season for mature berry malic acid. There are significant SNPs with a LOD over three for population 10 female.

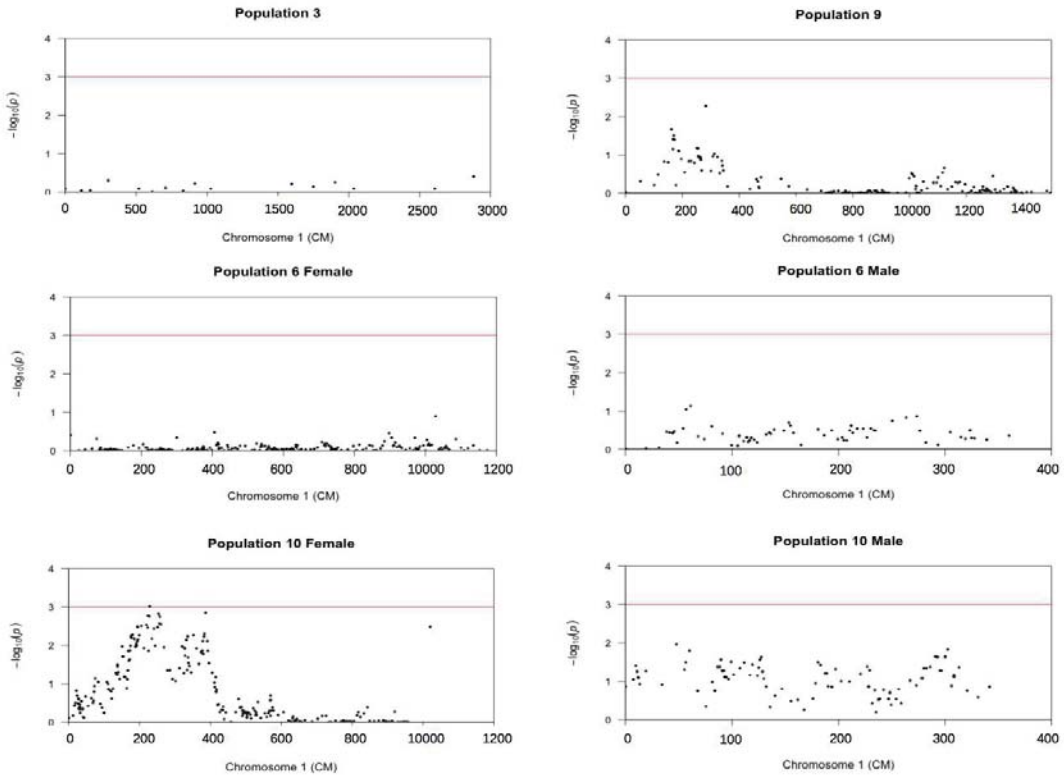


Figure 4.24: QTL analysis results for four populations, chromosome 1, in the 2013-2014 season for mature berry malic acid. There are significant SNPs with a LOD over three for population 10 female.

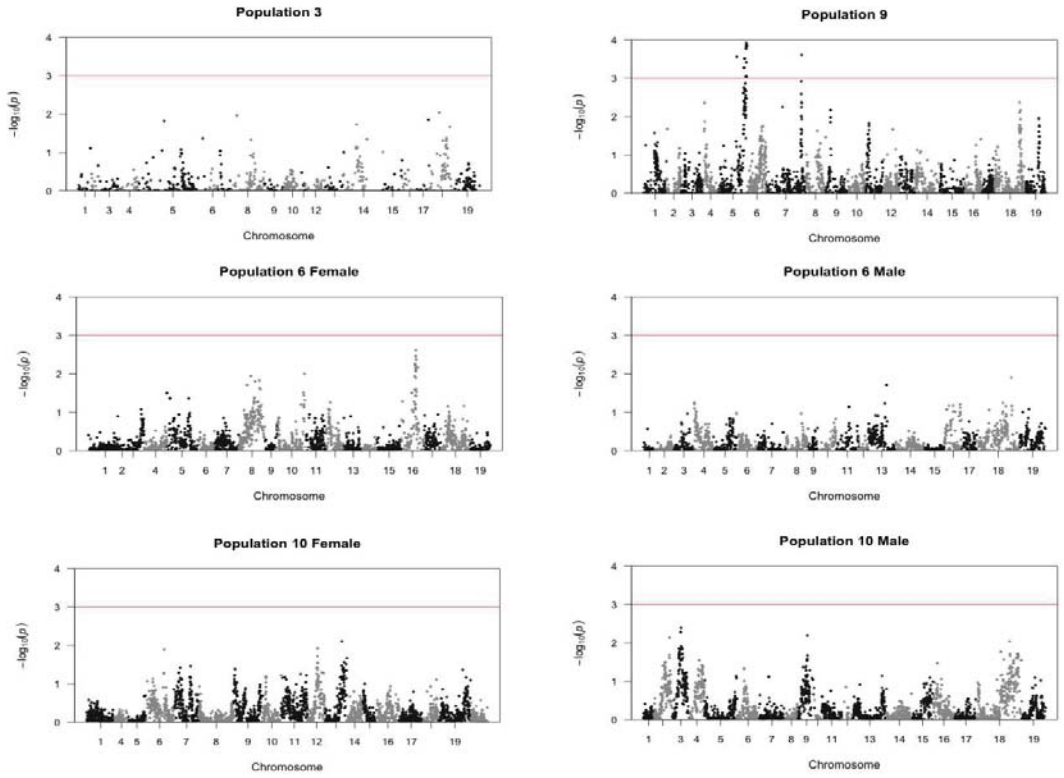


Figure 4.25: QTL analysis results for four populations in the 2013-2014 season for pre-veraison tartaric acid. There are significant SNPs with a LOD over three for population 9.

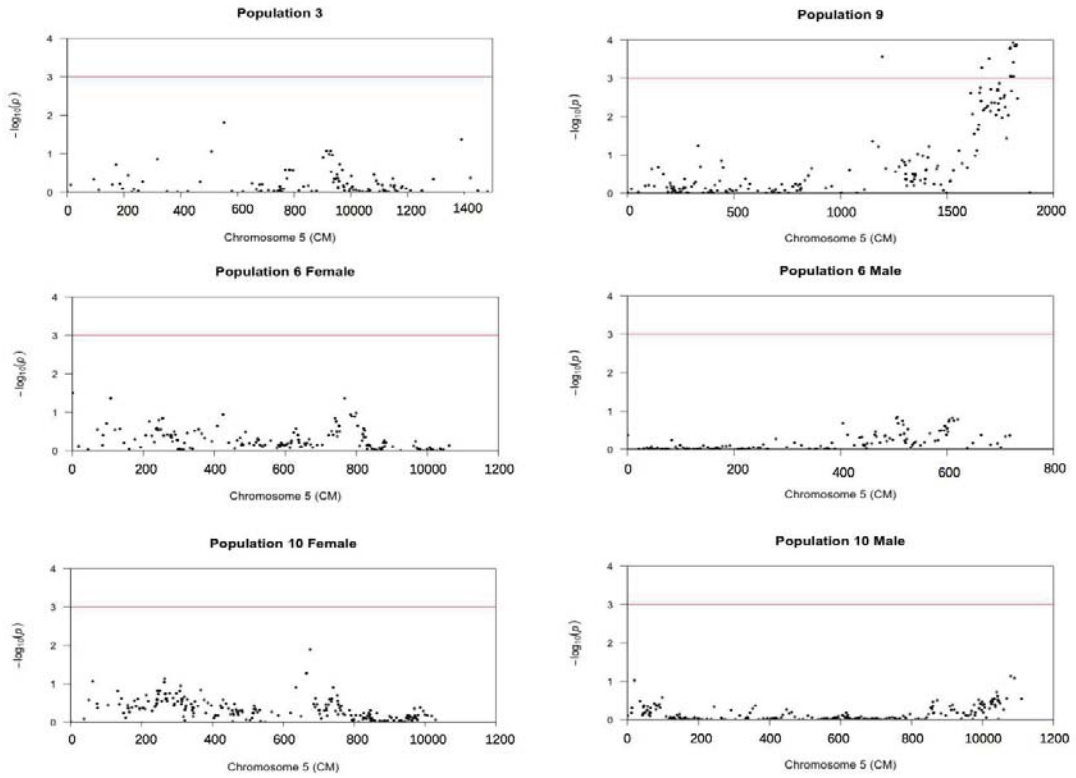


Figure 4.26: QTL analysis results for four populations, chromosome 5, in the 2013-2014 season for pre-veraison tartaric acid. There are significant SNPs with a LOD over three for population 9.

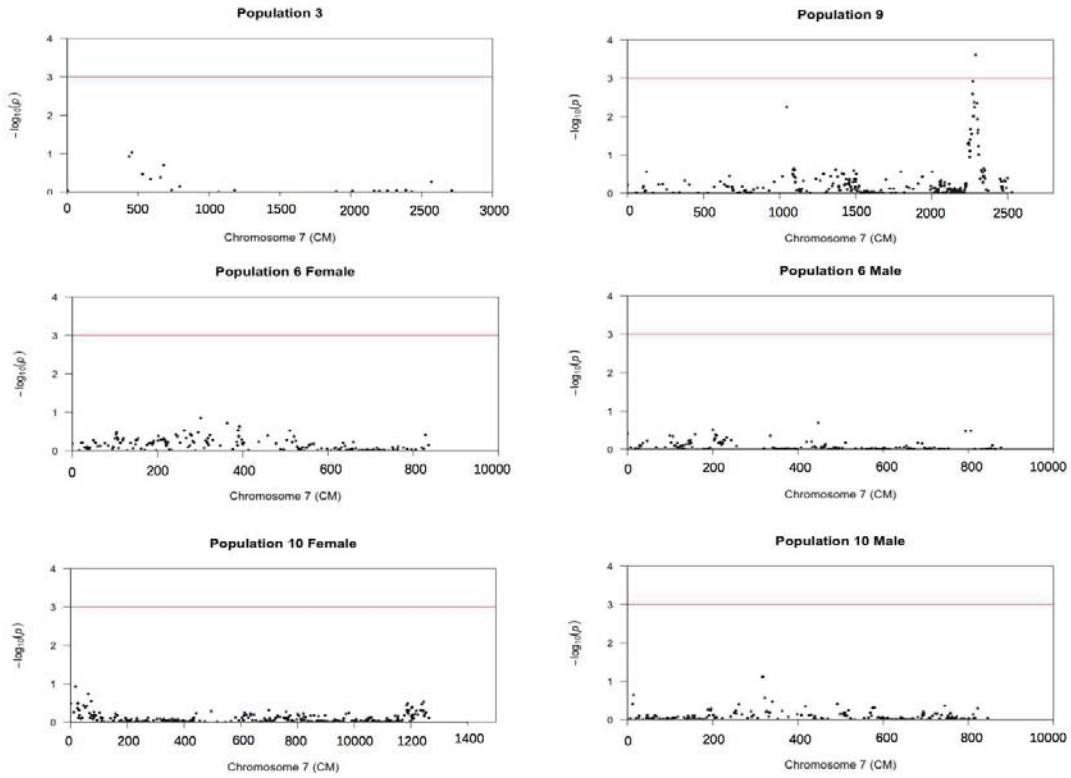


Figure 4.27: QTL analysis results for four populations, chromosome 7, in the 2013-2014 season for pre-veraison tartaric acid. There are significant SNPs with a LOD over three for population 9.

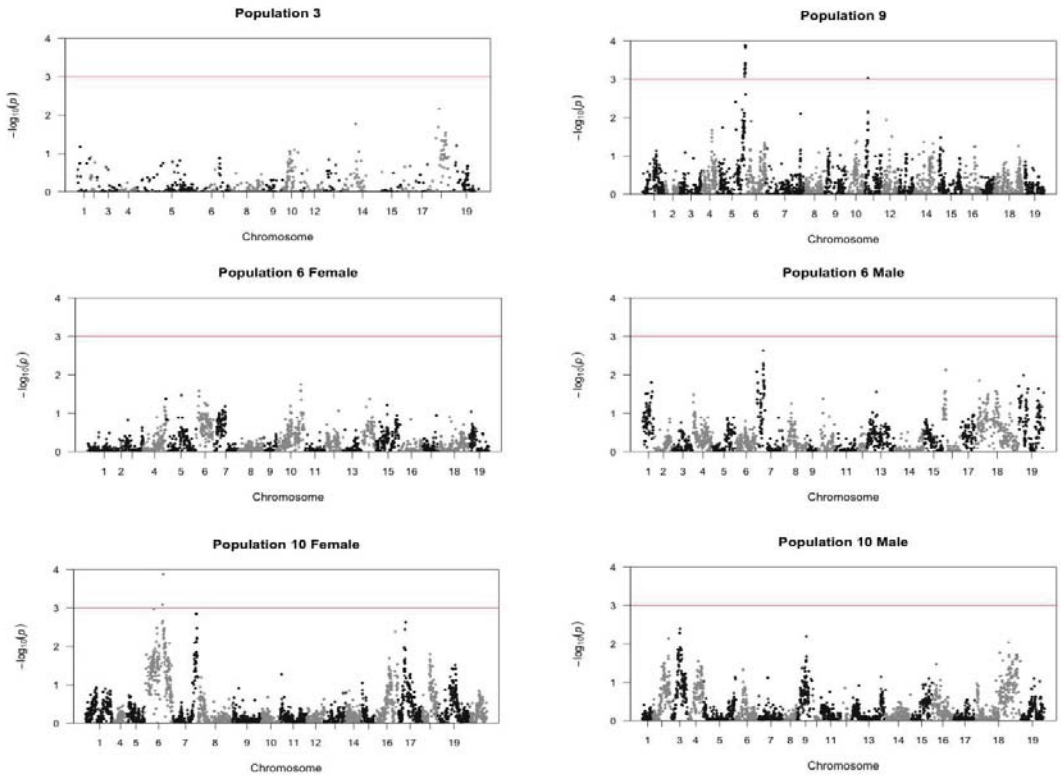


Figure 4.28: QTL analysis results for four populations in the 2013-2014 season for mature berry tartaric acid. There are significant SNPs with a LOD over three for population 9 and population 10 female.

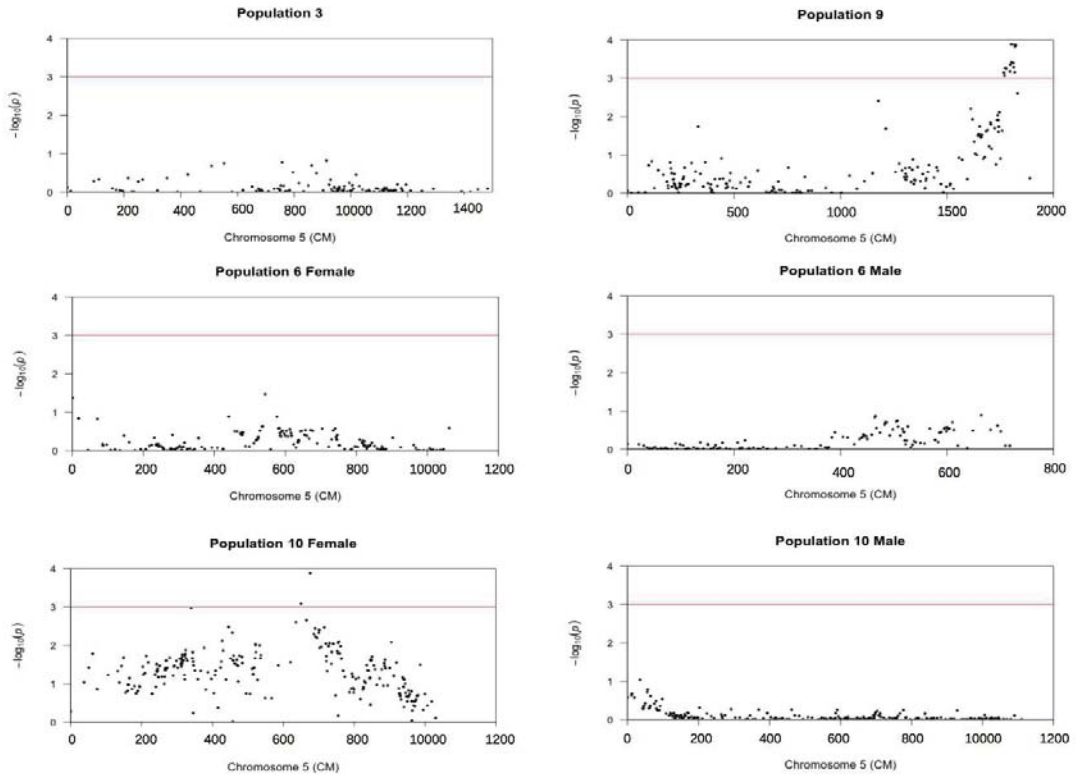


Figure 4.29: QTL analysis results for four population, chromosome 5, in the 2013-2014 season for mature berry tartaric acid. There are significant SNPs with a LOD over three for population 9 and population 10 female.

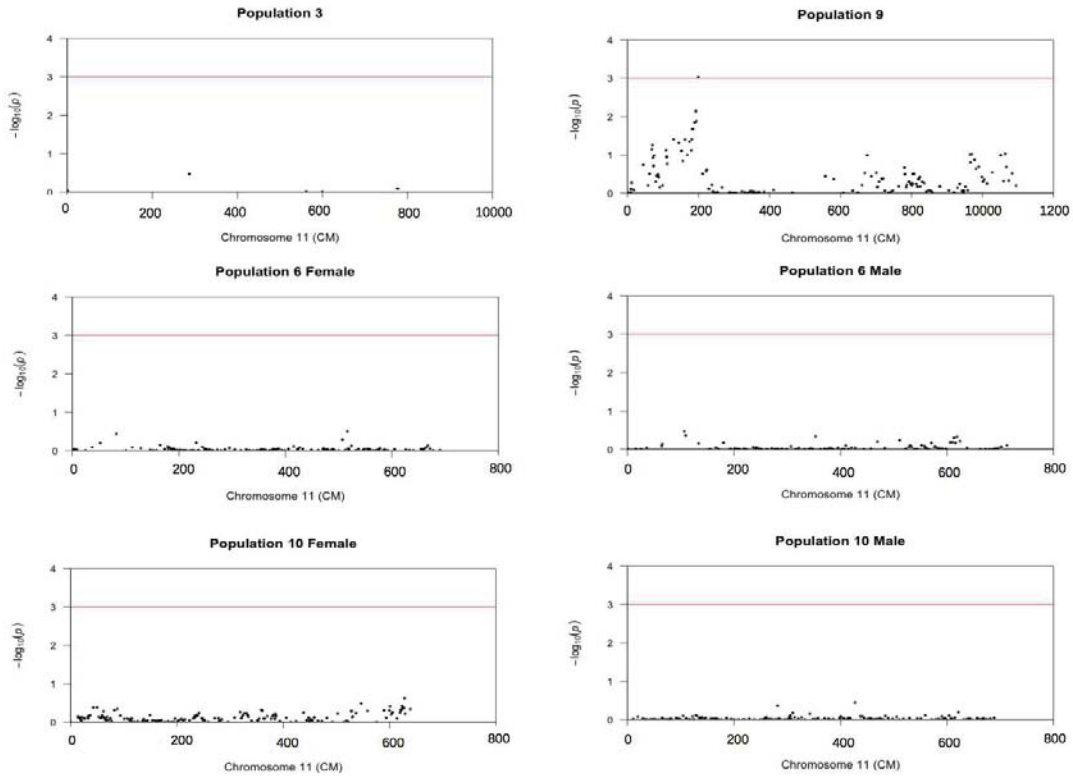


Figure 4.30: QTL analysis results for four population, chromosome 11, in the 2013-2014 season for mature berry tartaric acid. There are significant SNPs with a LOD over three for population 9.

4.4 Discussion

The long generation time for grapevines made it impossible to develop and characterise specific crosses for tartaric and malic acid during the period of this PhD study. The populations used for this study were developed for other studies but due to the number of individuals in each population it was thought they could be used in this study also. In the 2011-2012 season a subset of individuals from each of the ten populations was analysed, as well as the parents, to determine the variation in malic acid and tartaric acid within each population so that the best populations with the most variation could be chosen for sampling in subsequent years (see Chapter 3).

Missing data and errors created by the low sequencing depth of GBS methods can create problems in GWAS and QTL. Missing data can be imputed based on haplotypes and other related individuals but this was not done here as imputing without a full knowledge of the process may not have been useful. Instead sequencing depth was increased by re-sequencing the plates of prepared DNA fragments and merging this with the first sequencing run. The parents were also sequenced multiple times in each sequencing run, the female parent was sequenced eight times per run as it was common to each of the populations, and each male parent was sequenced twice per run. Increasing the number of times an individual is included in a sequencing run increases the coverage for that individual and the parents were considered important.

Many computer programs cannot handle very large SNPs sets so a smaller, higher quality set was determined to be better. To begin with there were over 80,000 SNPs between individuals in the four progeny populations, hence many were able to be discarded if they had missing data or a high error rate. Any SNPs left with impossible genotypes based on

the parent genotypes were re-placed with an NA and not analysed in either the GWAS or the QTL analysis. SNPs were also filtered on their usefulness in determining phase. Only SNPs that were heterozygous in one parent and homozygous in the other were retained. SNPs were filtered for each of the four populations separately. While the SNPs set for each population was only a fraction of the size of the original pool of SNPs, the most useful, highest quality SNPs remained for each. Population 10 had the largest number of SNPs with over 11,000, and population 3 had the smallest number with over 1,000. Due to the alignment process it is highly likely that the number of SNPs discovered is linked to the size of the population as population 10 is the largest population and population 3 is the smallest. After this stringent filtering there were still a large number of SNPs in our sets, especially for populations 10 and 6.

For GWAS the program Tassel was chosen because it was developed by the Buckler Laboratory at Cornell University for use in GWAS/GBS studies with large numbers of SNP markers (Bradbury et al. 2007). It was used by Barba *et al.* (2014) for discovering QTLs for resistance and susceptibility to powdery mildew in progeny of a cross between *V. rupestris* B38 and Chardonnay. For the QTL analysis, it was much more difficult to find a suitable computer program that could handle large numbers of markers. Many of these programs were developed before the development of NGS. The R package Onemap was chosen for the genetic mapping as it could handle population types other than inbred lines and relatively large numbers of SNP markers. The package qtl was used for the QTL analysis.

For the GWAS, a GLM was chosen due to limitations within Tassel. Mixed Linear Models (MLM) require a kinship matrix but this can only be generated in Tassel with homozygous

inbred lines and since the populations analysed here are heterozygous, MLM could not be used. Since GLM does not account for population structure any results may contain false positives (Fodor et al. 2014). To remove false positives from the GLM analysis the P-values generated were used in resampling-based multiple hypothesis tests using the Benjamini Hochberg method (Benjamini & Hochberg 1995).

For the GWAS, the same marker S19_151379 appeared in the mature berry tartaric acid analysis for both years and was the only significant marker in the 2012-2013 year (Table 4.3). For the QTL analysis, there are no markers that were significant in both years (Table 4.4). Environmental effects may play a large role in the difference between the two years' data. The environment is known to affect acid concentration and berry ripening rate. This is especially true for malic acid which is known to be influenced by temperature (Kliewer 1968). Both years had a large difference in acid amounts and had different climatic conditions over the growing season. To negate these effects, vine populations would need to be grown in a glasshouse or growth cabinet with controlled climatic conditions. However the large size of grapevines makes this difficult.

Other factors such as fruit maturity when picked also varied between the two years. It was impossible to collect all fruit at the same maturity while trying to juggle other factors such as climate variations. Instead berries from the whole population were picked on the same day to try and limit differences in climatic conditions, of particular concern would be heat waves that occur over the period of ripening. This difference in maturity would affect malic acid as it decreases after veraison unlike tartaric acid amount (mg/berry) which remains relatively stable.

For population 9 the same significant SNPs on chromosome 5 were seen in pre-veraison and mature berry tartaric acid analysis in the 2013-2014 season for the QTL and GWAS. SNP S19_105578867 appeared in the pre-veraison and mature berry tartaric acid analysis for the QTL and GWAS (Tables 4.3 and 4.4). However, in the QTL analysis this SNP was clustered in the linkage group for chromosome 5. It may be that this SNP was misaligned during the alignment to the genome process. Two SNPs on chromosome 7 for population 9 pre-veraison tartaric acid were also seen in both the QTL and GWAS and SNP S11_5230059 was seen in population 9 mature berry tartaric analysis for both the GWAS and the QTL analysis.

Of the two genes known to participate in tartaric acid synthesis L-IDH and 2-keto L-gulonate reductase, neither are located on chromosome 5, 7, 11 or 19. The L-IDH isoforms are located on chromosome 16 and 2-keto L-gulonate reductase is on chromosome 9. This means that there is the potential for a gene to be found within the QTL that was previously not known to participate in tartaric acid synthesis. Of the genes located in this region, according to NCBI, almost all of them have a predicted function based on similarity to other genes. When function cannot be predicted, the gene has an unknown function. Many genes in this region had an unknown function. There were also genes in this region that had a predicted function in DNA binding and transcription factors. These types of genes may be involved in regulation of the tartaric acid synthesis pathway.

Different significant SNPs appeared in different populations as each population had its own unique set of SNPs that were useful and error free in that population. For pre-veraison and mature berry malic acid population 10 female parent, population 9 and population 6 male parent had significant SNPs on chromosome 1 in the QTL analysis. While none of

these SNPs overlap they are within a 1MB region of each other and may be indicating a role of the same gene or genes in the variation in both mature berry and pre-veraison malic acid.

GWAS and QTL both have different limitations which explains why the QTL showed different and more significant markers than the GWAS. QTL analysis relies on recombination frequency of the markers while GWAS depends on phenotypic variation within a population (Deschamps et al. 2012, Korte & Farlow 2013). Using both techniques together can reduce some of the limitations associated with each technique separately.

To date, there is one other study using genetic mapping and QTL analysis of malic and tartaric acid (Chen et al. 2015). These authors used QTL analysis to determine genome regions involved in tartaric and malic acid concentration as well as sugar concentrations. From a population with a complex parental background, grown in China, they found QTLs on chromosome 18 and chromosome 6 for malic acid and no QTLs for tartaric acid. This contrasts with the results presented here in which several QTLs were found for tartaric acid on chromosomes 5, 7, 11 and 19. QTLs for malic acid were found on chromosomes 1, 3, 8, 9 and 18. The study by Chen *et al.* (2015) used only QTL analysis. A combination of GWAS and QTL analysis was used here which may explain why more QTLs were found in this study compared to Chen *et al.* (2015). A UHPLC-MS/MS method was used for measuring the tartaric and malic acid in the individuals of the populations, unlike Chen *et al.* (2015) who used HPLC with filtration to separate sugars and acids. The structures of the populations used could also explain the differences in the QTLs discovered in this study and in the study by Chen *et al.* (2015). In common with this study, Chen *et al.* (2015) also saw instability in their QTLs across the three years they measured and attributed

climate fluctuations as a cause.

Both GWAS and QTL have the potential to find genes that may be involved in malic acid and tartaric acid metabolism in grapevines. However, more work would need to be done to determine if any genes in the QTL regions found here are involved in malic and tartaric acid. The QTLs found on chromosome 5 for tartaric acid would need to be narrowed down to a smaller area containing fewer genes for further investigation as no genes stood out as being involved in tartaric acid when a gene search was conducted. Several genes in this region were transcription factors and DNA binding proteins and may be involved in regulation of tartaric acid. There was also a large number of genes in this region with uncharacterised functions, leading to the possibility of unknown genes participating in tartaric acid synthesis. To narrow this region down further, a larger population would need to be created, possibly kept in a greenhouse or growth room to lessen any climatic effects that could affect the discovery of QTLs. More GWAS studies could also be performed with additional vine populations. This may provide a narrower region on this chromosome or it may discount it completely.

In this study QTLs for tartaric acid were found which had not been achieved previously with the only other grapevine acid study finding no QTLs for tartaric acid but several for malic acid. If some of these significant SNPs can be positively identified as being linked with acid traits in these populations, they may be able to be used for marker assisted selection in breeding programs selecting for higher or lower malic or tartaric acid. However, more work would need to be carried out to determine if these markers are connected to acid traits and if they are also useful for other populations of vines.

Chapter 5

Tartaric acid concentration and L-Idonate dehydrogenase expression in grapevine tissues

5.1 introduction

While there is a large body of published data on tartaric acid concentrations in berry samples including the pattern of accumulation over development and how acid concentrations differ across different cultivars (Iland & Coombe 1988, Liu et al. 2006, Sabir et al. 2010, Munoz-Robredo et al. 2011, Pavlousek & Kumsta 2011, Preiner et al. 2013), little is known about tartaric acid concentrations in other tissues of the grapevine (see Chapter 1). There has been one study looking at tartaric acid concentration in leaves of V*itis vinifera* and other closely related species (Stafford 1959) and another examining the concentration in leaves, berries and other tissues (Kliewer 1966). Grapevines only produce fruit once a year and this makes the study of fruit traits such as acid metabolism difficult. The occurrence of tartaric acid metabolism in leaves may provide an alternative means of studying this unusual biochemical pathway.

It is known, from experiments feeding ^{14}C labelled ascorbic acid and ^{3}H labelled 5-keto-D-gluconic acid into berries and leaves, that grapevine leaves synthesise tartaric acid using the same pathway as berries (Saito & Kasai 1969, Saito 1994). Tartaric acid is formed in grapevines from ascorbic acid via L-idonic acid and D-gluconic acid; this differs from *Pelargonium crispum* which forms tartaric acid from ascorbic acid via L-threonate (Helsper & Loewus 1982) and *Phaseolus vulgaris* which forms tartaric acid from D-*xylo*-5-Hexulosonate (Saito & Loewus 1989).

This study explored the tartaric acid concentration in grapevine tissues other than the berry, which raised questions about the function of tartaric acid in the grapevine. The acid concentration in leaves was investigated in further detail to determine if leaf tartaric acid could be used as an estimate of berry tartaric acid. As grapevines have a long generation

time and only produce fruit annually, there are benefits in being able to estimate how much tartaric acid is expected in the berries before the vine has produced fruit.

The expression of L-IDH isoforms in different tissues of the grapevine was also studied to determine if the L-IDH concentrations varied between tissues. Understanding more about the expression of the three isoforms would further the understanding of the different roles of the three isoforms.

Of the three L-IDH isoforms in the grapevine genome only one isoform, now known as L-IDH 3, is known to participate in tartaric acid synthesis (DeBolt 2006). A second, L-IDH 2, has recently been found to be a sorbitol dehydrogenase (Jia et al. 2015). The final isoform is very similar to L-IDH 3 and has not yet been characterised (see Chapter 2). L-IDH mRNA and protein expression have been followed through berry development by the use of an L-IDH antibody, microarray analysis and qRT-PCR (DeBolt et al. 2006, Wen et al. 2010, 2014, Sweetman et al. 2012).

Wen *et al.* (2014) used an L-IDH antibody to determine the L-IDH protein expression in different wine grapes, table grapes, native Chinese grapes and hybrids. They found that the amount of L-IDH decreased over development in all cultivars tested and that this decrease was more pronounced in the native Chinese and wine grape varieties. This decrease was consistent with the accumulation of tartaric acid in these vines over development (Wen et al. 2014) and is also consistent with general published trends of tartaric acid accumulation in berries, which accumulates before veraison (Kliewer 1966, Iland & Coombe 1988, DeBolt 2006, Munoz-Robredo et al. 2011).

Sweetman *et al.* (2012) found that two of the three L-IDH isoforms, L-IDH 1 and L-IDH 3 were highly expressed in young fruit while L-IDH 2 was highly expressed from veraison onwards. This is consistent with what is known about tartaric acid accumulation and the role of L-IDH 2 as a sorbitol dehydrogenase. They also validated their results using qRT-PCR and found that L-IDH 1 does decrease in expression over development, being only highly expressed in early development.

Microarray data presented in Chapter 2 indicated that L-IDH 1 and 3 were highly expressed pre-veraison, with expression sharply dropping around veraison. This is consistent with the findings of other published works examining at expression of L-IDH isoforms (DeBolt 2006, Sweetman et al. 2012, Wen et al. 2014).

In this research, pyrosequencing was used to determine the percentage of each of the three isoforms expressed in different tissues of the grapevine. This may give us insights into the function of the two isoforms in tartaric acid synthesis since one isoform has yet to be fully characterised. Pyrosequencing is a DNA and RNA sequencing method which detects the pyrophosphate released during DNA or RNA synthesis and through a variety of chemical reactions is converted to light which can be detected (Ronaghi 2001). During the process each base is separately passed over the template and when a base incorporates into the sequence a signal is produced. To detect polymorphisms a pyrosequencing primer is used so that the 3' end hybridises a few bases from the polymophic site. Each of the different variants will give a different signal and can be easily distinguished. Since the three isoforms of L-IDH are very similar to each other, especially in the 3' region, a method of allele-specific expression profiling targeting one SNP in the sequence was used. Similar methods have been used for allele specific expression of genes in strawberries (Schaart et

al. 2005) and humans (Sun et al. 2005). However, it is believed that this is the first time it has been used to distinguish the levels of transcripts in genes with highly similar sequences.

5.2 Materials and Methods

Tissue sampling

Preliminary leaf sampling

Two leaf samples were collected at random from two V6 vines.

Leaf developmental study

Leaves were collected from shoots that had six to eight leaves present. Leaves were collected in sequential order with number 1 being the newest leaf next to the growing shoot (Figure 5.1). Leaves from four different cultivars were collected.

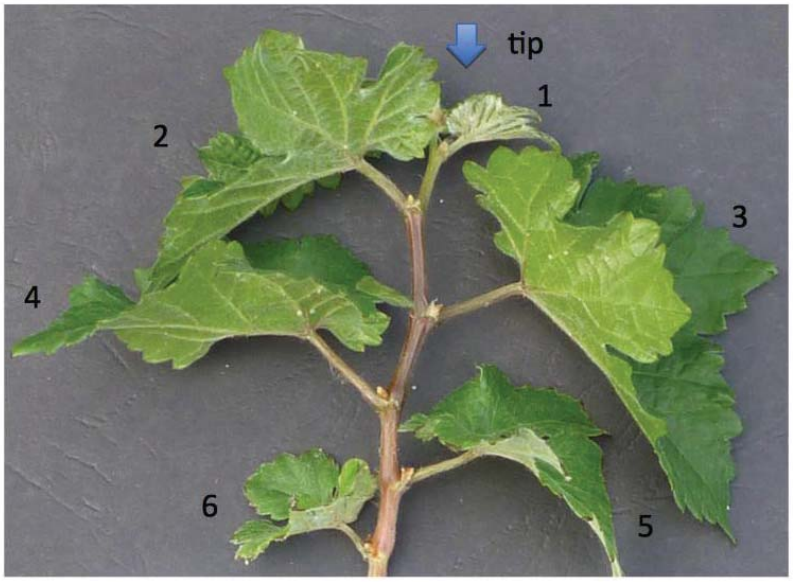


Figure 5.1: Example of a shoot for leaf collection. Leaves were collected from the youngest leaf closest to the growing tip (leaf 1) to the oldest leaf (leaf 6).

-

Leaf collection for measurement of leaf tartaric acid concentration to use as a predictor of berry tartaric acid concentration

Leaves were collected from ten individuals of progeny populations, grown in the field, during the 2013-2014 growing season. A sample consisted of two leaves that were the closest to the growing tip. Three separate samples per vine were collected for this analysis. Leaves were collected at the same time as mature berry samples (See Chapter 3). Leaves were stored at -40 ^{o}C

Berry collection

Berries were collected as part of the berry progeny populations experiment (see Chapter 3 for details). Acid concentrations from these berry samples were used here for comparison.

V6 and V10 tissue samples

Tissues were collected from two non-transformed microvine lines (V6 or V10) grown in the glasshouse or growth room. Pre-veraison berries were collected when the first few berries of the bunch had begun to soften. Four samples were collected for pyrosequencing. Mature berries were collected three to four weeks after the pre-veraison samples. Four samples were collected for pyrosequencing. Roots were dug up and collected; all roots were washed to remove excess soil and then dried before freezing. For the stem samples, shoots with four to six leaves were selected, and the leaves were removed before freezing. For the leaf samples the first few leaves below the shoot tip were collected. All samples were snap frozen in liquid nitrogen and then stored at – 80 ^{o}C.

Preparation of tissue samples

Acid extraction protocol

The acid extraction protocol is the same as the protocol used for berries (see Chapter 3 for details). All samples were stored at -20ºC until use. For the samples from the V10 vine used for cDNA synthesis, samples were ground to a fine powder using a mortar and pestle so that the same sample could be used for tartaric acid concentration measurement and RNA extraction. Tissue was stored at -80 ºC.

UHPLC-MS/MS

Dilutions for UHPLC-MS/MS run

All samples were diluted 1:200 with sterile water in a 96-well plate using a robot. Some samples were re-diluted by hand to a 1:500 dilution as the acid concentration was found to be outside the limits of the MS at a 1:200 dilution.

The UHPLC-MS/MS procedure was the same as the one used for berry samples. See Chapter 3 for details.

RNA extraction

See Chapter 2.2 for details

cDNA synthesis

For cDNA synthesis 1 µg of RNA in 12.5 µl of H_2O was added to 10 mM dNTPs and 100 µl of oligo-dT primer. This mix was incubated at 65 ºC for five minutes then on ice for one minute. To this mix was added 5x first strand buffer, 0.1 M DTT and 1 µl of superscript III (Invitrogen). The mix was incubated for one hour at 50 ºC and the superscript III was then

deactivated at 70 ºC for ten minutes. The cDNA was diluted in 180 µl of H_2O.

Pyrosequencing

A pyrosequencing method was selected to determine the levels of each of the L-IDH isoform transcripts in grapevine tissue and berry samples. This method was chosen since it could distinguish L-IDH 1 and L-IDH 3 transcripts from each other, a differentiation which could not be achieved using qRT-PCR due to the high sequence similarity at the 3' end of the transcripts.

The pyrosequencing was conducted by the AGRF (Perth, WA node). The sequence of the last 200 base pairs at the 3' end of each of the three isoforms was sent to the AGRF for primer design. This region of the sequences was very similar having only one SNP, at position 1,004bp, which differed between the three alleles (L-IDH 1 = A, L-IDH 2 = C and L-IDH 3 = G). Primers were designed and tested by the AGRF. For testing and validation, cloned cDNA sequences of the three isoforms were pyrosequenced individually and in mixtures of set amounts to determine if the pyrosequencing method could distinguish the three isoforms. Once verified, the cDNA from each tissue (stems, roots, leaf and the two berry time points) was sent for the pyrosequencing. To confirm the initial results the pyrosequencing was preformed a second time with triplicate samples of pre-veraison and mature berry samples.

5.3 Results

5.3.1 Tartaric acid concentration and L-idonate dehydrogenase expression in grapevine

Grapevine tissue samples were collected from two different microvines, V6 and V10, grown in a growth cabinet or in the glasshouse. The samples collected were root, stem and leaf. Berry samples were collected at a pre-veraison and a mature berry time point, for comparison to the three other tissues. For the V6 microvine, a sample of embryogenic callus was also collected; this could not be collected from the V10 vine as embryogenic callus was not created from this line.

The tartaric acid concentration in each of the samples was measured using the UHPLC-MS/MS method developed for measuring acid concentrations in berries (see Chapter 3) (Figure 5.2 and 5.3). For the V6 embryogenic callus sample there was no detectable amount of tartaric acid. For the V6 and V10 root samples, tartaric acid was very low in both lines. For the stem samples, the concentrations for both lines were high compared to the pre-veraison berry sample. For V6 this was the tissue with the second highest concentration of tartaric acid whereas for V10 it was the sample with the highest concentration. For leaf samples, there was a pronounced difference in the tartaric acid concentrations between the two lines. For V6, the leaf acid concentrations were relatively low, whereas in the V10 sample the leaf tartaric acid concentration was high, almost as high as the pre-veraison berry sample. It is known that tartaric acid stops accumulating in the berry around veraison and that as the berry grows, post-veraison, the acid concentration decreases. Therefore, it is unsurprising that pre-veraison berry tartaric acid concentrations in these samples were so high for both V6 and V 10 lines or that the acid concentration in the mature berry samples was lower. It is also interesting to note that the

acid concentrations in the V10 line were much higher than in the V6 line.

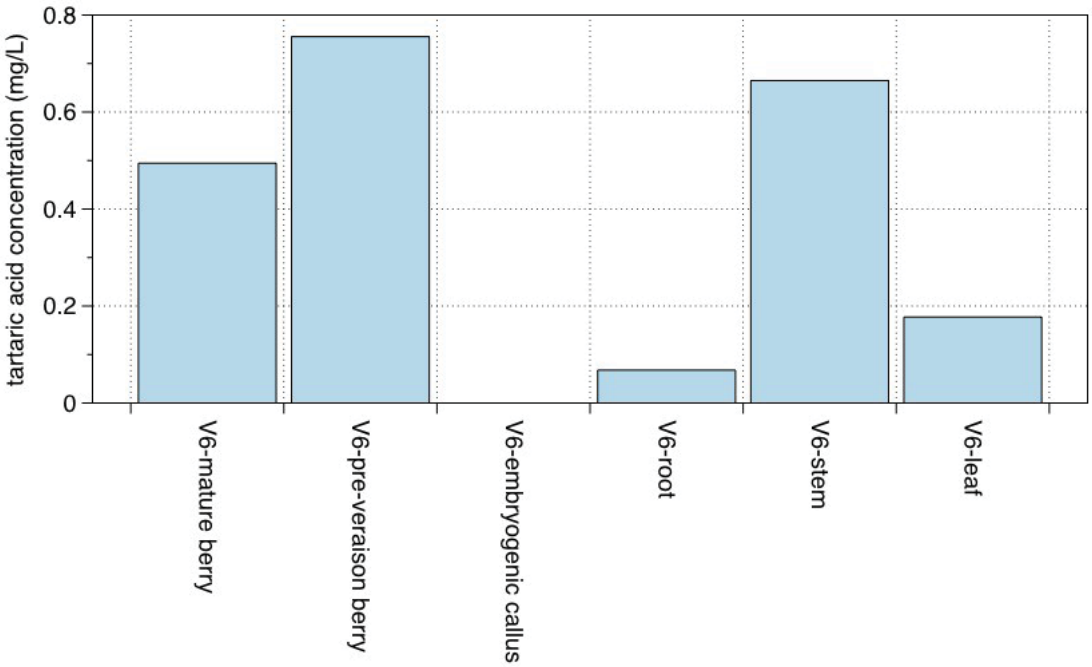


Figure 5.2: Tartaric acid concentrations in mg/L from different tissues of the V6 microvine line. Two berry samples are included from different time points in berry development.

L-IDH is a gene known to participate in tartaric acid synthesis. One of the isoforms has been characterised and is known to participate in the pathway, however the other two isoforms have not been characterised and very little is known about them. At first qRT-PCR was attempted with the aim of determining the differences in expression of the three isoforms across different tissues in the grapevine. However, isoforms L-IDH 1 and L-IDH 3 could not be separated as their 3' regions are very similar and no primers could be created that were able to distinguish them. Hence, a pyrosequencing method was used to determine the percentage of each isoform present in a sample.

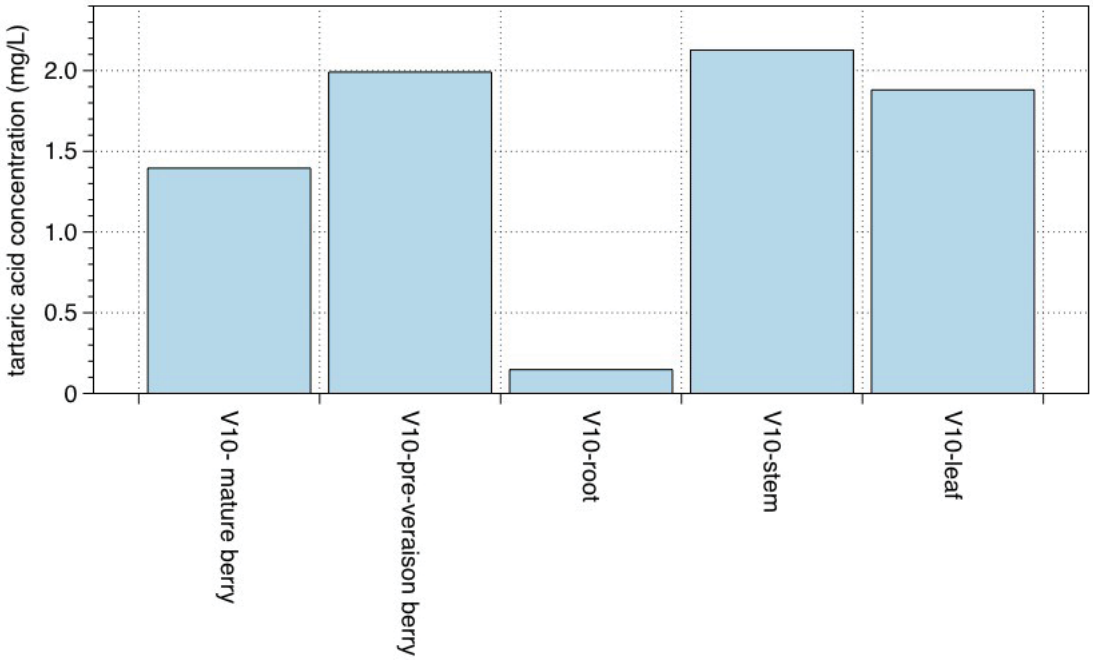


Figure 5.3: Tartaric acid concentrations in mg/L from different tissues of the V10 microvine line. Two berry samples are included from different time points in berry development.

The method developed used a polymorphism at a single site (position 1,004) at the 3' end of each isoform sequence. Root, stem and leaf tissue were examined in addition to two berry samples, one from a pre-veraison time point and the other from a mature berry time point (Table 5.1). L-IDH 2 was the most expressed isoform in roots and had a relatively consistent expression level in both berry time points. It was the lowest expressed isoform in leaves and had similar expression to L-IDH 3 in stem tissue.

L-IDH 1 was the highest expressed isoform in leaves, stems and pre-veraison berries (Table 5.1). by contrast, L-IDH 2 and L-IDH 3 were the highest expressed isoforms in mature berries in almost equal percentages. Both L-IDH 1 and 3 had low expression in roots compared to the L-IDH 2 isoform.

Table 5.1: The percentage of each L-IDH isoform transcript in different grapevine tissues based on the SNP located at position 1,004.

Grapevine tissue	Percentage of isoform		
	L-IDH 1 (A)	L-IDH 2 (C)	L-IDH 3 (G)
Roots	18	79	3
Stem	75	14	11
Leaf	49	16	36
Pre-veraison berries (range for four samples)	50-69	30-35	14-18
Mature berries (range for four samples)	9-13	41-61	31-50

5.3.2 Tartaric acid concentration in leaves as a predictor of berry tartaric acid concentrations

The leaves from two microvines from the V6 line were collected for a preliminary study of leaf acid. V6 was selected as this vine was used to create the V6 over-expression microvines described in Chapter 2. It was thought that a base level of acid in leaves could be determined, which could then be used to determine if over-expression of L-IDH produced higher concentrations of tartaric acid. However, the leaf tartaric and malic acid concentrations in these two similar vines were very variable (Table 5.2). Leaf samples from vine 2 contained acid concentrations higher than any berry samples, with the result that the acid concentrations in further leaf samples were measured to determine if there any were errors in the measurements and to see if the age of the leaf made any difference in the acid concentration. Since the leaf acid was so variable in these two samples it was decided to collect a series of leaves across developmental stages to see if the age of the leaf had any influence on tartaric or malic acid concentration.

Table 5.2: Preliminary leaf sample study shows that leaf acid concentrations (mg/ml) are highly variable and in one case higher than berry acid concentrations (mg/ml).

Leaf sample	Preliminary leaf samples		Range of berry samples	
	Malic acid mg/ml	Tartaric acid mg/ml	Malic acid mg/ml	Tartaric acid mg/ml
1	1.89	3.87	0.01-2.29	0.83-3.34
2	5.72	8.24		

Leaf samples were removed from a shoot starting with the first and second leaves after the growing tip, down to the last (oldest) leaf (Figure 5.1). Leaves were collected from four different vines. The malic and tartaric acid concentrations were measured using UHPLC-MS/MS. There appeared to be no correlation between acid concentration and leaf age, and

each vine exhibited a different pattern of acid concentrations across the leaf ages (Figure 5.4). Vine 3 had very high concentrations of both tartaric acid and malic acid in leaf 4.

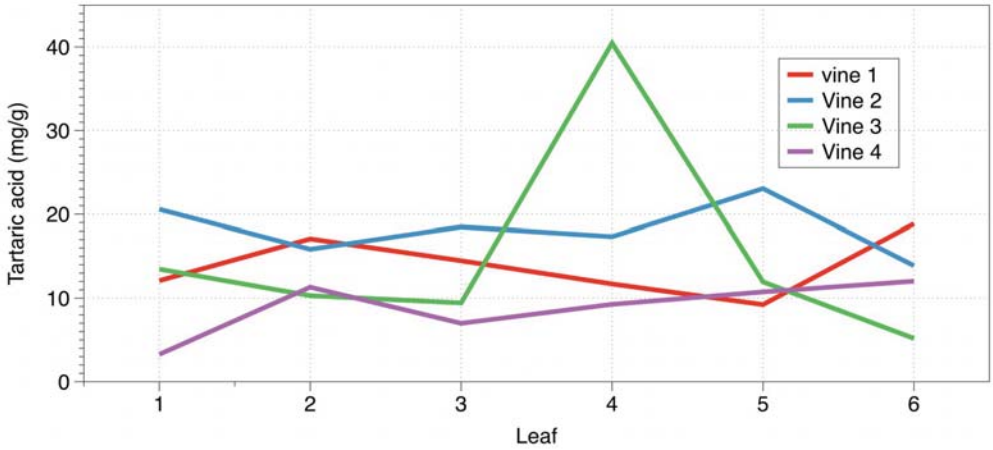


Figure 5.4: Tartaric acid concentrations in mg/g from leaves of different ages collected off the same shoot. Shoots were collected from four different vines. 1 corresponds to the youngest leaf and 6 corresponds to the oldest leaf from the shoot tip.

Three leaves were collected from ten individuals from one progeny population to determine whether leaf acid could be used as a predictor of berry acid. The acid concentrations in these leaves were measured individually using the UHPLC-MS/MS method, then averaged. This average was compared to an average of the mature berry acid concentration across the two years for each individual. Leaf tartaric and malic acids vary more than mature berry acid concentrations with a standard deviation for malic acid in leaf samples of 1.04 mg/g compared to 0.36 mg/g for berry samples (Table 5.3). For tartaric acid the standard deviation is 2.55 mg/g for the leaf samples and 0.67 mg/g for the berry samples. This was then compared to the average mature berry tartaric acid amounts from these individuals from the 2012-2013 and 2013-2014 seasons (see chapter 3 for mature berry tartaric acid amounts) (Figure 5.5).

Table 5.3: Tartaric and malic acid concentrations (mg/g) in leaf samples from a sample of individuals from population 9. The concentration was measured in three leaves for each vine, except vines 2, 5 and 9 which had measurements from two leaves only.

Vine No.	Tartaric acid (mg/g)
1	19.03-24.86
2	25.27-28.93
3	21.75-23.76
4	22.19-25.49
5	22.00-35.21
6	24.06-25-63
7	24.15-26.22
8	22.03-26.66
9	22.80-23.21
10	17.76-21.05

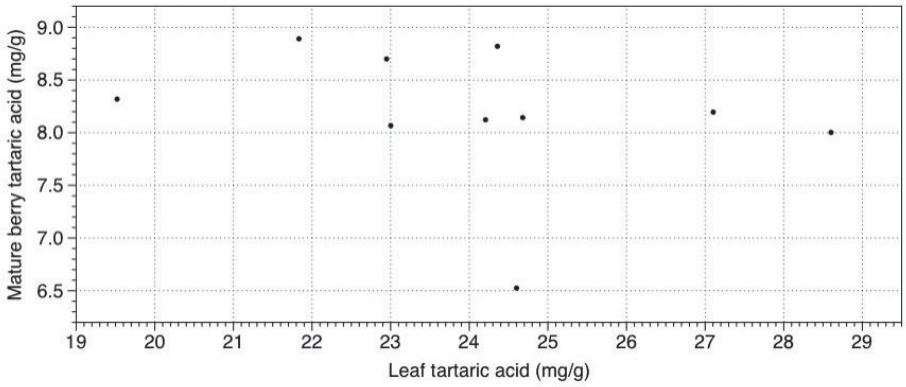


Figure 5.5: Leaf tartaric acid and mature berry tartaric acid from selected individuals of population 9.

5.4 Discussion

Tartaric acid concentration in grapevine tissues

Although there are several published studies measuring tartaric acid concentrations in grapevine berries (Iland & Coombe 1988, Liu et al. 2006, Sabir et al. 2010, Munoz-Robredo et al. 2011, Pavlousek & Kumsta 2011, Preiner et al. 2013) (see Chapter 1 for further details on these references), only one published study has measured the concentration in leaves (Stafford 1959) and another has measured acid concentrations in berries, leaves and other grapevine tissues (Kliewer 1966).

In the data presented here, the tartaric acid concentrations in roots, shoots, leaves and embryogenic callus were measured and compared to the concentrations in pre-veraison and mature berry samples (Figures 5.2 and 5.3). Acid concentrations in roots were low in the two microvine lines tested while acid concentrations in leaves were high in one line and low in the other. Tartaric acid concentrations in the stems were high in both lines and in one line the concentration was higher than the pre-veraison berries. Kliewer (1966) measured tartaric acid concentrations (g/100g FW) in multiple tissues from the grapevine at multiple time points. He found that green berries had the most tartaric acid and that acid in berries reduced over ripening; this is consistent with other reports (Iland & Coombe 1988, DeBolt 2006, Munoz-Robredo et al. 2011) and with what is shown here. Kliewer (1966) also reported low concentrations of tartaric acid in roots and which again is consistent with our data. Acid concentration was measured in mg/L in this study, however, this does not take into account the differences in sizes of the different tissues and future studies may wish to measure concentration in mg/g of fresh weight or measure leaf acid concentration in mg/mm^2. This was just a small study of the acid concentration in

grapevine tissues and a larger study may need to be undertaken in the future to determine if there is any link between acid concentration in these tissues and berry maturity.

Why grapevines contain tartaric acid in tissues other than the berry is still unknown as the full function of tartaric acid in the berries is unknown. It may be that human breeding of the grapevine over time has produced vines with high tartaric acid because it is very useful for winemaking. However, this increase in tartaric acid was not just targeted at the berry and a side effect is high tartaric acid concentration elsewhere in the vine.

Tartaric acid concentration in grapevine leaves

The two preliminary samples showed that the concentration of leaf tartaric acid was highly variable, with one sample having a much higher tartaric acid concentration than pre-veraison berries (Table 5.2). More leaf samples were collected from differently aged leaves from a single shoot to determine whether there was a difference in tartaric acid concentration over different leaf ages. All vines had a diverse pattern of tartaric acid accumulation across development and all leaves from the one shoot had different concentrations of tartaric acid with vine 3, leaf 4 having very high tartaric acid (Figure 5.4). Kleiwer (1966) compared the concentration of tartaric acid in leaves to the leaf area to determine if there was any pattern of tartaric acid accumulation over the growing season. He found that concentrations remained relatively constant across the growing season with a decrease at the end of the season as leaves became senescent. The results presented here are more variable. Three of the four vines tested (Vines 1,2 and 4) showed a relatively constant level of tartaric acid, and vine 3 showed a relatively constant concentration except for leaf 4 which showed extremely high concentrations of tartaric

acid. The possibility that a mistake was made in the determination of acid concentration in this sample cannot be ruled out and if this was the case, would most likely have occurred in the sample weight measurement. Another possibility is that any number of environmental factors, such as water status, light intensity and leaf damage, may have created variation in acid concentration in this leaf sample.

Tartaric acid was measured from the leaves of a selection of individuals from one progeny population and compared to mature berry tartaric acid concentrations to determine if leaf tartaric acid could be used as a predictor of berry tartaric acid at maturity. As grapevines only produce fruit annually and can take several years to produce fruit for the first time, it would be beneficial to have an indicator of how high the acid concentration in the berries might be before the vines produce fruit. This would be especially useful in grapevine breeding programs. However, there does not appear to be any correlation between the acid concentration in berries and leaf tartaric acid concentrations (Figure 5.5). For tartaric acid, the means were 24.09 mg/g for leaf tartaric acid and 8.18 mg/g for berry tartaric acid. Two individuals had above average tartaric acid concentrations for both the leaf and berry while one had below average tartaric acid concentrations. However, one sample had a berry tartaric acid concentration that was much lower than the others; when this sample was excluded the mean berry tartaric acid concentration was 8.36 mg/g. Two samples had below average leaf and berry tartaric acid concentrations, and one sample had high leaf and berry tartaric acid concentrations. This indicated that, in most samples, leaf acid could not be used to predict whether the berries would have high or low acid for both tartaric or malic acids. This is not surprising. Very little information is known about the regulation of the tartaric acid synthesis pathway in berries or leaves. It is known that tartaric acid is produced in the same pathway (Saito 1994) in both tissues but whether they are regulated

in the same way is unknown. Berries do have a very controlled regulation of all processes across development, with veraison acting as a very quick switch (Coombe 1992). Several papers looking at gene expression over berry development have reported the sudden increase or decrease of expression of genes involved in may important processes of the berry rapidly changing at veraison (Deluc et al. 2007, Pilati et al. 2007, Lund et al. 2008, Sweetman et al. 2012).

Expression of the three L-IDH isoforms in different tissues of the grapevine

In Chapter 2 it was determined that there are three L-IDH isoforms present in the grapevine genome. After sequencing of the three isoforms it was decided to measure the expression of the three isoforms. As qRT-PCR had failed to distinguish all three transcripts due to their similarity in sequence at the 3' end of the coding sequence pyrosequencing was used to determine the percentage of transcript of the three L-IDH isoforms present in different tissues of the grapevine.

L-IDH 2 was the most abundant of the three transcripts in the roots and also had a very consistent level in both berry samples. This aligns with what is known about L-IDH 2 (Table 5.1). From phylogeny studies and molecular evolution analysis, it is thought that L-IDH 2 is actually a sorbitol dehydrogenase (Jia et al. 2015). In apples there are many sorbitol dehydrogenase genes which are expressed in many tissues such as the flowers, fruits, stems and roots (Park et al. 2002, Wu et al. 2010). In the fruit the expression of the sorbitol dehydrogenase genes increases as fruit ripens (Park et al. 2002, Nosarzewski & Archbold 2007, Wu et al. 2010). In berries the sorbitol dehydrogenase (formally L-IDH 2) also increases in expression as the berry ripens as seen by published expression studies

(Sweetman et al. 2012) and the microarray data presented in Chapter 2.

It is unlikely that the transcript levels of L-IDH 1 and 3 have ever been measured separately before. Several studies have measured the transcript and protein expression levels across development. Wen *et. al.* (2010, 2014) created an antibody to L-IDH 1 and used it to look at protein amounts across development but could not differentiate between L-IDH 1 and L-IDH 3. Sweetman *et. al.* (2012) used RNAseq to look at the expression of transcripts across development but it is unlikely that the alignment process used could differentiate between L-IDH 1 and L-IDH 3 due to their high sequence similarity. This is the same as the microarray data presented in Chapter 2. By developing a pyrosequencing method the levels of the L-IDH transcripts for the three isoforms could be measured separately. This method could be used to measure the amount of transcripts in other genes families where the gene sequences have high similarity to each other. There is some background noise created in the pyrosequencing as it is rare that one allele is present at 100% concentration even when measuring a pure sample of one isoform. For example a pure sample of L-IDH 2 might have a 97 % allele concentration. L-IDH 2 had the least noise of the three isoforms and L-IDH 3 had the most. It was expected that there would be at lease 5 % of the A (L-IDH 1) in each L-IDH 3 samples according to control tests on pure L-IDH 3 samples. This could be due to noise or that the amplification of L-IDH 3 was not completely specific and always amplified a small amount of L-IDH 1 as they are very similar in sequence. However, the changes between pre-veraison and mature berry samples are much higher than 5% indicating a real change in the concentrations of the isoforms.

In the studies mentioned in the previous paragraph, the expression of L-IDH 1 and L-IDH 3 together is high pre-veraison when tartaric acid is being synthesised and the expression

then drops sharply around veraison. However, this was not observed for both isoforms when they were measured separately in this study (Table 5.1). L-IDH 3 had the lowest level in pre-veraison berries and was higher in mature berries, the opposite of what was expected if the expression was to mirror tartaric acid synthesis. L-IDH 1, however, was highest in pre-veraison berries, stems and leaves and lower in mature berry samples as expected. The involvement of L-IDH 3 in tartaric acid synthesis is known, however there is no conclusive proof that L-IDH 1 participates in tartaric acid synthesis or is an active protein. Therefore, more work needs to be conducted to show that L-IDH 1 is involved in tartaric acid synthesis and investigate the contribution of L-IDH 1 and L-IDH 3 to tartaric acid synthesis. These results raise interesting questions about the regulation of tartaric acid which is also unknown. It may be that levels of L-IDH 1 dropped significantly in the mature berry sample which then increased the percentage of L-IDH 3. Or it may be that L-IDH 3 is actually stable across all stages of development and that L-IDH 1 is regulated and decreases at veraison, however, this is speculation.

A larger study of L-IDH 1 and 3 needs to be conducted. As suggested in Chapter 2 a larger study of the sequence of these two isoforms needs to be conducted in different cultivars of grapevine with varied maximum concentrations of tartaric acid. This could be coupled with a more in depth look at the expression of L-IDH 1 and L-IDH 3 in these cultivars to determine if there is a link between sequence differences, maximum tartaric acid concentration and expression patterns of the two isoforms. It would be especially important to look at cultivars with a range of tartaric acid concentrations. Another future study could focus on the promotors of L-IDH 1 and L-IDH 3 which have never been studied before. The differences in the promoters may shed some light on how L-IDH 1 and L-HD 3 are differently regulated. These differences in promotor sequence could then be

linked back the expression pattern of the genes in different cultivars and at different times of berry development.

Tartaric acid was found to be present in all tissues of the grapevine except the embryogenic callus. In leaves the concentrations varied considerably and in some cases were higher than berry tartaric acid. L-IDH was expressed in the four tissues examined using a pyrosequencing method that could distinguish the expression of L-IDH 1 and L-IDH 3 separately. Over the four tissues examined the three isoforms had differing levels of expression. This is especially important for L-IDH 1 and L-IDH 3 and shows that there may be regulation occurring within berries and across different grapevine tissues.

Chapter 6

Conclusions and future directions

6.1 Introduction

Tartaric acid and malic acid are the two most abundant acids in grape berries, contributing to 90 % of acidity (Ruffner 1982). Malic acid has been extensively studied in grapes and other fruits as it is a very common acid across the plant kingdom, with an important role in metabolism. Tartaric acid has not been well studied because it is not a major acid in fruits, except grapes, and does not have an apparent role in metabolism. However, tartaric acid is more important for pH balance in wines than malic acid, as it has a more pleasant taste and is not broken down in malo-lactic fermentation. Tartaric acid is relatively stable throughout berry development whereas malic acid decreases in concentration (Iland & Coombe 1988).

While the pathway of tartaric acid synthesis in grapes is known, little is known about the genes that participate in and regulate this pathway during berry development. So far only two genes have been shown to have some involvement in the tartaric acid synthesis pathway, of which L-IDH is one. When this research began three *L-IDH* genes were identified in the grapevine genome based on *in silico* annotation. By investigating the relevant databases for the grapevine genome on NCBI, it was shown that the three L-IDH isoforms are in tandem on chromosome 16 (Figure 2.1). *L-IDH 1* and *L-IDH 3* are very similar in sequence and *L-IDH 2* is more diverse in sequence and annotated in the opposite direction.

6.2 There are three L-idonate dehydrogenase isoforms in the grapevine genome

In the work described in Chapter 2, comparisons were made between the amino acid sequences of the three L-IDH isoforms. Only three amino acids were different between L-

IDH 3 and L-IDH 1 (Figure 2.2), but despite these small differences this work was able to show that there are three genes encoded in the grapevine genome. The amino acid sequences of L-IDH 3 and L-IDH 1 were compared to the original L-IDH characterised by DeBolt *et al.* (2006) and it was found that L-IDH 3 was the closest isoform in sequence to the original enzyme (Figure 2.4). PCR amplification of the L-IDH genes from gDNA of two homozygous lines and their wild type parent were used to determine if these two isoforms were two different loci in the genome or had been annotated incorrectly and were alleles of the same gene. As the L-IDH 1 and L-IDH 3 genes were both present in the two homozygous lines and were different sizes it was concluded that they are two different loci (Figure 2.5). Sequencing of the two isoforms confirmed this as it was discovered that there were many differences between the two sequences. There were also differences between the sequences and the published reference genome for both isoforms. Most of these differences were contained in the introns and could possibly create differences in regulation (Belostotsky & Rose 2005). Future work could examine these sequence differences in the introns and determine if they do regulate expression.

The protein activity assays presented in Chapter 2 continue on from work by DeBolt *et al.* (2006). Since this early research was conducted, two new isoforms of L-IDH have been discovered. Recent evidence has been published by Jia *et al*. (2015) suggesting that L-IDH 2 is likely to be a sorbitol dehydrogenase (closely related to L-IDH) and not an L-idonate dehydrogenase. The protein activity of L-IDH 1 was explored to determine if it was involved in tartaric acid synthesis. Protein activity assays proved inconclusive as only small amounts of protein could be purified (Figure 2.7). Further work is needed on the optimisation of L-IDH 1 protein expression. While a small amount of protein was purified, there was not enough to test its activity. The final method of protein expression that was

used here managed to generate a small amount of protein, however this method was not optimised. Optimisation of this method could increase yields of protein to such a level that if could be used for protein activity assays to test the activity of L-IDH 1 in the future.

The function of the L-IDH isoforms was also tested through the creation of transgenic microvines which over-expressed each of the isoforms. The aim was to determine if over-expression led to an increase in tartaric acid concentration in berries. Higher acid was only seen in two of the four L-IDH 3 over expressing plants and only in berries at the pre-veraison time point (Table 2.15 to 2.18). L-IDH is involved in the rate limiting step of the pathway, so it was hypothesised that over-expression of L-IDH may lead to an increase in tartaric acid. However, the regulatory mechanisms of the tartaric acid synthesis pathway are not known in grapevines or other species.

Since results were not conclusive as to whether over-expression of L-IDH in microvines leads to an increase in tartaric acid in berries further work is needed. ^{14}C bunch feeding experiments may be needed to trace the intermediates and determine if they increase in concentration in the transgenic vines. The measurement of the intermediates by HPLC or UHPLC-MS/MS could also be attempted to determine if any of the intermediates are increased due to over-expression of L-IDH, especially 5-keto-D-gluconic acid. This was not attempted in this work because of the lack of L-idonate standards. L-IDH protein levels could also be investigated in these transgenic microvines to determine if there is an increase in protein. This could be done with the use of an L-IDH antibody which has been done before to measure L-IDH protein levels (Wen et al. 2010, 2014).

While the microvine has been used in several studies (Dunlevy et al. 2013, Fernandez et al.

2013, Rienth, Romieu, et al. 2014, Rienth, Torregrosa, et al. 2014), it is believed that, to date, this is the first reported work to transform the microvine to investigate gene function (Chaib et al. 2010). Its small stature and continuous fruiting make it ideal for transgenic studies in the future. As well as over-expression studies L-IDH knockouts could be created using the microvine with RNAi.

There is a large number of published studies documenting the differences in acid concentration (and amount in mg/berry) in different species of *Vitis* and cultivars of *Vitis vinifera* (Kliewer et al. 1967, DeBolt 2006, Liu et al. 2006, Sabir et al. 2010, Munoz-Robredo et al. 2011, Pavlousek & Kumsta 2011). This variation is thought to have environmental aspects, such as affects by temperature (Kliewer 1968, Sadras et al. 2013, Sweetman et al. 2014) and sun exposure (DeBolt et al. 2008), as well as genetic aspects (Duchene et al. 2014). Before the genetic aspects of the variation can be studied, the tartaric and malic acid must be measured in the individuals of the populations included in the study. Variation was also studied in individual berries from the same vine to determine if there was any variation and how much variation there exists within one vine.

6.3 Variation in tartaric and malic acids in individual berries from the same vine and samples of multiple berries from different vines

In the work described in Chapter 3 the concentrations of tartaric and malic acid were measured in single berries from three bunches from two vines. Variation was seen in the concentration of both acids between berries from the same bunch, in the mean acid concentration of bunches, and in the total acid concentration from each vine (Figure 3.1). Using this data the variance components were calculated for vine, berries and bunches (Table 3.1). The standard error of tartaric and malic acid concentration measurements for

different sample sizes from 3 to 50 berries was then estimated (Figure 3.2). As expected, when the sample size increased, the standard error of the measurement decreased.

The tartaric and malic acid concentration was measured in a total of 60 individual berries at one time point. Future studies could include more berries and examine the variation in acid concentration between berries from different positions in the bunch, between berries in different positions on the vine and between different cultivars. Berry sampling at multiple time points could be incorporated into the studies to determine if there are points in development when the amount of variation is larger or smaller than at harvest. All of these additional measurements could determine if the ideal sampling size estimated here holds true throughout berry development and for other cultivars.

While the concentration of tartaric acid and malic acid had been measured in many different species of *Vitis* and cultivars of *Vitis vinifera* (Kliewer et al. 1967, DeBolt 2006, Liu et al. 2006, Sabir et al. 2010, Munoz-Robredo et al. 2011, Pavlousek & Kumsta 2011) this is the first study showing the variation in tartaric and malic acid in berries from the same vine. The acid concentrations from the individual berries were then used to estimate the standard error in the acid measurement by collecting subsets of berries from a single vine. This information could be useful for designing future studies, such as developmental series, where all the fruit from one vine cannot be collected in a single sample at a single time point.

Tartaric and malic acid concentrations were also measured in samples of individuals of ten progeny populations at a pre-veraison and mature berry time point in the 2011-2012 season (Tables 3.2 and 3.3). From this data four progeny populations were chosen for further

sampling based on the variation of both acids and the number of individuals in the population. All individuals in these four populations were sampled at the two time points across two seasons (2012-2013 and 2013-2014) (Tables 3.4 to 3.7). Variation was found between the individuals in each progeny population, between the progeny populations and between the two seasons for both acids. The data from these four populations were used for GWAS and QTL analysis looking for genes involved in acid metabolism and is detailed in Chapter 4.

Tartaric acid is known to be synthesised before veraison and is relatively stable across the remainder of berry development. However, this study found a difference in tartaric acid amount (mg/berry) between the pre-veraison and mature berry samples in both years. The most likely cause of this difference is that the pre-veraison samples were collected before tartaric acid had finished accumulating. In future experiments data should be collected across berry development to determine (after the event) when the maximum tartaric acid concentration occurs. Many published studies have documented the concentrations of tartaric acid in berries from one vine across development but they found a large variation in the flowering time and date of the maximum tartaric acid concentration from year to year (Melino et al. 2009). Therefore it would be advantageous to collect berry samples across development every season.

6.4 GWAS and QTL analysis to find genes involved in the variation of tartaric and malic acids in four progeny populations of vines

In the research described in Chapter 4, acid concentration data from four progeny populations, along with SNP markers found through genotyping by sequencing was used for GWAS and QTL analysis. In the GWAS, significant SNPs (adjusted P-value lower than

0.05) were found on chromosome 19, chromosome 7, chromosome 11 and chromosome 5 for tartaric acid in population 9 (Table 4.3). One of the SNPs found on chromosome 19 was significant in the analysis for both years. The forty SNPs on chromosome 5 and the other SNP from chromosome 19 were significant in both pre-veraison and mature berry tartaric acid analyses.

More significant SNPs were found in the QTL analysis. For details on the location of these SNPs please see table 4.4. The same forty SNPs that were significant for tartaric acid in the GWAS on chromosome 5, the two from chromosome 7 and the one from chromosome 11, were also significant in the QTL analysis. There was a SNP that was significant on chromosome 19 in the GWAS for tartaric acid that was also significant, but it was mapped to chromosome 5. This SNP may have been incorrectly assigned to chromosome 19 when aligning the GBS data to find SNPs.

The set of forty significant SNPs from chromosome 5 all fall between approximately 400,000 bp and 3,000,000 bp on the chromosome. There are many genes in this region when compared to the NCBI genome annotations. A large number have unknown functions which raises the possibility of new uncharacterised genes. The genes in these region need to be investigated in more detail and a shortlist of candidates investigated in further detail.

The measurement of tartaric and malic acid in more grapevine populations may help to narrow down the QTL regions and confirm the results presented here. To do this, existing populations could be genotyped and their acid concentration measured. Alternately, new populations could be created, especially for the purpose, to find associations with acid

traits.

The process of GWAS and QTL could be simplified if the steps of the tartaric acid synthesis pathway could be measured separately. However, at this point in time, there are no methods to measure all of the intermediates and it is not known how long any intermediates exist in the grape berry. L-idonate is very rare, so creating good standards to measure this by UHPLC-MS/MS or HPLC is not feasible.

6.5 Tartaric acid concentration and L-IDH isoform expression in the tissues of the grapevine

While the concentration of tartaric acid has been studied extensively in berries, very little is known about the concentration in other tissues of the grapevine. One study reported the acid concentrations from several grapevine tissues (Kliewer 1966) and another documented the acid concentration in leaves from a variety of samples including *Vitis vinifera* and related species (Stafford 1959). In the work described in chapter 5, the tartaric and malic acid concentrations were measured in several different tissues of two microvine lines (Figure 5.1 and 5.2). Acid concentrations varied greatly across the vines and there were high concentrations of tartaric acid in leaves and stems as well as berries. The amount of each of the three L-IDH isoform transcripts was also measured in different tissues of the V10 microvine line using a pyrosequencing method (Table 5.1). The pyrosequencing results showed that L-IDH 3 and L-IDH 1 had different expression patterns in pre-veraison and mature berries. This pattern was also different from what was expected from previous expression studies where L-IDH 1 and 3 were measured together. It may be that L-IDH 1 and L-IDH 3 are regulated differently and a study of the up-stream regions of the two genes may shed some light on any differential regulation. It can not be

conclusively stated that L-IDH 3 is the original active L-IDH (reported by DeBolt *et al.* (2006)) as a mix up of these two isoforms may explain these results. When L-IDH 1 and L-IDH3 were compared to the L-IDH characterised by DeBolt et al. (2006) there was one amino acid difference between L-IDH 3 and the original L-IDH and two between L-IDH 1 and the original. Further sequencing of the L-IDH transcripts is needed to determine which isoform was originally characterised.

The L-IDH transcripts could be measured throughout berry development in more cultivars to determine if the differences in L-IDH 1 and L-IDH 3 are seen in wild type vines and across all cultivars. Unfortunately the activity of the L-IDH 1 protein could not be determined and it is still unknown if the transcript produces a protein that participates in tartaric acid synthesis. The expression pattern of L-IDH 1 matches the accumulation of tartaric acid in the berry much better than the expression pattern of L-IDH 3, so it is possible that it is active. It is also possible that there has been some confusion in naming L-IDH 3 as the active transcript. The sequence of L-IDH 3 has only one amino acid difference compared to the original characterised L-IDH where as L-IDH 1 had two amino acid's difference. Since the similarity between L-IDH 1 and L-IDH 3 is so high any sequencing errors may have had a huge impact and further sequencing is needed to confirm what was the originally characterised isoform as the original cloned cDNA from DeBolt *et al.* (2006) is still available.

Tartaric acid concentration was measured to determine if the concentration changes as the leaves aged (Figure 5.3). Leaf samples were taken from one growing shoot, starting with a sample of the first two newest leaves below the growing tip through to the oldest leaf sample. Leaf tartaric acid was mostly steady across development except for one vine

which had a sample with a very high concentration of tartaric acid. There may have been a mistake in the measurement or preparation of this single sample.

Tartaric acid concentration was measured in leaves from ten individuals from population 9 to compare to berry tartaric acid and examine any link between berry acid and leaf acid (Table 5.3). There was no correlation between leaf tartaric acid concentration and mature berry tartaric acid concentration (average of two years data) (Figure 5.4).

While tartaric acid is known to be made in the same pathway in leaves and berries it may be regulated differently in these tissues. The expression pattern of L-IDH 1 and L-IDH 3 was only measured in new leaf samples using pyrosequencing. Measuring of the level of L-IDH 1 and L-IDH 3 across leaf development might provide further information on any differential regulation occurring in leaves compared to berries.

The study of the relationship between leaf tartaric acid and mature berry tartaric acid was only preliminary and involved three separate samples from ten vines. This could be expanded to add extra samples which might provide some correlation between leaf and berry acid but from the data obtained here this looks unlikely. A method to predict berry tartaric acid concentration would be very useful, however it does not look like leaf tartaric acid provides that link.

It is believed that this is the first time that L-IDH 1 and L-IDH 3 transcripts have been measured separately. The two isoforms could not be amplified separately using qRT-PCR due to their similarity in sequence at the 3' end of the transcripts. L-IDH 1 and L-IDH 3 have been measured together in a variety of studies (Sweetman et al. 2012, Wen et al.

2014) and their expression pattern matched tartaric acid accumulation. In chapter 5, L-IDH 1 follows the pattern that was predicted in past studies, while L-IDH 3 follows a pattern not previously described.

The pyrosequencing method developed as part of this work can distinguish the expression L-IDH 1 and L-IDH 3. This will allow the further investigation of the expression of these two isoforms. The promoters of L-IDH 1 and L-IDH 3 have never been investigated. The next step in this research would be to link the differences in promoters of these two genes to differences in expression across different grapevine tissues and time points across berry development. This may help us to further understand the distinct roles of L-IDH 1 and L-IDH 3 in the grapevine.

6.6 Potassium, pH and tartaric acid concentration

Tartaric acid concentration in wines is influenced by pH and potassium concentration. In chapter 3 section 3.3, pH was measured in the four progeny populations at two time points over two years (Tables 3.8 to 3.11). pH varied more in mature berry samples than in pre-veraison samples. Potassium concentration was measured in the mature berry samples from population 10 in the 2013-2104 season (Figure 3.4).

More berries need to be measured to determine if potassium and pH are correlated to tartaric and malic acid concentrations and total acid concentrations. While no correlations were found here, it is possible that some correlation might be seen if more samples were included. This is particularly true for potassium which was only measured in one population (under 100 samples).

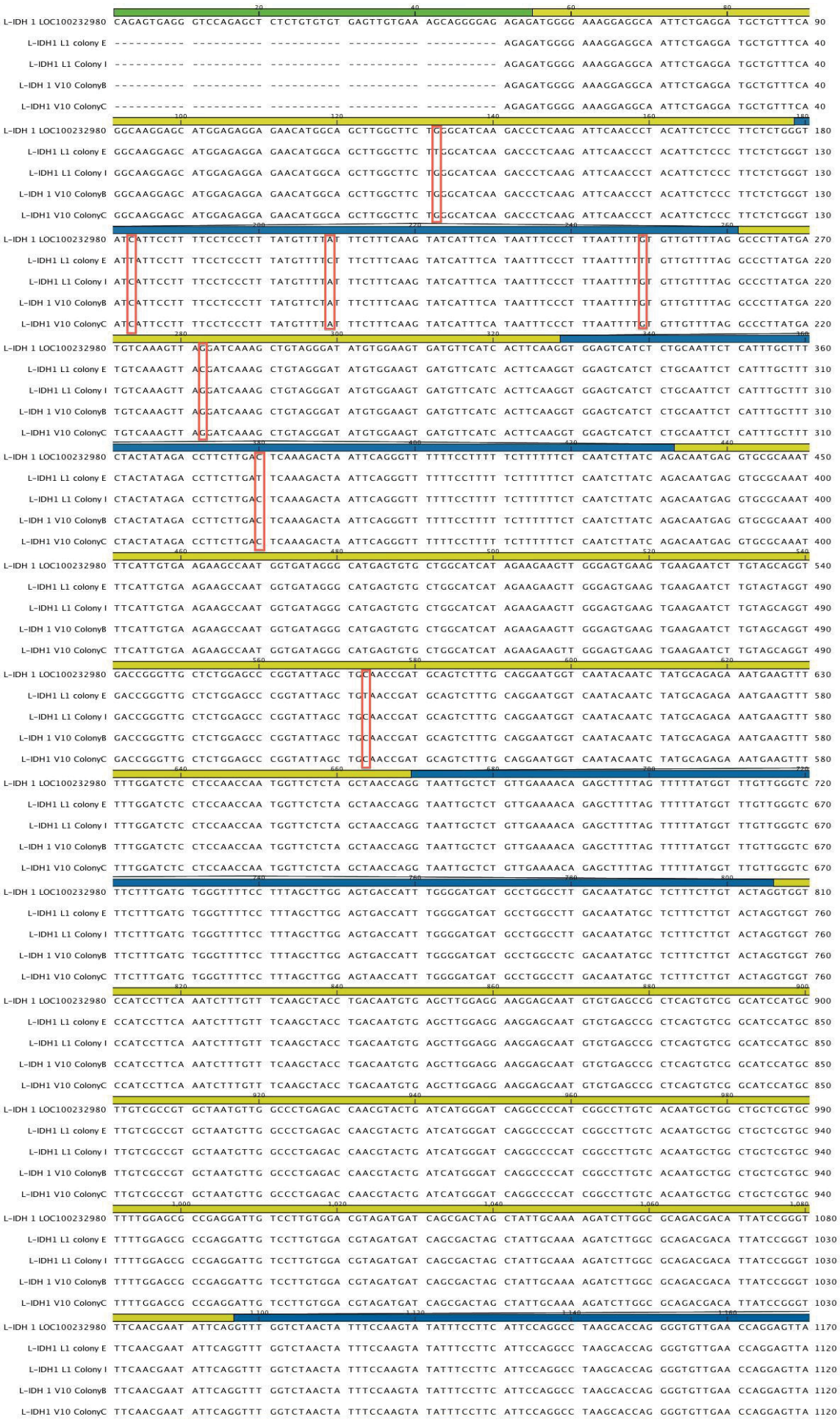
Potassium and tartaric acid in the juice of crushed grapes can form potassium bitartrate which is insoluble (Mpelasoka et al. 2003). This lowers the acid concentration and increases the pH in the wine which can lead to microbial spoilage. Tartaric acid is added to decrease the pH of wines during fermentation. If new grape cultivars with higher concentrations of tartaric acid in grape berries are to be developed it would also be advantageous to make sure that the new cultivars do not also have high concentrations of potassium which would negate the effects of increasing the tartaric acid concentration.

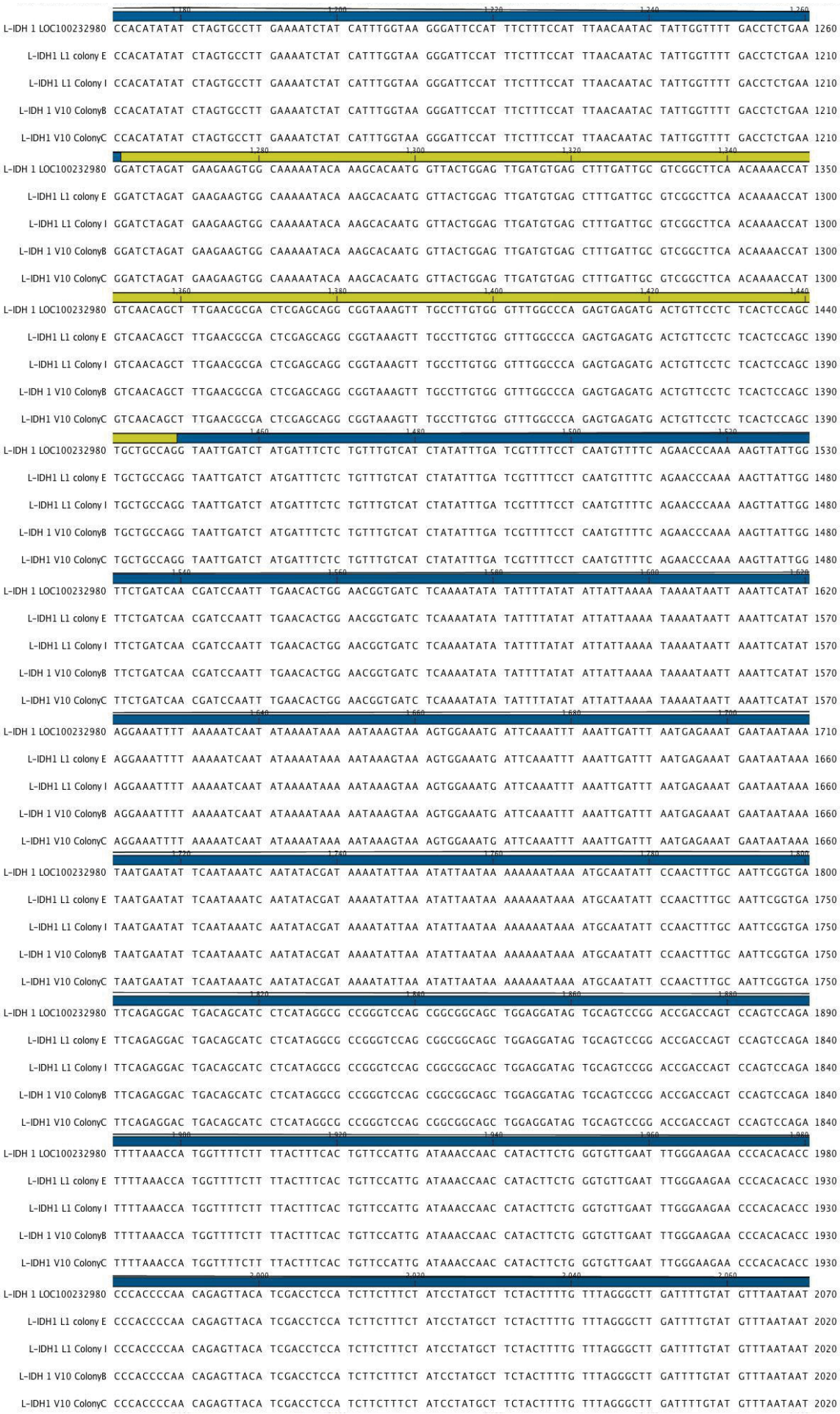
6.7 Summary

Tartaric acid is added in Australian wineries to decrease the pH during fermentation which decreases oxidation and microbial spoilage. Adding tartaric acid is a large cost for winemakers. Due to global warming the pH of grape berries is expected to increase due to the increased malic acid breakdown that occurs when temperatures are high. If acidity decreases, more acid would need to be added in the winery to achieve the desired pH which will increase costs to winemakers.

Breeding new grapevine cultivars with higher tartaric acid concentrations could help to negate some of these costs. However, very little is known about tartaric acid synthesis in grapevines. The research presented here provides more insights into the three isoforms of L-IDH; their involvement in tartaric acid synthesis; and expression patterns across tissues of the grapevine. The discovery of chromosome regions involved in acid metabolism through GWAS and QTL analysis could lead to the discovery of new genes involved in acid metabolism in the grapevine and markers for grapevine breeding.

Appendices


L-IDH 1 LOC100232980 CAGAGTGAGG GTCCAGAGCT CTCTGTGTGT GAGTTGTGAA AGCAGGGGAG AGAGATGGGG AAAGGAGGCA ATTCTGAGGA TGCTGTTTCA 90
L-IDH1 L1 colony E ---------- ---------- ---------- ---------- ---------- AGAGATGGGG AAAGGAGGCA ATTCTGAGGA TGCTGTTTCA 40
L-IDH1 L1 Colony I ---------- ---------- ---------- ---------- ---------- AGAGATGGGG AAAGGAGGCA ATTCTGAGGA TGCTGTTTCA 40
L-IDH 1 V10 ColonyB ---------- ---------- ---------- ---------- ---------- AGAGATGGGG AAAGGAGGCA ATTCTGAGGA TGCTGTTTCA 40
L-IDH1 V10 ColonyC ---------- ---------- ---------- ---------- ---------- AGAGATGGGG AAAGGAGGCA ATTCTGAGGA TGCTGTTTCA 40
L-IDH 1 LOC100232980 GGCAAGGAGC ATGGAGAGGA GAACATGGCA GCTTGGCTTC TGGGCATCAA GACCCTCAAG ATTCAACCCT ACATTCTCCC TTCTCTGGGT 180
L-IDH1 L1 colony E GGCAAGGAGC ATGGAGAGGA GAACATGGCA GCTTGGCTTC TGGGCATCAA GACCCTCAAG ATTCAACCCT ACATTCTCCC TTCTCTGGGT 130
L-IDH1 L1 Colony I GGCAAGGAGC ATGGAGAGGA GAACATGGCA GCTTGGCTTC TGGGCATCAA GACCCTCAAG ATTCAACCCT ACATTCTCCC TTCTCTGGGT 130
L-IDH 1 V10 ColonyB GGCAAGGAGC ATGGAGAGGA GAACATGGCA GCTTGGCTTC TGGGCATCAA GACCCTCAAG ATTCAACCCT ACATTCTCCC TTCTCTGGGT 130
L-IDH1 V10 ColonyC GGCAAGGAGC ATGGAGAGGA GAACATGGCA GCTTGGCTTC TGGGCATCAA GACCCTCAAG ATTCAACCCT ACATTCTCCC TTCTCTGGGT 130
L-IDH 1 LOC100232980 ATCATTCCTT TTCCTCCCTT TATGTTTTAT TTCTTTCAAG TATCATTTCA TAATTTCCCT TTAATTTTGT GTTGTTTTAG GCCCTTATGA 270
L-IDH1 L1 colony E ATTATTCCTT TTCCTCCCTT TATGTTTTCT TTCTTTCAAG TATCATTTCA TAATTTCCCT TTAATTTTTT GTTGTTTTAG GCCCTTATGA 220
L-IDH1 L1 Colony I ATCATTCCTT TTCCTCCCTT TATGTTTTAT TTCTTTCAAG TATCATTTCA TAATTTCCCT TTAATTTTGT GTTGTTTTAG GCCCTTATGA 220
L-IDH 1 V10 ColonyB ATCATTCCTT TTCCTCCCTT TATGTTCTAT TTCTTTCAAG TATCATTTCA TAATTTCCCT TTAATTTTGT GTTGTTTTAG GCCCTTATGA 220
L-IDH1 V10 ColonyC ATCATTCCTT TTCCTCCCTT TATGTTTTAT TTCTTTCAAG TATCATTTCA TAATTTCCCT TTAATTTTGT GTTGTTTTAG GCCCTTATGA 220
L-IDH 1 LOC100232980 TGTCAAAGTT AGGATCAAAG CTGTAGGGAT ATGTGGAAGT GATGTTCATC ACTTCAAGGT GGAGTCATCT CTGCAATTCT CATTTGCTTT 360
L-IDH1 L1 colony E TGTCAAAGTT ACGATCAAAG CTGTAGGGAT ATGTGGAAGT GATGTTCATC ACTTCAAGGT GGAGTCATCT CTGCAATTCT CATTTGCTTT 310
L-IDH1 L1 Colony I TGTCAAAGTT AGGATCAAAG CTGTAGGGAT ATGTGGAAGT GATGTTCATC ACTTCAAGGT GGAGTCATCT CTGCAATTCT CATTTGCTTT 310
L-IDH 1 V10 ColonyB TGTCAAAGTT AGGATCAAAG CTGTAGGGAT ATGTGGAAGT GATGTTCATC ACTTCAAGGT GGAGTCATCT CTGCAATTCT CATTTGCTTT 310
L-IDH1 V10 ColonyC TGTCAAAGTT AGGATCAAAG CTGTAGGGAT ATGTGGAAGT GATGTTCATC ACTTCAAGGT GGAGTCATCT CTGCAATTCT CATTTGCTTT 310
L-IDH 1 LOC100232980 CTACTATAGA CCTTCTTGAC TCAAAGACTA ATTCAGGGTT TTTTCCTTTT TCTTTTTTCT CAATCTTATC AGACAATGAG GTGCGCAAAT 450
L-IDH1 L1 colony E CTACTATAGA CCTTCTTGAT TCAAAGACTA ATTCAGGGTT TTTTCCTTTT TCTTTTTTCT CAATCTTATC AGACAATGAG GTGCGCAAAT 400
L-IDH1 L1 Colony I CTACTATAGA CCTTCTTGAC TCAAAGACTA ATTCAGGGTT TTTTCCTTTT TCTTTTTTCT CAATCTTATC AGACAATGAG GTGCGCAAAT 400
L-IDH 1 V10 ColonyB CTACTATAGA CCTTCTTGAC TCAAAGACTA ATTCAGGGTT TTTTCCTTTT TCTTTTTTCT CAATCTTATC AGACAATGAG GTGCGCAAAT 400
L-IDH1 V10 ColonyC CTACTATAGA CCTTCTTGAC TCAAAGACTA ATTCAGGGTT TTTTCCTTTT TCTTTTTTCT CAATCTTATC AGACAATGAG GTGCGCAAAT 400
L-IDH 1 LOC100232980 TTCATTGTGA AGAAGCCAAT GGTGATAGGG CATGAGTGTG CTGGCATCAT AGAAGAAGTT GGGAGTGAAG TGAAGAATCT TGTAGCAGGT 540
L-IDH1 L1 colony E TTCATTGTGA AGAAGCCAAT GGTGATAGGG CATGAGTGTG CTGGCATCAT AGAAGAAGTT GGGAGTGAAG TGAAGAATCT TGTAGTAGGT 490
L-IDH1 L1 Colony I TTCATTGTGA AGAAGCCAAT GGTGATAGGG CATGAGTGTG CTGGCATCAT AGAAGAAGTT GGGAGTGAAG TGAAGAATCT TGTAGCAGGT 490
L-IDH 1 V10 ColonyB TTCATTGTGA AGAAGCCAAT GGTGATAGGG CATGAGTGTG CTGGCATCAT AGAAGAAGTT GGGAGTGAAG TGAAGAATCT TGTAGCAGGT 490
L-IDH1 V10 ColonyC TTCATTGTGA AGAAGCCAAT GGTGATAGGG CATGAGTGTG CTGGCATCAT AGAAGAAGTT GGGAGTGAAG TGAAGAATCT TGTAGCAGGT 490
L-IDH 1 LOC100232980 GACCGGGTTG CTCTGGAGCC CGGTATTAGC TGCAACCGAT GCAGTCTTTG CAGGAATGGT CAATACAATC TATGCAGAGA AATGAAGTTT 630
L-IDH1 L1 colony E GACCGGGTTG CTCTGGAGCC CGGTATTAGC TGTAACCGAT GCAGTCTTTG CAGGAATGGT CAATACAATC TATGCAGAGA AATGAAGTTT 580
L-IDH1 L1 Colony I GACCGGGTTG CTCTGGAGCC CGGTATTAGC TGCAACCGAT GCAGTCTTTG CAGGAATGGT CAATACAATC TATGCAGAGA AATGAAGTTT 580
L-IDH 1 V10 ColonyB GACCGGGTTG CTCTGGAGCC CGGTATTAGC TGCAACCGAT GCAGTCTTTG CAGGAATGGT CAATACAATC TATGCAGAGA AATGAAGTTT 580
L-IDH1 V10 ColonyC GACCGGGTTG CTCTGGAGCC CGGTATTAGC TGCAACCGAT GCAGTCTTTG CAGGAATGGT CAATACAATC TATGCAGAGA AATGAAGTTT 580
L-IDH 1 LOC100232980 TTTGGATCTC CTCCAACCAA TGGTTCTCTA GCTAACCAGG TAATTGCTCT GTTGAAAACA GAGCTTTTAG TTTTTATGGT TTGTTGGGTC 720
L-IDH1 L1 colony E TTTGGATCTC CTCCAACCAA TGGTTCTCTA GCTAACCAGG TAATTGCTCT GTTGAAAACA GAGCTTTTAG TTTTTATGGT TTGTTGGGTC 670
L-IDH1 L1 Colony I TTTGGATCTC CTCCAACCAA TGGTTCTCTA GCTAACCAGG TAATTGCTCT GTTGAAAACA GAGCTTTTAG TTTTTATGGT TTGTTGGGTC 670
L-IDH 1 V10 ColonyB TTTGGATCTC CTCCAACCAA TGGTTCTCTA GCTAACCAGG TAATTGCTCT GTTGAAAACA GAGCTTTTAG TTTTTATGGT TTGTTGGGTC 670
L-IDH1 V10 ColonyC TTTGGATCTC CTCCAACCAA TGGTTCTCTA GCTAACCAGG TAATTGCTCT GTTGAAAACA GAGCTTTTAG TTTTTATGGT TTGTTGGGTC 670
L-IDH 1 LOC100232980 TTCTTTGATG TGGGTTTTCC TTTAGCTTGG AGTGACCATT TGGGGATGAT GCCTGGCCTT GACAATATGC TCTTTCTTGT ACTAGGTGGT 810
L-IDH1 L1 colony E TTCTTTGATG TGGGTTTTCC TTTAGCTTGG AGTGACCATT TGGGGATGAT GCCTGGCCTT GACAATATGC TCTTTCTTGT ACTAGGTGGT 760
L-IDH1 L1 Colony I TTCTTTGATG TGGGTTTTCC TTTAGCTTGG AGTGACCATT TGGGGATGAT GCCTGGCCTT GACAATATGC TCTTTCTTGT ACTAGGTGGT 760
L-IDH 1 V10 ColonyB TTCTTTGATG TGGGTTTTCC TTTAGCTTGG AGTGACCATT TGGGGATGAT GCCTGGCCTC GACAATATGC TCTTTCTTGT ACTAGGTGGT 760
L-IDH1 V10 ColonyC TTCTTTGATG TGGGTTTTCC TTTAGCTTGG AGTAACCATT TGGGGATGAT GCCTGGCCTT GACAATATGC TCTTTCTTGT ACTAGGTGGT 760
L-IDH 1 LOC100232980 CCATCCTTCA AATCTTTGTT TCAAGCTACC TGACAATGTG AGCTTGGAGG AAGGAGCAAT GTGTGAGCCG CTCAGTGTCG GCATCCATGC 900
L-IDH1 L1 colony E CCATCCTTCA AATCTTTGTT TCAAGCTACC TGACAATGTG AGCTTGGAGG AAGGAGCAAT GTGTGAGCCG CTCAGTGTCG GCATCCATGC 850
L-IDH1 L1 Colony I CCATCCTTCA AATCTTTGTT TCAAGCTACC TGACAATGTG AGCTTGGAGG AAGGAGCAAT GTGTGAGCCG CTCAGTGTCG GCATCCATGC 850
L-IDH 1 V10 ColonyB CCATCCTTCA AATCTTTGTT TCAAGCTACC TGACAATGTG AGCTTGGAGG AAGGAGCAAT GTGTGAGCCG CTCAGTGTCG GCATCCATGC 850
L-IDH1 V10 ColonyC CCATCCTTCA AATCTTTGTT TCAAGCTACC TGACAATGTG AGCTTGGAGG AAGGAGCAAT GTGTGAGCCG CTCAGTGTCG GCATCCATGC 850
L-IDH 1 LOC100232980 TTGTCGCCGT GCTAATGTTG GCCCTGAGAC CAACGTACTG ATCATGGGAT CAGGCCCCAT CGGCCTTGTC ACAATGCTGG CTGCTCGTGC 990
L-IDH1 L1 colony E TTGTCGCCGT GCTAATGTTG GCCCTGAGAC CAACGTACTG ATCATGGGAT CAGGCCCCAT CGGCCTTGTC ACAATGCTGG CTGCTCGTGC 940
L-IDH1 L1 Colony I TTGTCGCCGT GCTAATGTTG GCCCTGAGAC CAACGTACTG ATCATGGGAT CAGGCCCCAT CGGCCTTGTC ACAATGCTGG CTGCTCGTGC 940
L-IDH 1 V10 ColonyB TTGTCGCCGT GCTAATGTTG GCCCTGAGAC CAACGTACTG ATCATGGGAT CAGGCCCCAT CGGCCTTGTC ACAATGCTGG CTGCTCGTGC 940
L-IDH1 V10 ColonyC TTGTCGCCGT GCTAATGTTG GCCCTGAGAC CAACGTACTG ATCATGGGAT CAGGCCCCAT CGGCCTTGTC ACAATGCTGG CTGCTCGTGC 940
L-IDH 1 LOC100232980 TTTTGGAGCG CCGAGGATTG TCCTTGTGGA CGTAGATGAT CAGCGACTAG CTATTGCAAA AGATCTTGGC GCAGACGACA TTATCCGGGT 1080
L-IDH1 L1 colony E TTTTGGAGCG CCGAGGATTG TCCTTGTGGA CGTAGATGAT CAGCGACTAG CTATTGCAAA AGATCTTGGC GCAGACGACA TTATCCGGGT 1030
L-IDH1 L1 Colony I TTTTGGAGCG CCGAGGATTG TCCTTGTGGA CGTAGATGAT CAGCGACTAG CTATTGCAAA AGATCTTGGC GCAGACGACA TTATCCGGGT 1030
L-IDH 1 V10 ColonyB TTTTGGAGCG CCGAGGATTG TCCTTGTGGA CGTAGATGAT CAGCGACTAG CTATTGCAAA AGATCTTGGC GCAGACGACA TTATCCGGGT 1030
L-IDH1 V10 ColonyC TTTTGGAGCG CCGAGGATTG TCCTTGTGGA CGTAGATGAT CAGCGACTAG CTATTGCAAA AGATCTTGGC GCAGACGACA TTATCCGGGT 1030
L-IDH 1 LOC100232980 TTCAACGAAT ATTCAGGTTT GGTCTAACTA TTTCCAAGTA TATTTCCTTC ATTCCAGGCC TAAGCACCAG GGGTGTTGAA CCAGGAGTTA 1170
L-IDH1 L1 colony E TTCAACGAAT ATTCAGGTTT GGTCTAACTA TTTCCAAGTA TATTTCCTTC ATTCCAGGCC TAAGCACCAG GGGTGTTGAA CCAGGAGTTA 1120
L-IDH1 L1 Colony I TTCAACGAAT ATTCAGGTTT GGTCTAACTA TTTCCAAGTA TATTTCCTTC ATTCCAGGCC TAAGCACCAG GGGTGTTGAA CCAGGAGTTA 1120
L-IDH 1 V10 ColonyB TTCAACGAAT ATTCAGGTTT GGTCTAACTA TTTCCAAGTA TATTTCCTTC ATTCCAGGCC TAAGCACCAG GGGTGTTGAA CCAGGAGTTA 1120
L-IDH1 V10 ColonyC TTCAACGAAT ATTCAGGTTT GGTCTAACTA TTTCCAAGTA TATTTCCTTC ATTCCAGGCC TAAGCACCAG GGGTGTTGAA CCAGGAGTTA 1120


L-IDH 1 LOC100232980 CCACATATAT CTAGTGCCTT GAAAATCTAT CATTTGGTAA GGGATTCCAT TTCTTTCCAT TTAACAATAC TATTGGTTTT GACCTCTGAA 1260
L-IDH1 L1 colony E CCACATATAT CTAGTGCCTT GAAAATCTAT CATTTGGTAA GGGATTCCAT TTCTTTCCAT TTAACAATAC TATTGGTTTT GACCTCTGAA 1210
L-IDH1 L1 Colony I CCACATATAT CTAGTGCCTT GAAAATCTAT CATTTGGTAA GGGATTCCAT TTCTTTCCAT TTAACAATAC TATTGGTTTT GACCTCTGAA 1210
L-IDH 1 V10 ColonyB CCACATATAT CTAGTGCCTT GAAAATCTAT CATTTGGTAA GGGATTCCAT TTCTTTCCAT TTAACAATAC TATTGGTTTT GACCTCTGAA 1210
L-IDH1 V10 ColonyC CCACATATAT CTAGTGCCTT GAAAATCTAT CATTTGGTAA GGGATTCCAT TTCTTTCCAT TTAACAATAC TATTGGTTTT GACCTCTGAA 1210
L-IDH 1 LOC100232980 GGATCTAGAT GAAGAAGTGG CAAAAATACA AAGCACAATG GTTACTGGAG TTGATGTGAG CTTTGATTGC GTCGGCTTCA ACAAAACCAT 1350
L-IDH1 L1 colony E GGATCTAGAT GAAGAAGTGG CAAAAATACA AAGCACAATG GTTACTGGAG TTGATGTGAG CTTTGATTGC GTCGGCTTCA ACAAAACCAT 1300
L-IDH1 L1 Colony I GGATCTAGAT GAAGAAGTGG CAAAAATACA AAGCACAATG GTTACTGGAG TTGATGTGAG CTTTGATTGC GTCGGCTTCA ACAAAACCAT 1300
L-IDH 1 V10 ColonyB GGATCTAGAT GAAGAAGTGG CAAAAATACA AAGCACAATG GTTACTGGAG TTGATGTGAG CTTTGATTGC GTCGGCTTCA ACAAAACCAT 1300
L-IDH1 V10 ColonyC GGATCTAGAT GAAGAAGTGG CAAAAATACA AAGCACAATG GTTACTGGAG TTGATGTGAG CTTTGATTGC GTCGGCTTCA ACAAAACCAT 1300
L-IDH 1 LOC100232980 GTCAACAGCT TTGAACGCGA CTCGAGCAGG CGGTAAAGTT TGCCTTGTGG GTTTGGCCCA GAGTGAGATG ACTGTTCCTC TCACTCCAGC 1440
L-IDH1 L1 colony E GTCAACAGCT TTGAACGCGA CTCGAGCAGG CGGTAAAGTT TGCCTTGTGG GTTTGGCCCA GAGTGAGATG ACTGTTCCTC TCACTCCAGC 1390
L-IDH1 L1 Colony I GTCAACAGCT TTGAACGCGA CTCGAGCAGG CGGTAAAGTT TGCCTTGTGG GTTTGGCCCA GAGTGAGATG ACTGTTCCTC TCACTCCAGC 1390
L-IDH 1 V10 ColonyB GTCAACAGCT TTGAACGCGA CTCGAGCAGG CGGTAAAGTT TGCCTTGTGG GTTTGGCCCA GAGTGAGATG ACTGTTCCTC TCACTCCAGC 1390
L-IDH1 V10 ColonyC GTCAACAGCT TTGAACGCGA CTCGAGCAGG CGGTAAAGTT TGCCTTGTGG GTTTGGCCCA GAGTGAGATG ACTGTTCCTC TCACTCCAGC 1390
L-IDH 1 LOC100232980 TGCTGCCAGG TAATTGATCT ATGATTTCTC TGTTTGTCAT CTATATTTGA TCGTTTTCCT CAATGTTTTC AGAACCCAAA AAGTTATTGG 1530
L-IDH1 L1 colony E TGCTGCCAGG TAATTGATCT ATGATTTCTC TGTTTGTCAT CTATATTTGA TCGTTTTCCT CAATGTTTTC AGAACCCAAA AAGTTATTGG 1480
L-IDH1 L1 Colony I TGCTGCCAGG TAATTGATCT ATGATTTCTC TGTTTGTCAT CTATATTTGA TCGTTTTCCT CAATGTTTTC AGAACCCAAA AAGTTATTGG 1480
L-IDH 1 V10 ColonyB TGCTGCCAGG TAATTGATCT ATGATTTCTC TGTTTGTCAT CTATATTTGA TCGTTTTCCT CAATGTTTTC AGAACCCAAA AAGTTATTGG 1480
L-IDH1 V10 ColonyC TGCTGCCAGG TAATTGATCT ATGATTTCTC TGTTTGTCAT CTATATTTGA TCGTTTTCCT CAATGTTTTC AGAACCCAAA AAGTTATTGG 1480
L-IDH 1 LOC100232980 TTCTGATCAA CGATCCAATT TGAACACTGG AACGGTGATC TCAAAATATA TATTTTATAT ATTATTAAAA TAAAATAATT AAATTCATAT 1620
L-IDH1 L1 colony E TTCTGATCAA CGATCCAATT TGAACACTGG AACGGTGATC TCAAAATATA TATTTTATAT ATTATTAAAA TAAAATAATT AAATTCATAT 1570
L-IDH1 L1 Colony I TTCTGATCAA CGATCCAATT TGAACACTGG AACGGTGATC TCAAAATATA TATTTTATAT ATTATTAAAA TAAAATAATT AAATTCATAT 1570
L-IDH 1 V10 ColonyB TTCTGATCAA CGATCCAATT TGAACACTGG AACGGTGATC TCAAAATATA TATTTTATAT ATTATTAAAA TAAAATAATT AAATTCATAT 1570
L-IDH1 V10 ColonyC TTCTGATCAA CGATCCAATT TGAACACTGG AACGGTGATC TCAAAATATA TATTTTATAT ATTATTAAAA TAAAATAATT AAATTCATAT 1570
L-IDH 1 LOC100232980 AGGAAATTTT AAAAATCAAT ATAAAATAAA AATAAAGTAA AGTGGAAATG ATTCAAATTT AAATTGATTT AATGAGAAAT GAATAATAAA 1710
L-IDH1 L1 colony E AGGAAATTTT AAAAATCAAT ATAAAATAAA AATAAAGTAA AGTGGAAATG ATTCAAATTT AAATTGATTT AATGAGAAAT GAATAATAAA 1660
L-IDH1 L1 Colony I AGGAAATTTT AAAAATCAAT ATAAAATAAA AATAAAGTAA AGTGGAAATG ATTCAAATTT AAATTGATTT AATGAGAAAT GAATAATAAA 1660
L-IDH 1 V10 ColonyB AGGAAATTTT AAAAATCAAT ATAAAATAAA AATAAAGTAA AGTGGAAATG ATTCAAATTT AAATTGATTT AATGAGAAAT GAATAATAAA 1660
L-IDH1 V10 ColonyC AGGAAATTTT AAAAATCAAT ATAAAATAAA AATAAAGTAA AGTGGAAATG ATTCAAATTT AAATTGATTT AATGAGAAAT GAATAATAAA 1660
L-IDH 1 LOC100232980 TAATGAATAT TCAATAAATC AATATACGAT AAAATATTAA ATATTAATAA AAAAAATAAA ATGCAATATT CCAACTTTGC AATTCGGTGA 1800
L-IDH1 L1 colony E TAATGAATAT TCAATAAATC AATATACGAT AAAATATTAA ATATTAATAA AAAAAATAAA ATGCAATATT CCAACTTTGC AATTCGGTGA 1750
L-IDH1 L1 Colony I TAATGAATAT TCAATAAATC AATATACGAT AAAATATTAA ATATTAATAA AAAAAATAAA ATGCAATATT CCAACTTTGC AATTCGGTGA 1750
L-IDH 1 V10 ColonyB TAATGAATAT TCAATAAATC AATATACGAT AAAATATTAA ATATTAATAA AAAAAATAAA ATGCAATATT CCAACTTTGC AATTCGGTGA 1750
L-IDH1 V10 ColonyC TAATGAATAT TCAATAAATC AATATACGAT AAAATATTAA ATATTAATAA AAAAAATAAA ATGCAATATT CCAACTTTGC AATTCGGTGA 1750
L-IDH 1 LOC100232980 TTCAGAGGAC TGACAGCATC CTCATAGGCG CCGGGTCCAG CGGCGGCAGC TGGAGGATAG TGCAGTCCGG ACCGACCAGT CCAGTCCAGA 1890
L-IDH1 L1 colony E TTCAGAGGAC TGACAGCATC CTCATAGGCG CCGGGTCCAG CGGCGGCAGC TGGAGGATAG TGCAGTCCGG ACCGACCAGT CCAGTCCAGA 1840
L-IDH1 L1 Colony I TTCAGAGGAC TGACAGCATC CTCATAGGCG CCGGGTCCAG CGGCGGCAGC TGGAGGATAG TGCAGTCCGG ACCGACCAGT CCAGTCCAGA 1840
L-IDH 1 V10 ColonyB TTCAGAGGAC TGACAGCATC CTCATAGGCG CCGGGTCCAG CGGCGGCAGC TGGAGGATAG TGCAGTCCGG ACCGACCAGT CCAGTCCAGA 1840
L-IDH1 V10 ColonyC TTCAGAGGAC TGACAGCATC CTCATAGGCG CCGGGTCCAG CGGCGGCAGC TGGAGGATAG TGCAGTCCGG ACCGACCAGT CCAGTCCAGA 1840
L-IDH 1 LOC100232980 TTTTAAACCA TGGTTTTCTT TTACTTTCAC TGTTCCATTG ATAAACCAAC CATACTTCTG GGTGTTGAAT TTGGGAAGAA CCCACACACC 1980
L-IDH1 L1 colony E TTTTAAACCA TGGTTTTCTT TTACTTTCAC TGTTCCATTG ATAAACCAAC CATACTTCTG GGTGTTGAAT TTGGGAAGAA CCCACACACC 1930
L-IDH1 L1 Colony I TTTTAAACCA TGGTTTTCTT TTACTTTCAC TGTTCCATTG ATAAACCAAC CATACTTCTG GGTGTTGAAT TTGGGAAGAA CCCACACACC 1930
L-IDH 1 V10 ColonyB TTTTAAACCA TGGTTTTCTT TTACTTTCAC TGTTCCATTG ATAAACCAAC CATACTTCTG GGTGTTGAAT TTGGGAAGAA CCCACACACC 1930
L-IDH1 V10 ColonyC TTTTAAACCA TGGTTTTCTT TTACTTTCAC TGTTCCATTG ATAAACCAAC CATACTTCTG GGTGTTGAAT TTGGGAAGAA CCCACACACC 1930
L-IDH 1 LOC100232980 CCCACCCCAA CAGAGTTACA TCGACCTCCA TCTTCTTTCT ATCCTATGCT TCTACTTTTG TTTAGGGCTT GATTTTGTAT GTTTAATAAT 2070
L-IDH1 L1 colony E CCCACCCCAA CAGAGTTACA TCGACCTCCA TCTTCTTTCT ATCCTATGCT TCTACTTTTG TTTAGGGCTT GATTTTGTAT GTTTAATAAT 2020
L-IDH1 L1 Colony I CCCACCCCAA CAGAGTTACA TCGACCTCCA TCTTCTTTCT ATCCTATGCT TCTACTTTTG TTTAGGGCTT GATTTTGTAT GTTTAATAAT 2020
L-IDH 1 V10 ColonyB CCCACCCCAA CAGAGTTACA TCGACCTCCA TCTTCTTTCT ATCCTATGCT TCTACTTTTG TTTAGGGCTT GATTTTGTAT GTTTAATAAT 2020
L-IDH1 V10 ColonyC CCCACCCCAA CAGAGTTACA TCGACCTCCA TCTTCTTTCT ATCCTATGCT TCTACTTTTG TTTAGGGCTT GATTTTGTAT GTTTAATAAT 2020

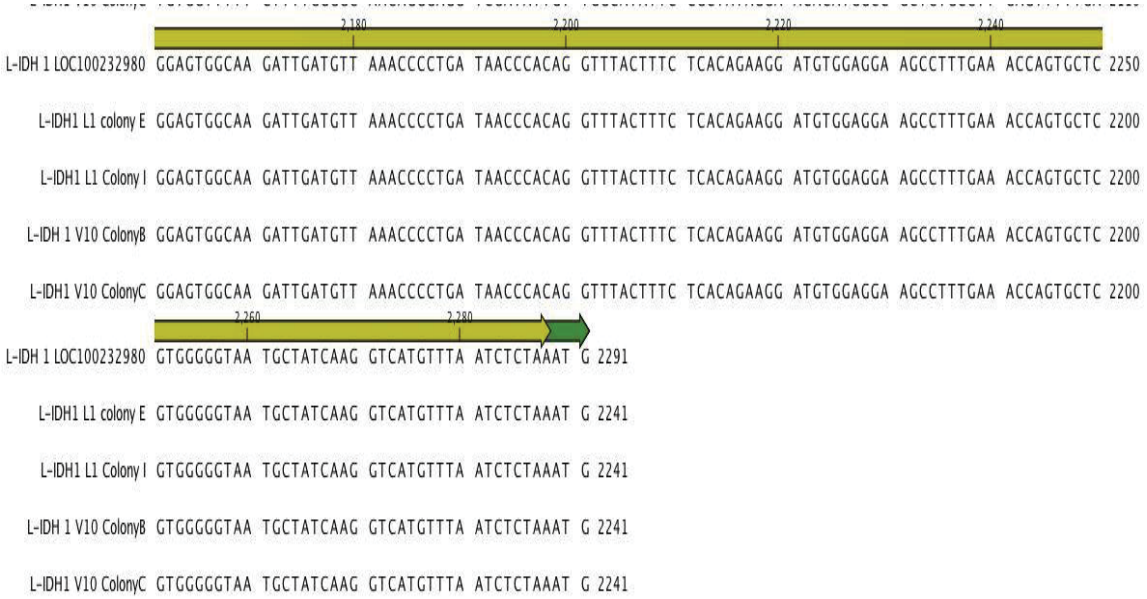
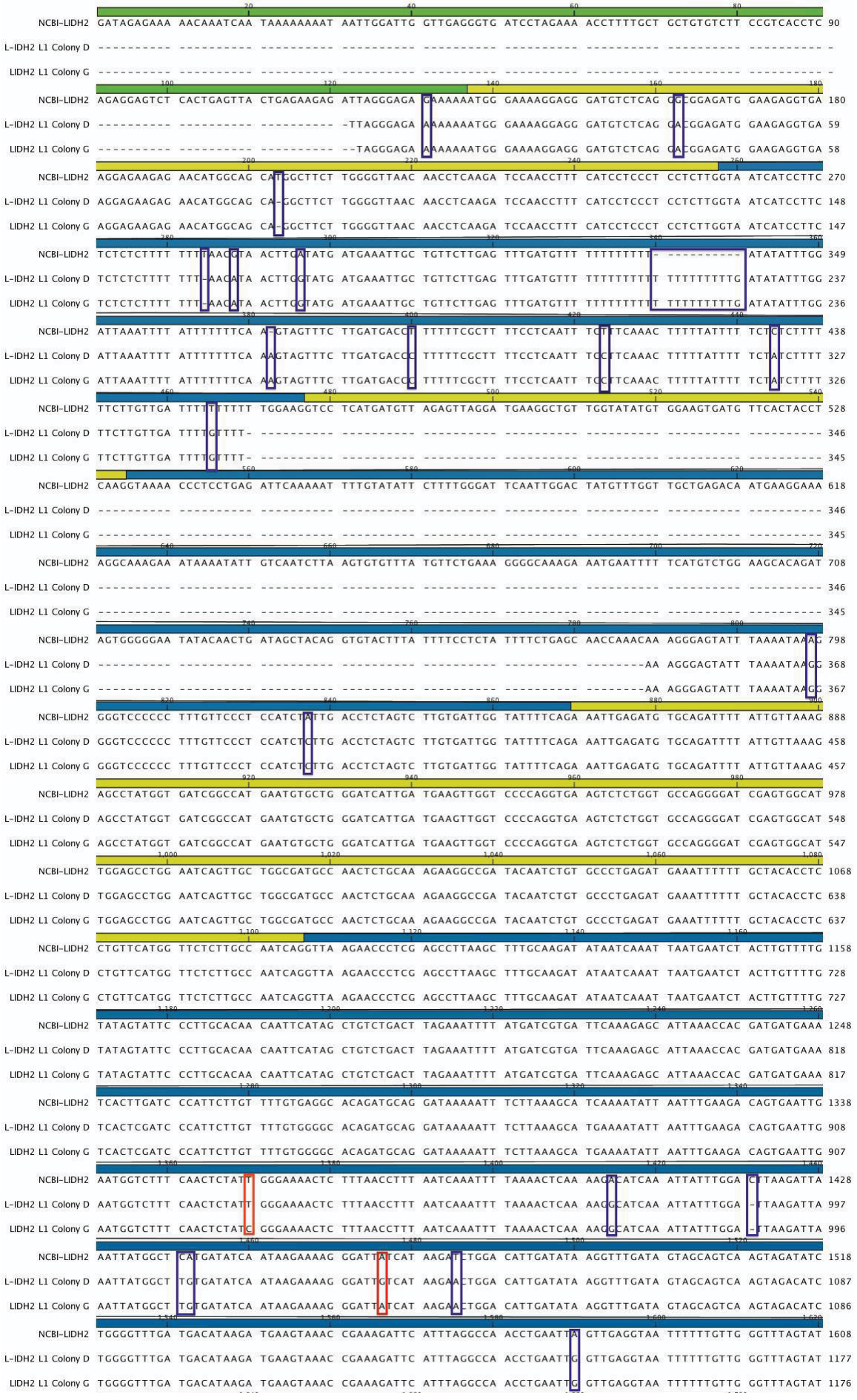


Figure A.1: L-idonate dehydrogenase 1 sequence from the V10 homozygous microvine line and the heterozygous L1 parent.


NCBI-LIDH2
L-IDH2 L1 Colony D
LIDH2 L1 Colony G

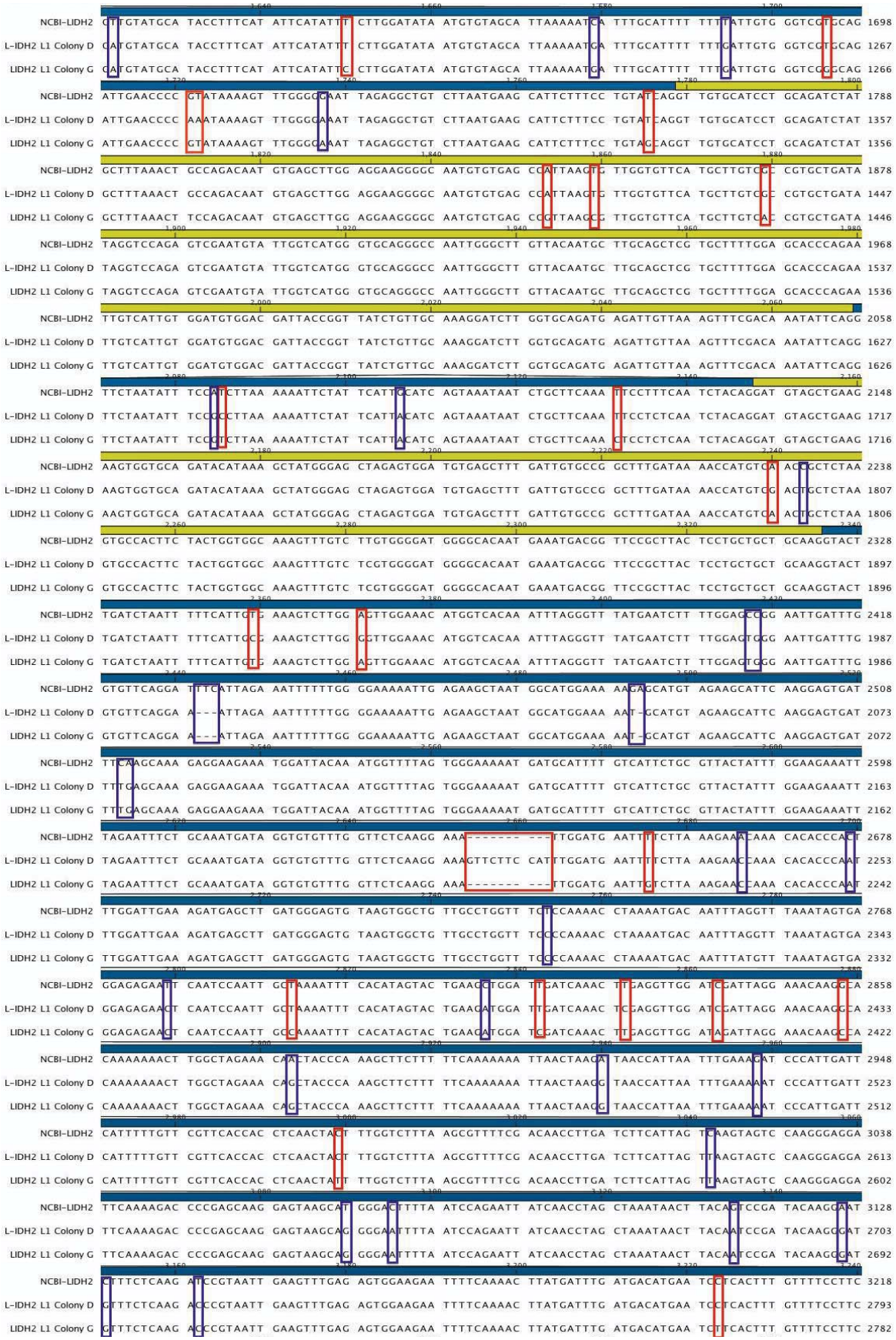
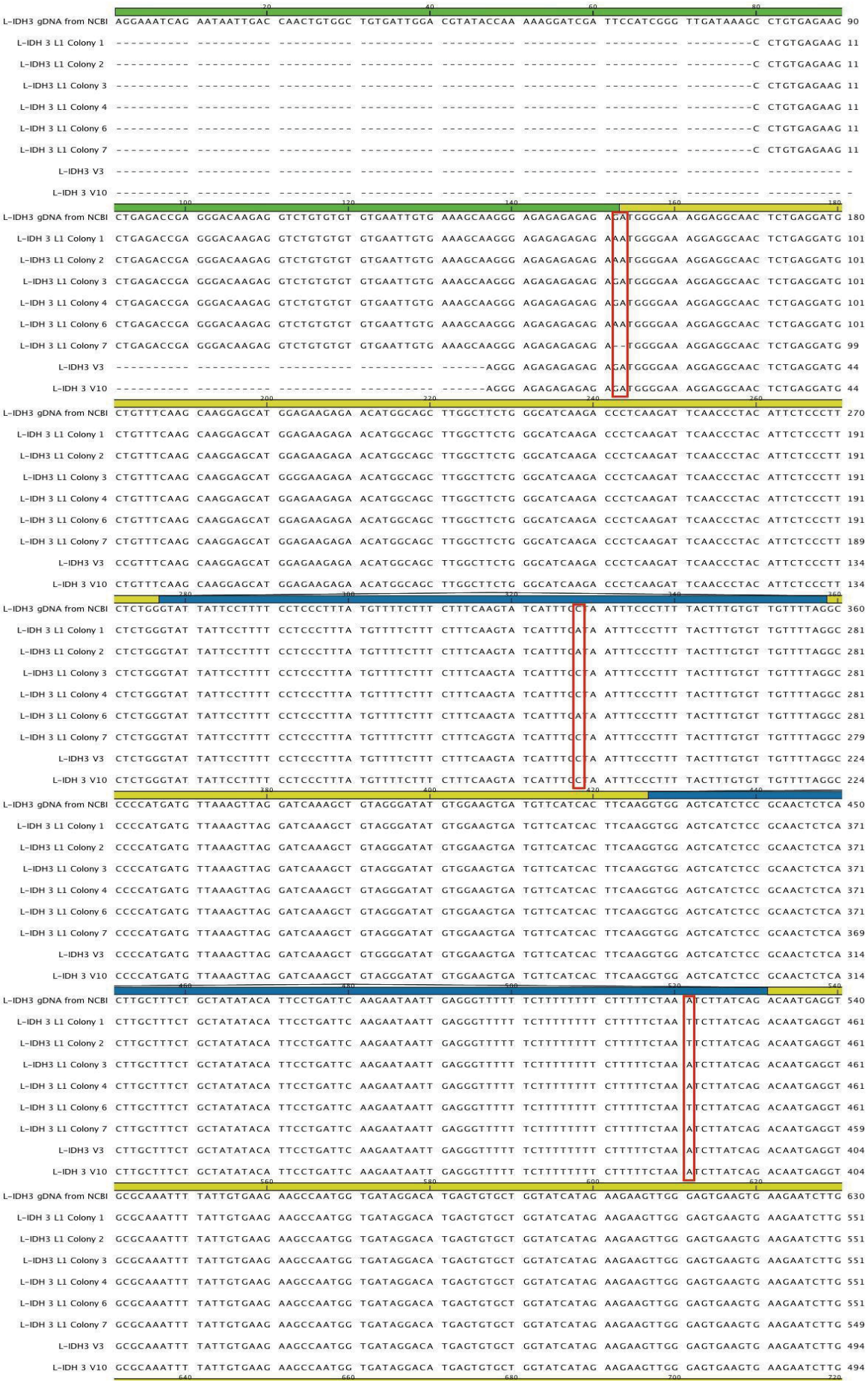
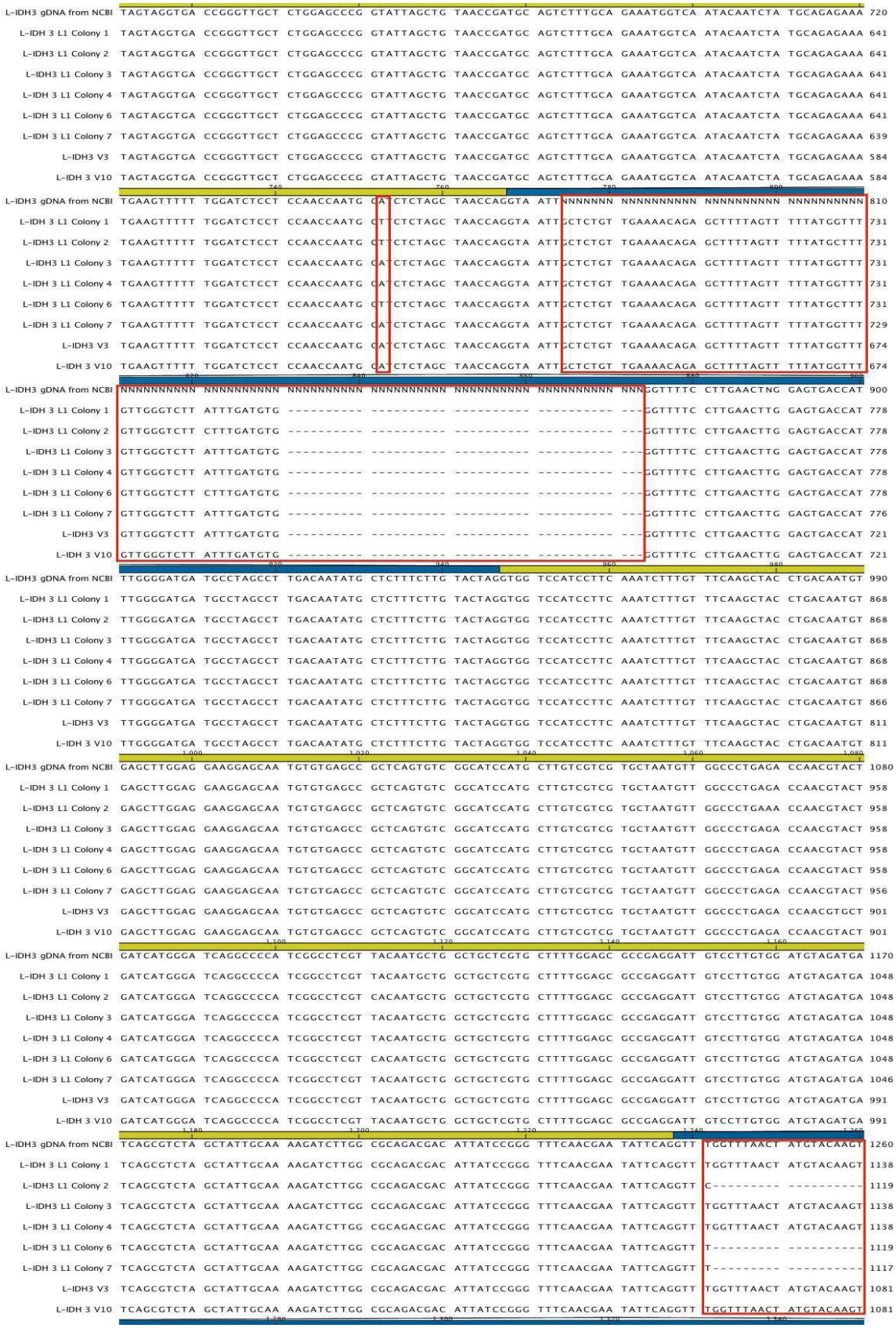
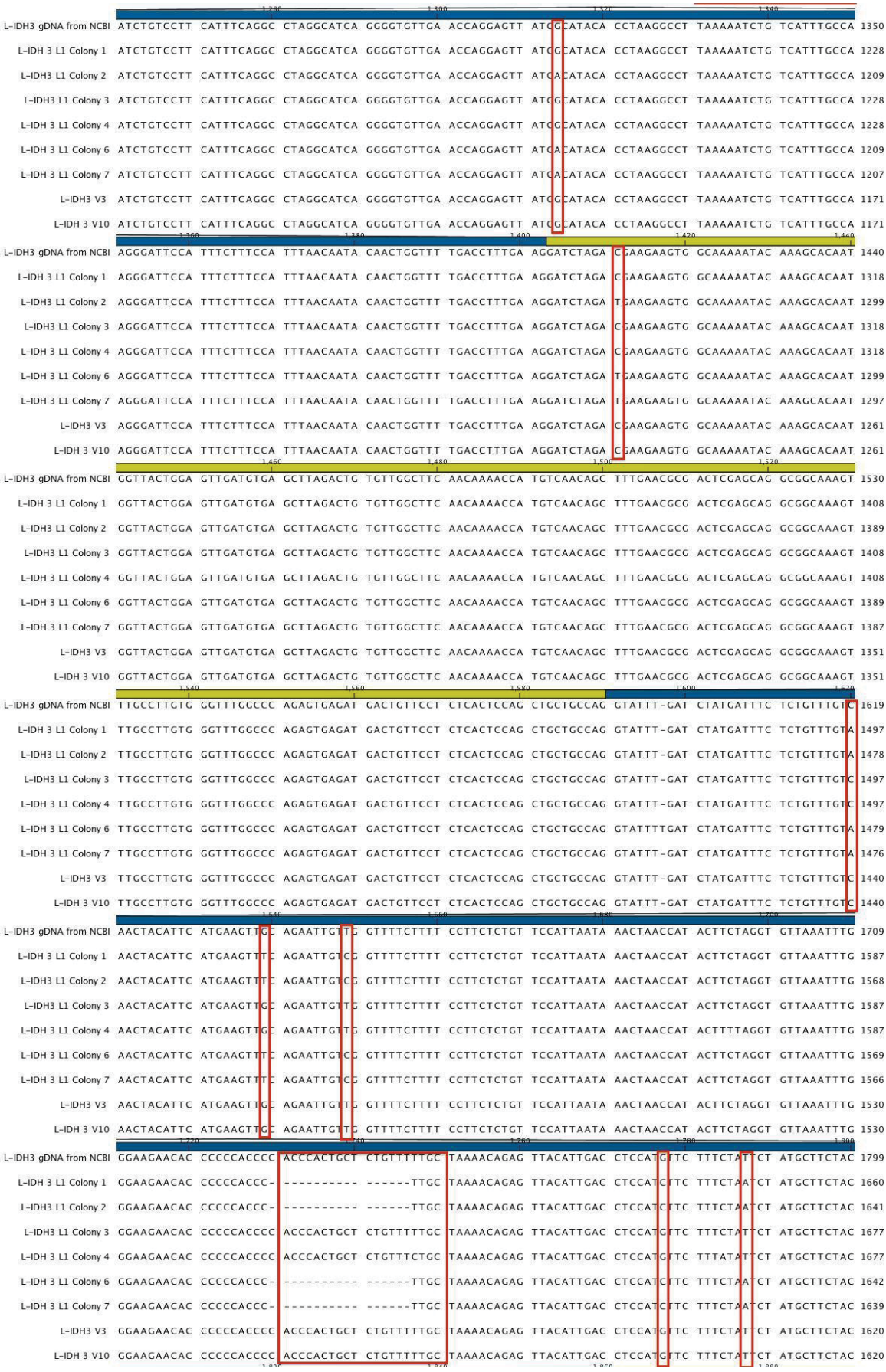

NCBI-LIDH2
L-IDH2 L1 Colony D
LIDH2 L1 Colony G

Figure A.2: L-idonate dehydrogenase 2 sequence from the heterozygous L1 vine.


L-IDH3 gDNA from NCBI
L-IDH 3 L1 Colony 1
L-IDH3 L1 Colony 2
L-IDH3 L1 Colony 3
L-IDH 3 L1 Colony 4
L-IDH 3 L1 Colony 6
L-IDH 3 L1 Colony 7
L-IDH3 V3
L-IDH 3 V10


L-IDH3 gDNA from NCBI
L-IDH 3 L1 Colony 1
L-IDH3 L1 Colony 2
L-IDH3 L1 Colony 3
L-IDH 3 L1 Colony 4
L-IDH 3 L1 Colony 6
L-IDH 3 L1 Colony 7
L-IDH3 V3
L-IDH 3 V10


L-IDH3 gDNA from NCBI ATCTGTCCTT CATTTCAGGC CTAGGCATCA GGGGTGTTGA ACCAGGAGTT ATCGCATACA CCTAAGGCCT TAAAAATCTG TCATTTGCCA 1350
L-IDH 3 L1 Colony 1 ATCTGTCCTT CATTTCAGGC CTAGGCATCA GGGGTGTTGA ACCAGGAGTT ATCGCATACA CCTAAGGCCT TAAAAATCTG TCATTTGCCA 1228
L-IDH3 L1 Colony 2 ATCTGTCCTT CATTTCAGGC CTAGGCATCA GGGGTGTTGA ACCAGGAGTT ATCACATACA CCTAAGGCCT TAAAAATCTG TCATTTGCCA 1209
L-IDH3 L1 Colony 3 ATCTGTCCTT CATTTCAGGC CTAGGCATCA GGGGTGTTGA ACCAGGAGTT ATCGCATACA CCTAAGGCCT TAAAAATCTG TCATTTGCCA 1228
L-IDH 3 L1 Colony 4 ATCTGTCCTT CATTTCAGGC CTAGGCATCA GGGGTGTTGA ACCAGGAGTT ATCGCATACA CCTAAGGCCT TAAAAATCTG TCATTTGCCA 1228
L-IDH 3 L1 Colony 6 ATCTGTCCTT CATTTCAGGC CTAGGCATCA GGGGTGTTGA ACCAGGAGTT ATCACATACA CCTAAGGCCT TAAAAATCTG TCATTTGCCA 1209
L-IDH 3 L1 Colony 7 ATCTGTCCTT CATTTCAGGC CTAGGCATCA GGGGTGTTGA ACCAGGAGTT ATCACATACA CCTAAGGCCT TAAAAATCTG TCATTTGCCA 1207
L-IDH3 V3 ATCTGTCCTT CATTTCAGGC CTAGGCATCA GGGGTGTTGA ACCAGGAGTT ATCGCATACA CCTAAGGCCT TAAAAATCTG TCATTTGCCA 1171
L-IDH 3 V10 ATCTGTCCTT CATTTCAGGC CTAGGCATCA GGGGTGTTGA ACCAGGAGTT ATCGCATACA CCTAAGGCCT TAAAAATCTG TCATTTGCCA 1171
L-IDH3 gDNA from NCBI AGGGATTCCA TTTCTTTCCA TTTAACAATA CAACTGGTTT TGACCTTTGA AGGATCTAGA CGAAGAAGTG GCAAAAATAC AAAGCACAAT 1440
L-IDH 3 L1 Colony 1 AGGGATTCCA TTTCTTTCCA TTTAACAATA CAACTGGTTT TGACCTTTGA AGGATCTAGA CGAAGAAGTG GCAAAAATAC AAAGCACAAT 1318
L-IDH3 L1 Colony 2 AGGGATTCCA TTTCTTTCCA TTTAACAATA CAACTGGTTT TGACCTTTGA AGGATCTAGA TGAAGAAGTG GCAAAAATAC AAAGCACAAT 1299
L-IDH3 L1 Colony 3 AGGGATTCCA TTTCTTTCCA TTTAACAATA CAACTGGTTT TGACCTTTGA AGGATCTAGA CGAAGAAGTG GCAAAAATAC AAAGCACAAT 1318
L-IDH 3 L1 Colony 4 AGGGATTCCA TTTCTTTCCA TTTAACAATA CAACTGGTTT TGACCTTTGA AGGATCTAGA CGAAGAAGTG GCAAAAATAC AAAGCACAAT 1318
L-IDH 3 L1 Colony 6 AGGGATTCCA TTTCTTTCCA TTTAACAATA CAACTGGTTT TGACCTTTGA AGGATCTAGA TGAAGAAGTG GCAAAAATAC AAAGCACAAT 1299
L-IDH 3 L1 Colony 7 AGGGATTCCA TTTCTTTCCA TTTAACAATA CAACTGGTTT TGACCTTTGA AGGATCTAGA TGAAGAAGTG GCAAAAATAC AAAGCACAAT 1297
L-IDH3 V3 AGGGATTCCA TTTCTTTCCA TTTAACAATA CAACTGGTTT TGACCTTTGA AGGATCTAGA CGAAGAAGTG GCAAAAATAC AAAGCACAAT 1261
L-IDH 3 V10 AGGGATTCCA TTTCTTTCCA TTTAACAATA CAACTGGTTT TGACCTTTGA AGGATCTAGA CGAAGAAGTG GCAAAAATAC AAAGCACAAT 1261
L-IDH3 gDNA from NCBI GGTTACTGGA GTTGATGTGA GCTTAGACTG TGTTGGCTTC AACAAAACCA TGTCAACAGC TTTGAACGCG ACTCGAGCAG GCGGCAAAGT 1530
L-IDH 3 L1 Colony 1 GGTTACTGGA GTTGATGTGA GCTTAGACTG TGTTGGCTTC AACAAAACCA TGTCAACAGC TTTGAACGCG ACTCGAGCAG GCGGCAAAGT 1408
L-IDH3 L1 Colony 2 GGTTACTGGA GTTGATGTGA GCTTAGACTG TGTTGGCTTC AACAAAACCA TGTCAACAGC TTTGAACGCG ACTCGAGCAG GCGGCAAAGT 1389
L-IDH3 L1 Colony 3 GGTTACTGGA GTTGATGTGA GCTTAGACTG TGTTGGCTTC AACAAAACCA TGTCAACAGC TTTGAACGCG ACTCGAGCAG GCGGCAAAGT 1408
L-IDH 3 L1 Colony 4 GGTTACTGGA GTTGATGTGA GCTTAGACTG TGTTGGCTTC AACAAAACCA TGTCAACAGC TTTGAACGCG ACTCGAGCAG GCGGCAAAGT 1408
L-IDH 3 L1 Colony 6 GGTTACTGGA GTTGATGTGA GCTTAGACTG TGTTGGCTTC AACAAAACCA TGTCAACAGC TTTGAACGCG ACTCGAGCAG GCGGCAAAGT 1389
L-IDH 3 L1 Colony 7 GGTTACTGGA GTTGATGTGA GCTTAGACTG TGTTGGCTTC AACAAAACCA TGTCAACAGC TTTGAACGCG ACTCGAGCAG GCGGCAAAGT 1387
L-IDH3 V3 GGTTACTGGA GTTGATGTGA GCTTAGACTG TGTTGGCTTC AACAAAACCA TGTCAACAGC TTTGAACGCG ACTCGAGCAG GCGGCAAAGT 1351
L-IDH 3 V10 GGTTACTGGA GTTGATGTGA GCTTAGACTG TGTTGGCTTC AACAAAACCA TGTCAACAGC TTTGAACGCG ACTCGAGCAG GCGGCAAAGT 1351
L-IDH3 gDNA from NCBI TTGCCTTGTG GGTTTGGCCC AGAGTGAGAT GACTGTTCCT CTCACTCCAG CTGCTGCCAG GTATTT-GAT CTATGATTTC TCTGTTTGTC 1619
L-IDH 3 L1 Colony 1 TTGCCTTGTG GGTTTGGCCC AGAGTGAGAT GACTGTTCCT CTCACTCCAG CTGCTGCCAG GTATTT-GAT CTATGATTTC TCTGTTTGTA 1497
L-IDH3 L1 Colony 2 TTGCCTTGTG GGTTTGGCCC AGAGTGAGAT GACTGTTCCT CTCACTCCAG CTGCTGCCAG GTATTT-GAT CTATGATTTC TCTGTTTGTA 1478
L-IDH3 L1 Colony 3 TTGCCTTGTG GGTTTGGCCC AGAGTGAGAT GACTGTTCCT CTCACTCCAG CTGCTGCCAG GTATTT-GAT CTATGATTTC TCTGTTTGTC 1497
L-IDH 3 L1 Colony 4 TTGCCTTGTG GGTTTGGCCC AGAGTGAGAT GACTGTTCCT CTCACTCCAG CTGCTGCCAG GTATTT-GAT CTATGATTTC TCTGTTTGTC 1497
L-IDH 3 L1 Colony 6 TTGCCTTGTG GGTTTGGCCC AGAGTGAGAT GACTGTTCCT CTCACTCCAG CTGCTGCCAG GTATTTTGAT CTATGATTTC TCTGTTTGTA 1479
L-IDH 3 L1 Colony 7 TTGCCTTGTG GGTTTGGCCC AGAGTGAGAT GACTGTTCCT CTCACTCCAG CTGCTGCCAG GTATTT-GAT CTATGATTTC TCTGTTTGTA 1476
L-IDH3 V3 TTGCCTTGTG GGTTTGGCCC AGAGTGAGAT GACTGTTCCT CTCACTCCAG CTGCTGCCAG GTATTT-GAT CTATGATTTC TCTGTTTGTC 1440
L-IDH 3 V10 TTGCCTTGTG GGTTTGGCCC AGAGTGAGAT GACTGTTCCT CTCACTCCAG CTGCTGCCAG GTATTT-GAT CTATGATTTC TCTGTTTGTC 1440
L-IDH3 gDNA from NCBI AACTACATTC ATGAAGTTGC AGAATTGTTG GTTTTCTTTT CCTTCTCTGT TCCATTAATA AACTAACCAT ACTTCTAGGT GTTAAATTTG 1709
L-IDH 3 L1 Colony 1 AACTACATTC ATGAAGTTTC AGAATTGTCG GTTTTCTTTT CCTTCTCTGT TCCATTAATA AACTAACCAT ACTTCTAGGT GTTAAATTTG 1587
L-IDH3 L1 Colony 2 AACTACATTC ATGAAGTTTC AGAATTGTCG GTTTTCTTTT CCTTCTCTGT TCCATTAATA AACTAACCAT ACTTCTAGGT GTTAAATTTG 1568
L-IDH3 L1 Colony 3 AACTACATTC ATGAAGTTGC AGAATTGTTG GTTTTCTTTT CCTTCTCTGT TCCATTAATA AACTAACCAT ACTTCTAGGT GTTAAATTTG 1587
L-IDH 3 L1 Colony 4 AACTACATTC ATGAAGTTGC AGAATTGTTG GTTTTCTTTT CCTTCTCTGT TCCATTAATA AACTAACCAT ACTTTTAGGT GTTAAATTTG 1587
L-IDH 3 L1 Colony 6 AACTACATTC ATGAAGTTTC AGAATTGTCG GTTTTCTTTT CCTTCTCTGT TCCATTAATA AACTAACCAT ACTTCTAGGT GTTAAATTTG 1569
L-IDH 3 L1 Colony 7 AACTACATTC ATGAAGTTTC AGAATTGTCG GTTTTCTTTT CCTTCTCTGT TCCATTAATA AACTAACCAT ACTTCTAGGT GTTAAATTTG 1566
L-IDH3 V3 AACTACATTC ATGAAGTTGC AGAATTGTTG GTTTTCTTTT CCTTCTCTGT TCCATTAATA AACTAACCAT ACTTCTAGGT GTTAAATTTG 1530
L-IDH 3 V10 AACTACATTC ATGAAGTTGC AGAATTGTTG GTTTTCTTTT CCTTCTCTGT TCCATTAATA AACTAACCAT ACTTCTAGGT GTTAAATTTG 1530
L-IDH3 gDNA from NCBI GGAAGAACAC CCCCCACCCC ACCCACTGCT CTGTTTTTGC TAAAACAGAG TTACATTGAC CTCCATGTTC TTTCTATTCT ATGCTTCTAC 1799
L-IDH 3 L1 Colony 1 GGAAGAACAC CCCCCACCC- ---------- ------TTGC TAAAACAGAG TTACATTGAC CTCCATCTTC TTTCTAATCT ATGCTTCTAC 1660
L-IDH3 L1 Colony 2 GGAAGAACAC CCCCCACCC- ---------- ------TTGC TAAAACAGAG TTACATTGAC CTCCATCTTC TTTCTAATCT ATGCTTCTAC 1641
L-IDH3 L1 Colony 3 GGAAGAACAC CCCCCACCCC ACCCACTGCT CTGTTTTTGC TAAAACAGAG TTACATTGAC CTCCATGTTC TTTCTATTCT ATGCTTCTAC 1677
L-IDH 3 L1 Colony 4 GGAAGAACAC CCCCCACCCC ACCCACTGCT CTGTTTCTGC TAAAACAGAG TTACATTGAC CTCCATGTTC TTTATATTCT ATGCTTCTAC 1677
L-IDH 3 L1 Colony 6 GGAAGAACAC CCCCCACCC- ---------- ------TTGC TAAAACAGAG TTACATTGAC CTCCATCTTC TTTCTAATCT ATGCTTCTAC 1642
L-IDH 3 L1 Colony 7 GGAAGAACAC CCCCCACCC- ---------- ------TTGC TAAAACAGAG TTACATTGAC CTCCATCTTC TTTCTAATCT ATGCTTCTAC 1639
L-IDH3 V3 GGAAGAACAC CCCCCACCCC ACCCACTGCT CTGTTTTTGC TAAAACAGAG TTACATTGAC CTCCATGTTC TTTCTATTCT ATGCTTCTAC 1620
L-IDH 3 V10 GGAAGAACAC CCCCCACCCC ACCCACTGCT CTGTTTTTGC TAAAACAGAG TTACATTGAC CTCCATGTTC TTTCTATTCT ATGCTTCTAC 1620

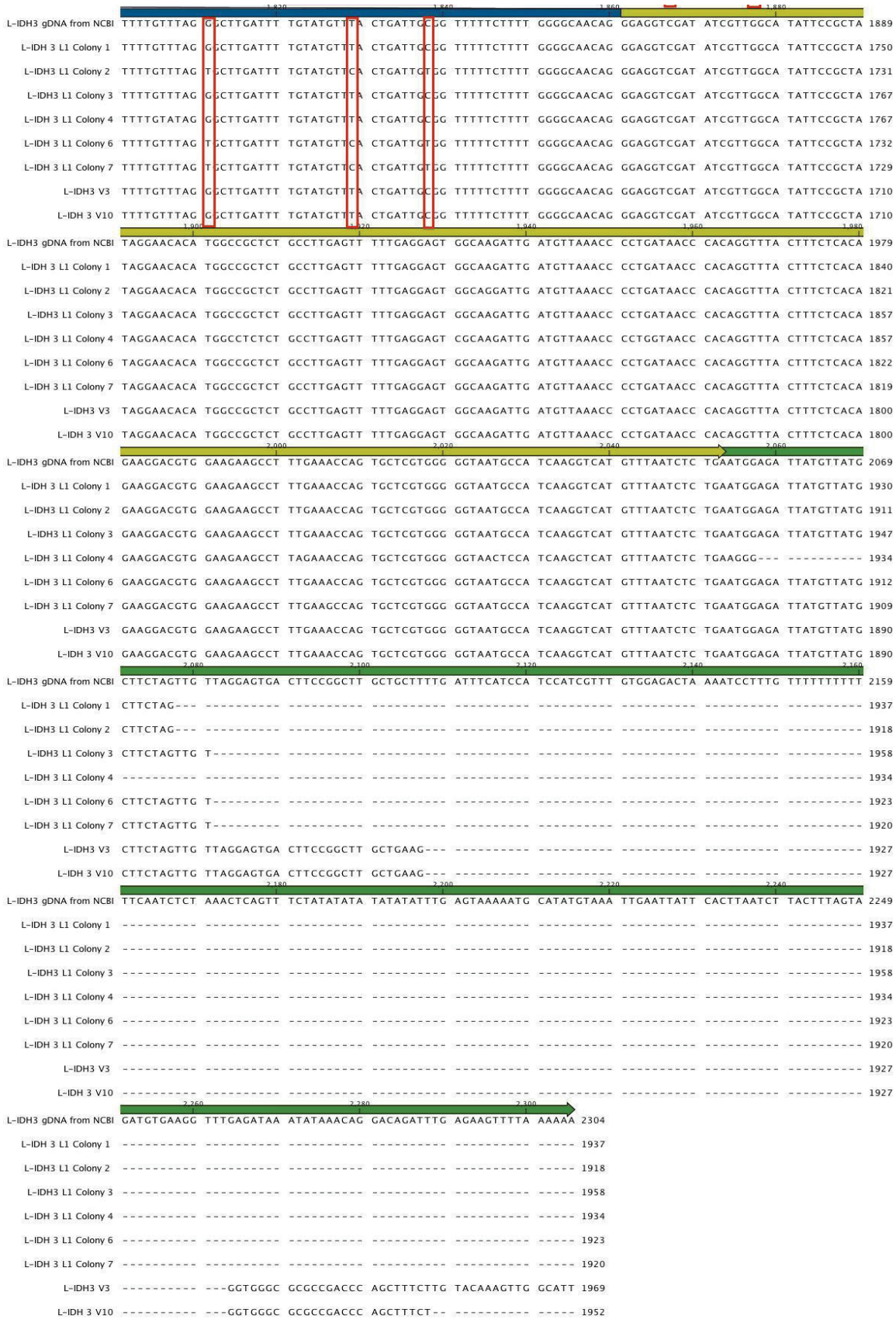


Figure A.3: L-idonate dehydrogenase 3 sequence from the V3 and V10 homozygous microvine line and the heterozygous L1 parent.

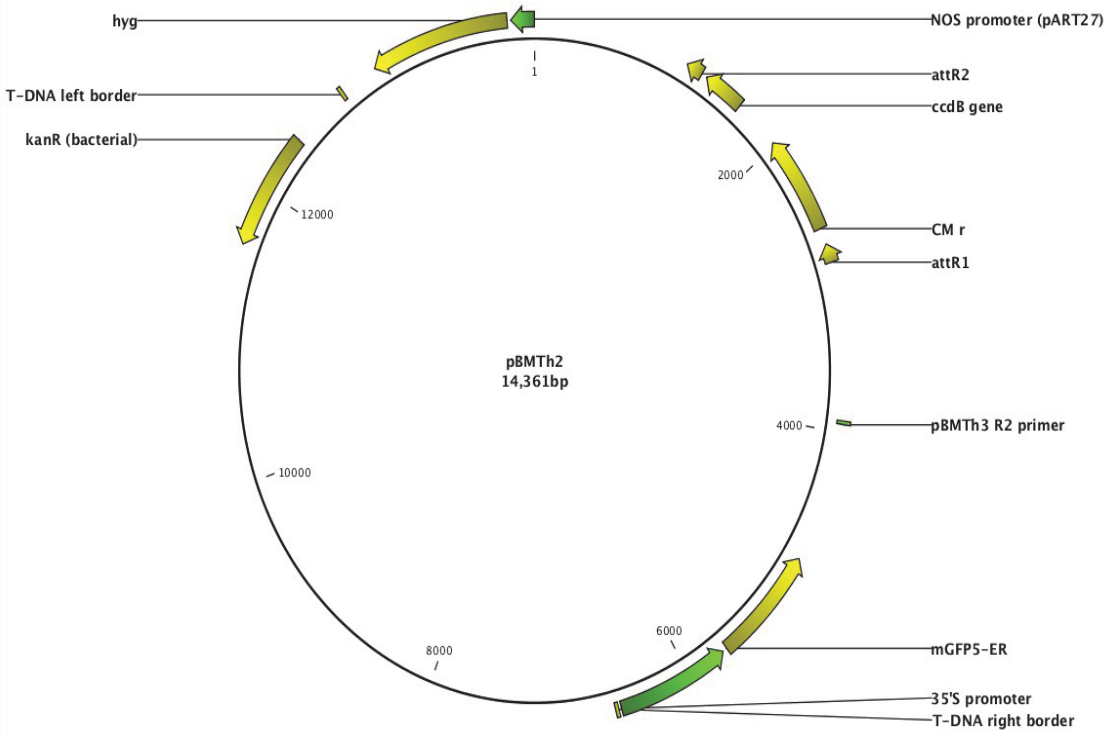


Figure A.4: Vector map of pBMTh2 vector used as the binary vector for the transformation of L-IDH isoforms into microvines.

Table A.1: Markers with P-values below 0.05 for the genome wide association study Only markers for population 9 tartaric acid showed any significance in the two seasons. P-values were corrected for false discovery rate using the Benjamini-Hochberg method.

Season	Acid/time point	Chromosome	Marker	P-value
2012-2013	Mature berry tartaric acid	19	S19_1513794	0.016
2013-2014	Pre-veraison tartaric acid	5	S5_487880	0.012
			S5_610996	0.012
			S5_681547	0.012
			S5_681548	0.012
			S5_753946	0.012
			S5_806219	0.012
			S5_984862	0.012
			S5_984889	0.012
			S5_988836	0.012
			S5_1310550	0.012
			S5_1310552	0.012
			S5_1338430	0.012
			S5_1421606	0.012
			S5_1421594	0.012
			S5_1425109	0.012
			S5_2907050	0.012
			S5_3137792	0.013
			S5_1419308	0.015
			S5_2074874	0.016
			S5_3149630	0.017
			S5_770050	0.022
			S5_2100686	0.023
			S5_5060527	0.024
			S5_5060542	0.024
			S5_5850104	0.033
			S5_5850143	0.033
			S5_2734946	0.039
			S5_2393425	0.039

			S5_2516576	0.039
			S5_2743133	0.039
			S5_2743201	0.039
2013-2014	Pre-veraison tartaric acid	5	S5_2502879	0.040
			S5_1419940	0.041
			S5_2100590	0.041
			S5_2724772	0.048
			S5_3245688	0.048
		7	S7_4125660	0.012
			S7_4125667	0.012
			S7_4670664	0.044
		19	S19_10557867	0.017
2013-2014	Mature berry tartaric acid	5	S5_487880	0.007
			S5_610996	0.007
			S5_681547	0.007
			S5_681548	0.007
			S5_753946	0.007
			S5_770050	0.007
			S5_984862	0.007
			S5_984889	0.007
			S5_988836	0.007
			S5_1310550	0.007
			S5_1310552	0.007
			S5_1338430	0.007
			S5_1421594	0.007
			S5_1421606	0.007
			S5_1419308	0.007
			S5_1425109	0.007
			S5_2074874	0.007
			S5_2100590	0.002
			S5_2393425	0.007
			S5_2502879	0.007

			S5_2516576	0.007
			S5_2660672	0.007
			S5_2665411	0.007
			S5_2665416	0.007
			S5_2724772	0.007
			S5_2734946	0.007
			S5_2735200	0.007
			S5_2743133	0.007
			S5_2743201	0.007
			S5_2100686	0.008
			S5_2548953	0.010
			S5_2378483	0.016
			S5_2907050	0.018
			S5_1419940	0.021
			S5_806219	0.025
			S5_2630064	0.025
			S5_3137792	0.025
			S5_2670397	0.027
			S5_2753449	0.037
			S5_3149630	0.037
		11	S11_5230059	0.040
		19	S19_10557867	0.007
			S19_1513794	0.025

Table A.2: Markers with LOD values above 3 for the QTL analysis. Included are the two seasons for both pre-veraison malic and tartaric acid, and mature berry malic and tartaric acid.

Season	Acid/time point	Population/ map	Chromosome	Marker	Position	LOD
2013-2013	Mature berry tartaric acid	10 male parent	18	S18_5834633	537.77	3.45
				S18_4308599	236.77	3.42
				S18_4069655	459.31	3.39
				S18_4268742	239.12	3.39
				S18_4098631	459.31	3.39
				S18_4098595	459.31	3.39
				S18_8646731	194.26	3.33
				S18_4859473	465.20	3.33
				S18_4316856	235.59	3.15
				S18_4728424	467.56	3.03
				S18_5173523	480.58	3.03
				S18_5193869	480.58	3.03
				S18_4160233	458.13	3.02
2013-2013	Mature berry tartaric acid	6 female parent	12	S12_8261721	321.66	3.76
				S12_10162340	415.10	3.45
				S12_9752921	413.69	3.07
				S12_9373036	413.69	3.07
				S12_9600067	443.36	3.20
				S12_10161764	443.36	3.20
				S12_10161807	443.36	3.20
				S12_9669737	443.36	3.20
				S1200_10307889	447.59	3.07
				S1200_305820	447.59	3.07
				S1200_130467	447.59	3.07

				S1200_130438	447.59	3.07
				S12_13364045	447.59	3.07
				S12_10203489	430.64	3.02
2012-2013	Mature berry malic acid	10 female parent	3	S3_12209946	158.06	3.30
				S3_10927331	158.06	3.18
				S3_10937305	158.06	3.18
			1	S1_3637717	1020.01	3.06
				S1_3843208	240.66	3.04
				S1_5314777	253.67	3.02
		3	8	S8_15322186	246.46	3.43
				S8_152322209	246.46	3.43
		6 male parent	1	S1_3882510	250.52	3.99
				S1_4244946	106.69	3.77
				S1_4478492	106.69	3.77
				S1_4467861	106.69	3.77
				S1_3168390	106.69	3.77
				S1_4946929	106.69	3.77
				S1_3664806	116.57	3.33
				S1_4948628	166.67	3.33
				S1_4189667	136.39	3.23
				S1_471979	110.93	3.06
				S1_3845512	113.75	3.06
				S1_4912640	113.75	3.06
				S1_4855597	113.75	3.06
				S1_3845544	113.75	3.06
2013-2014	Pre-veraison malic acid	10 male parent	9	S9_22302105	389.86	3.49
				S9_22501159	265.11	3.07

				S9_21898862	378.07	3.06
				S9_21947986	378.07	3.06
				S9_22437349	378.07	3.06
				S9_22302124	379.24	3.06
				S9_22302123	379.24	3.06
2013-2014	Pre-veraison malic acid	9	1	S1_3754326	282.76	3.17
				S1_3754308	282.76	3.17
		10 male	9	S9_22302105	389.86	3.47
				S9_22501159	365.11	3.07
				S9_21898862	378.07	3.06
				S9_22437349	378.07	3.06
				S9_22302124	378.07	3.06
				S9_22302123	378.07	3.06
				S9_21947986	378.07	3.06
2013-2014	Mature berry malic acid	10 female parent	1	S1_4189668	228.84	3.02
2013-2014	Pre-veraison tartaric acid	9	5	S5_1425109	1809.82	3.93
				S5_988836	1824.14	3.88
				S5_1419308	1824.14	3.88
				S5_1421606	1824.14	3.88
				S5_1310550	1825.72	3.88
				S5_1310552	1825.72	3.88
				S5_487880	1820.96	3.84
				S5_681547	1822.55	3.84
				S5_681548	1822.55	3.84
				S5_610996	1822.55	3.84
				S5_753946	1822.55	3.84
				S5_806219	1817.77	3.84
				S5_984889	1822.55	3.84
				S5_984864	1822.55	3.84
				S5_1338430	1822.55	3.84
				S5_1421594	1822.55	3.84
				S5_2074874	1820.96	3.84

2013-2014	Pre-veraison tartaric acid	9	5	S5_2100686	1819.37	3.84
				S5_2100590	1820.96	3.84
				S5_2907050	1797.08	3.80
				S5_3137792	1795.49	3.77
				S5_3149630	1795.49	3.77
				S19_10557867	1196.04	3.56
				S5_5060542	1697.64	3.52
				S5_5060527	1697.64	3.52
				S5_770050	1811.41	3.42
				S5_5850143	1662.96	3.28
				S5_5850104	1662.96	3.28
				S5_1419940	1798.67	3.05
				S5_2724772	1798.57	3.05
				S5_2743201	1798.57	3.05
				S5_2735133	1798.57	3.05
				S5_2735200	1798.57	3.05
				S5_2734946	1798.57	3.05
				S5_2548953	1800.26	3.03
				S5_2516576	1813.00	3.03
				S5_2393425	1800.26	3.02
			7	S7_4125660	2287.92	3.61
				S7_4125667	2287.92	3.61
	Mature berry tartaric acid	9	5	S5_487880	1820.96	4.65
				S5_2074874	1820.96	4.65
				S5_2100590	1820.96	4.65
				S19_10557867	1196.04	4.15
				S5_988836	1824.14	4.07
				S5_1310550	1825.72	4.07
				S5_1310552	1825.72	4.07
				S5_1421606	1824.14	4.07
				S5_1419308	1824.14	4.07

				S5_2660672	1787.49	4.07
				S5_2743201	1798.67	4.07
				S5_2735133	1798.67	4.07
				S5_2735200	1798.67	4.07
2013-2014	Mature berry tartaric acid	9	5	S5_2734946	1798.67	4.07
				S5_2724772	1789.67	4.07
				S5_1425109	1809.82	3.89
				S5_2393425	1800.26	3.89
				S5_2516576	1813.00	3.89
				S5_2548953	1800.26	3.89
				S5_681547	1822.55	3.87
				S5_681548	1822.55	3.87
				S5_610996	1822.55	3.87
				S5_1421594	1822.55	3.87
				S5_753946	1822.55	3.87
				S5_984889	1822.55	3.87
				S5_984864	1822.55	3.87
				S5_1338430	1822.55	3.87
				S5_2100686	1819.37	3.82
				S5_238483	1803.44	3.42
				S5_2502879	1801.85	3.42
				S5_770050	1811.41	3.39
				S5_290750	1797.03	3.36
				S5_1419940	1813.00	3.29
				S5_2665411	1773.08	3.26
				S5_2665416	1772.08	3.26
				S5_2630064	1779.48	3.26
				S5_3137792	1795.49	3.18
				S5_3149630	1795.49	3.18
				S5_806219	1817.77	3.17
				S5_2670397	1766.71	3.15
				S5_2753449	1769.89	3.07
			11	S11_5230059	199.16	3.02

2013-2014	Mature berry tartaric acid	10 female parent	5	S5_5290946	675.01	3.89
				S5_6756112	649.81	3.10

Included paper

“A high-throughput UHPLC-MS/MS method for evaluation of tartaric and malic acid concentration in individual grapevine berries”

Higginson, E.G., Lloyd, N.D.R., Kravchuk, O., Ford, C.M. and Thomas, M.R. (2015).
A high-throughput UHPLC MS/MS method for evaluation of tartaric and malic acid concentration in individual grapevine berries.
Australian Journal of Grape and Wine Research, 22(1), 16-23.

Reference List

Agudelo-Romero, P., Erban, A., Sousa, L., Pais, M.S., Kopka, J. and Fortes, A.M., (2013) Search for transcriptional and metabolic markers of grape pre-ripening and ripening and insights into specific aroma development in three Portuguese cultivars. PLoS One **8**, 21.

Arroyo-Garcia, R., Ruiz-Garcia, L., Bolling, L., Ocete, R., Lopez, M.A., Arnold, C., Ergul, A., Soylemezoglu, G., Uzun, H.I., Cabello, F., Ibanez, J., Aradhya, M.K., Atanassov, A., Atanassov, I., Balint, S., Cenis, J.L., Costantini, L., Goris-Lavets, S., Grando, M.S., Klein, B.Y., McGovern, P.E., Merdinoglu, D., Pejic, I., Pelsy, F., Primikirios, N., Risovannaya, V., Roubelakis-Angelakis, K.A., Snoussi, H., Sotiri, P., Tamhankar, S., This, P., Troshin, L., Malpica, J.M., Lefort, F. and Martinez-Zapater, J.M., (2006) Multiple origins of cultivated grapevine (Vitis vinifera L. ssp sativa) based on chloroplast DNA polymorphisms. Molecular Ecology **15**, 3707–3714.

Atwell, S., Huang, Y.S., Vilhjalmsson, B.J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A.M., Hu, T.T., Jiang, R., Muliyati, N.W., Zhang, X., Amer, M.A., Baxter, I., Brachi, B., Chory, J., Dean, C., Debieu, M., de Meaux, J., Ecker, J.R., Faure, N., Kniskern, J.M., Jones, J.D.G., Michael, T., Nemri, A., Roux, F., Salt, D.E., Tang, C., Todesco, M., Traw, M.B., Weigel, D., Marjoram, P., Borevitz, J.O., Bergelson, J. and Nordborg, M., (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature **465**, 627–631.

Australian Bureau of Meteorology, (2015) Climate Data online [WWW Document]. URL http://www.bom.gov.au/climate/data/?ref=ftr

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. and Johnson, E.A., (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One **3**, e3376.

Barba, P., Cadle-Davidson, L., Harriman, J., Glaubitz, J.C., Brooks, S., Hyma, K. and Reisch, B., (2014) Grapevine powdery mildew resistance and susceptibility loci identified on a high-resolution SNP map. Theoretical and Applied Genetics **127**, 73–84.

Battilana, J., Emanuelli, F., Gambino, G., Gribaudo, I., Gasperi, F., Boss, P.K. and Grando, M.S., (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.

Benjamini, Y. and Hochberg, Y., (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, B (methodological) **57**, 289–300.

Bergqvist, J., Dokoozlian, N. and Ebisuda, N., (2001) Sunlight exposure and temperature effects on berry growth and composition of Cabernet Sauvignon and Grenache in the central San Joaquin Valley of California. American Journal of Enology and Viticulture **52**, 1–7.

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

Boulton, R., (1980) The general relationship between potassium, sodium and pH in grape juice and wine. American Journal of Enology and Viticulture **31**, 182–186.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S., (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics **23**, 2633–2635.

Broman, K.W., Wu, H., Sen, S. and Churchill, G.A., (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics **19**, 889–890.

Burbidge, C.A., (2011) Identification and Characterisation of the Enzymes Involved in the

Biosynthetic Pathway of Tartaric Acid in Vitis vinifera. PhD Thesis, Flinders University, South Australia.
Castellari, M., Varsari, A., Spinabelli, U., Galassi, S. and Amati, A., (2000) An improved HPLC method for the analysis of organic acids, and carbohydrates, and alcohols in grape musts and wines. Journal of Liquid Chromatography and Related Technologies **23**, 2047–2056.
Chaib, J., Torregrosa, L., Mackenzie, D., Corena, P., Bouquet, A. and Thomas, M.R., (2010) The grape microvine- a model system for rapid forward and reverse genetics of grapevines. The Plant Journal **62**, 1083–1092.
Chen, J., Wang, N., Fang, L.-C., Liang, Z.-C., Li, S.-H. and Wu, B.-H., (2015) Construction of a high-density genetic map and QTLs mapping for sugars and acids in grape berries. BMC Plant Biology **15**, 28.
Chinnici, F., Spinabelli, U., Riponi, C. and Amati, A., (2005) Optimization of the determination of organic acids and sugars in fruit juices by ion-exclusion liquid chromatography. Journal of Food Composition and Analysis, **18**, 121–130.
Cohen, S., Tzuri, G., Harel-Beja, R., Itkin, M., Portnoy, V., Sa'ar, U., Lev, S., Yeselson, L., Petrikov, M., Rogachev, I., Aharoni, A., Ophir, R., Tadmor, Y., Lewinsohn, E., Burger, Y., Katzir, N. and Schaffer, A.A., (2012) Co-mapping studies of QTLs for fruit acidity and candidate genes of organic acid metabolism and proton transport in sweet melon (Cucumis melo L.). Theoretical and Applied Genetics, **125**, 343–353.
Coombe, B.. G., (1960) Relationship of growth and development to changes in sugars, auxins, gibberellins in fruit of seeded and seedless varieties of Vitis vinifera. Plant Physiology **35**, 241–250.
Coombe, B.G., (1992) Research on Development and ripening of the grape berry. American Journal of Enology and Viticulture, **43**, 101–110.
Coombe, B.G. and Jones, G.P., (1983) Development of the grape berry .4. measurement of the changes in the composition of developing undetected grape berries by using c-13 NMR techniques. Phytochemistry **22**, 2185–2187.
Correa, J.,Mamani, M., Munoz-Espinoza, C., Laborie, D., Munoz, C., Pinto, M., Hinrichsen, P., (2014) Heritability and identification of QTLs and underlying candidate genes associated with the architecture of the grapevine cluster (Vitis vinifera L.). Theoretical and Applied Genetics, **127**, 1143-1152.
Crippen, D.D. and Morrison, J.C., (1986) The effects of sun exposure on the compositional development of Cabernet-Sauvignon berries. Am. J. Enol. Vitic. **37**, 235–242.
Dai, Z.W., Ollat, N., Gomes, E., Decroocq, S., Tandonnet, J., van Leeuwen, C., Pieri, P., Hilbert, G., Kappel, C., Bordenave, L., Vivin, P. and Delrot, S., (2011) Ecophysical, genetic, and molecular causes of variation in grape berry weight and composition: a review. American Journal of Enology and Viticulture **62**, 413–425.
Davies, C., Shin, R., Liu, W., Thomas, M.R. and Schachtman, D.P., (2006) Transporters expressed during grape berry (Vitis vinifera L.) development are associated with an increase in berry size and berry potassium accumulation. Journal of Experimental Botany, **57**, 3209–3216.
De Angeli, A., Baetz, U., Francisco, R., Zhang, J.B., Chaves, M.M. and Regalado, A., (2013) The vacuolar channel VvALMT9 mediates malate and tartrate accumulation in berries of Vitis vinifera. Planta **238**, 283–291.
DeBolt, S., (2006) L-Tartaric acid synthesis in plants. PhD Thesis, The University of Adelaide, South Australia.
DeBolt, S., Cook, D.R. and Ford, C.M., (2006) L-Tartaric acid synthesis from vitamin C in higher plants. PLoS One, **103**, 5608–5613.

DeBolt, S., Hardie, J., Tyerman, S. and Ford, C.M., (2004) Composition and synthesis of raphide crystals and druse crystals in berries of Vitis vinifera L. cv. Cabernet Sauvignon: Ascorbic acid as precursor for both oxalic and tartaric acids as revealed by radiolabelling studies. Australian Journal of Grape Wine Research, **10**, 134–142.

DeBolt, S., Melino, V. and Ford, C.M., (2007) Ascorbate as a biosynthetic precursor in plants. Annuals of Botany, **99**, 3–8.

DeBolt, S., Ristic, R., Iland, P.G. and Ford, C.M., (2008) Altered light interception reduces grape berry weight and modulates organic acid biosynthesis during development. Hortscience **43**, 957–961.

Deluc, L.G., Grimplet, J., Wheatley, M.D., Tillett, R.L., Quilici, D.R., Osborne, C., Schooley, D.A., Schlauch, K.A., Cushman, J.C. and Cramer, G.R., (2007) Transcriptomic and metabolite analyses of Cabernet Sauvignon grape berry development. BMC Genomics **8**, 429.

Dempster, A.P., Laird, N.M. and Rubin, D.B., (1977) Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (methodological) 1–38.

Deschamps, S., Llaca, V. and May, G.D., (2012) Genotyping-by-sequencing in plants. Biology **1**, 460–483.

Duchene, E., Dumas, V., Jaegli, N. and Merdinoglu, D., (2014) Generic variability of descriptors for grapevine berry acidity in Riesling, Gewurztraminer and their progeny. Australian Journal of Grape and Wine Research **20**, 91–99.

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

Ehling, S. and Cole, S., (2011) Analysis of Organic Acids in Fruit Juices by Liquid Chromatography-Mass Spectrometry: An Enhanced Tool for Authenticity Testing. Journal of Agricultural and Food Chemistry, **59**, 2229–2234.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E., (2011) A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. PLoS ONE **6**, e19379.

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

Famiani, F., Walker, R.P., Tecsi, L., Chen, Z., Proietti, P. and Leegood, R.C., (2000) An immunohisochemical study of the compartmentation of metabolism during the development of grape (Vitis vinifera L.) berries. Journal of Experemental Botany **51**, 675–683.

Fernandez, L., Chaib, J., Martinez-Zapater, J.-M., Thomas, M.R. and Torregrosa, L., (2013) Mis-expression of a PISTILLATA-like MADS box gene prevents fruit development in grapevine. Plant J. **73**, 918–928.

Flores, P., Hellin, P. and Fenoll, J., (2012) Determination of organic acids in fruits and vegetables by liquid chromatography with tandem-mass spectrometry. Food Chemistry **132**, 1049–1054.

Fodor, A., Segura, V., Denis, M., Neuenschwander, S., Fournier-Level, A., Chatelet, P., Homa, F.A.A., Lacombe, T., This, P. and Le Cunff, L., (2014) genome-wide prediction methods in highly diverse and heterozygous species: proof-of-concept through simulation in grapevine. PLoS ONE **9**, e110436.

Fontes, N., Geros, H. and Delrot, D., (2011) Grapevine berry vacuole: a complex and heterogeneous membrane system specilized in the accumulation of solutes. American Journal of Enology and Viticulture **63**, 270–278.

Ford, C.M., (2012) The biochemistry of organic acids in the grape, in: The Biochemistry of the Grape. Bentham Science Publishers, pp. 67–88.
Franks, T., He, D.G. and Thomas, M., (1998) Regeneration of transgenic Vitis vinifera L. Sultana plants: genotypic and phenotypic analysis. Molecular Breeding **4**, 321–333.
Frayne, R.F., (1986) Direct analysis of the major organic-components in grape must and wine using high-performance liquid-chromatography. American Journal of Enology and Viticulture, **37**, 281–287.
Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., Sun, Q. and Buckler, E.S., (2014) TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One **9**, 11.
Gonzalez-Caballero, V., Perez-Marin, D., Lopez, M.I. and Sanchez, M.T., (2011) Optimization of NIR spectral data management for quality control of grape bunches during on-vine ripening. Sensors, **11**, 6109–6124.
Gouthu, S., O'Neil, S.T., Di, Y., Ansarolia, M., Megraw, M. and Deluc, L.G., (2014) A comparative study of ripening among berries of the grape cluster reveals an altered transcriptional program and enhanced ripening rate in delayed berries. Journal of Experimental Botany, **65**, 5889–5902.
Hall, A. and Jones, G.V., (2009) Effect of potential atmospheric warming on temperature-based indices describing Australian wine grape growing conditions. Australian Journal of Grape and Wine Research, **15**, 97–119.
Hancock, R.D. and Viola, R., (2005) Biosynthesis and catabolism of L-ascorbic acid in plants. Critical Reviews in Plant Science, **24**, 167–188.
He, J., Zhao, X., Laroche, A., Lu, Z.-X., Liu, H. and Li, Z., (2014) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in Plant Science, **5**, 484.
Helsper, J. and Loewus, F.A., (1982) Metabolism of L-threonic acid in Rumex X-acutus and Pelargonium crispum (L) her. Plant Physiology, **69**, 1365–1368.
Higginson, E.G., Lloyd, N.D.R., Kravchuk, O., Ford, C.M. and Thomas, M.R., (2015) A high throughput UHPLC MS/MS method for evaluation of tartaric and malic acid concentrations in individual grapevine berries. Australian Journal of Grape and Wine Research.
Hrazdina, G., Parsons, G. and Mattick, L., (1984) Physiological and biochemical events during development and maturation. American Journal of Enology and Viticulture, **35**, 220–227.
Hyma, K., Barba, P., Wang, M., Londo, J.P., Acharya, C.B., Mitchell, S.E., Sun, Q., Reisch, B., Cadel-Davidson, L., (2015) Heterozygous mapping strategy (HetMappS) for high resolution genotyping-by-sequencing markers: a case study in grapevine. PLoS One, **10.**
Iland, P.G., (2011) The Grapevine: From the science to the practice of growing vines for wine. Patrick Iland Wine Promotions.
Iland, P.G. and Coombe, B.G., (1988) Malate, tartrate, potassium, and sodium in flesh and skin of Shiraz grapes during ripening- concentration and compartmentation. American Journal of Enology and Viticulture, **39**, 71–76.
Iocco, P., Franks, T. and Thomas, M.R., (2001) Genetic transformation of major wine grape cultivars of Vitis vinifera L. Transgenic Research, **10**, 105–112.
Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyere, C., Billault, A., Segurens, B., Gouyvenoux, M., Ugarte, E., Cattonaro, F., Anthouard, V., Vico, V., Del Fabbro,

C., Alaux, M., Di Gaspero, G., Dumas, V., Felice, N., Paillard, S., Juman, I., Moroldo, M., Scalabrin, S., Canaguier, A., Le Clainche, I., Malacrida, G., Durand, E., Pesole, G., Laucou, V., Chatelet, P., Merdinoglu, D., Delledonne, M., Pezzotti, M., Lecharny, A., Scarpelli, C., Artiguenave, F., Pe, M.E., Valle, G., Morgante, M., Caboche, M., Adam-Blondon, A.-F., Weissenbach, J., Quetier, F. and Wincker, P., (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449**, 463–U5.

Jia, Y., Wong, D.C.J., Sweetman, C., Bruning, J.B. and Ford, C.M., (2015) New insights into the evolutionary history of plant sorbitol dehydrogenase. BMC Plant Biology, **15**, 101.

Johnson, T. and Nagel, C., (1976) Composition of central Washington grapes during maturation. American Journal of Enology and Viticulture, **27**, 15–20.

Kliewer, W.M., (1966) Sugars and organic acids of Vitis vinifera. Plant Physiology, **41**, 923.

Kliewer, W.M., (1968) Effect of temperature on composition of grapes grown under field and controlled conditions. Proceedings of the American Society for Horticultural Science, **93**, 797–806.

Kliewer, W.M., (1971) Effect of day temperature and light intensity on concentration of malic and tartaric acids in Vitis-vinifera L grapes. J. Am. Soc. Hortic. Sci. **96**, 372.

Kliewer, W.M., Howarth, L. and Omori, M., (1967) Concentrations of tartaric acid and malic acids and their salts in Vitis vinifera grapes. American Journal of Enology and Viticulture, **18**, 42–54.

Kliewer, W.M. and Lider, L., (1968) Influence of cluster exposure on composition of Thompson Seedless fruit. American Journal of Enology and Viticulture, **19**, 175.

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

Kyte, J. and Doolittle, R.F., (1982) A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology, **157**, 105–132.

Lakso, A.N. and Kliewer, W.M., (1978) The influence of temperature on malic acid metabolism in grape berries. II. responses of dark CO_2 fixation and malic acid pools. American Journal of Enology and Viticulture, **29**, 145–149.

Lander, E.S. and Botstein, D., (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics, **121**, 185–199.

Van Leeuwen, C., Friant, P., Chone, X., Tregoat, O., Koundouras, S. and Dubourdieu, D., (2004) Influence of climate, soil and culivar on Terroir. American Journal of Enology and Viticulture, **55**, 207–217.

Leigh, R. and Wyn Jones, R., (1984) A hypothesis relating critical potassium concentrations for growth to the distribution and functions of this ion in the plant-cell. New Phytologist, **97**, 1–13.

Liu, H.F., Wu, B.H., Fan, P.G., Li, S.H. and Li, L.S., (2006) Sugar and acid concentrations in 98 grape cultivars analyzed by principal component analysis. Journal of the Science of Food and Agriculture, **86**, 1526–1536.

Llorente, M., Villarroya, B. and Gomezcordoves, C., (1991) Reverse-Phase HPLC of organic-acids in musts. Chromatographia, **32**, 555–558.

Loewus, F.A., (1999) Biosynthesis and metabolism of ascorbic acid in plants and of analogs of ascorbic acid in fungi. Phytochemistry, **54**, 447–447.

Loewus, F.A. and Stafford, H.A., (1958) Observations on the incorporation of C^{14} into tartaric acid and the labeling pattern of D-glucose from an excised grape leaf administered L-ascorbic acid-6-C^{14}. Plant Physiology, **33**, 155–156.

Lund, S.T., Peng, F.Y., Nayar, T., Reid, K.E. and Schlosser, J., (2008) Gene expression

analyses in individual grape (Vitis vinifera L.) berries during ripening initiation reveal that pigmentation intensity is a valid indicator of developmental staging within the cluster. Plant Molecular Biology, **68**, 301–315.
Maathuis, F.J.M. and Sanders, D., (1996) Mechanisms of potassium absorption by higher plant roots. Physiologia Plantarum, **96**, 158–168.
Margarido, G.R.A., Souza, A.P. and Garcia, A.A.F., (2007) Onemap. Hereditas **144**, 78–79.
Martinez-Esteso, M.J., Selles-Marchart, S., Lijavetzky, D., Pedreno, M.A. and Bru-Martinez, R., (2011) A DIGE-based quantitative proteomic analysis of grape berry flesh development and ripening reveals key events in sugar and organic acid metabolism. Journal of Experimental Botany, **62**, 2521–2569.
Martinez-Zapater, J.M., Carmona, M.J., Diaz-Riquelme, J., Fernandez, L. and Lijavetzky, D., (2010) Grapevine genetics after the genome sequence: challenges and l imitations. Australian Journal of Grape Wine Research, **16**, 33–46.
Melino, V.J., Hayes, M.A., Soole, K.L. and Ford, C.M., (2011) The role of light in the regulation of ascorbate metabolism during berry development in the cultivated grapevine Vitis vinifera L. Journal of the Science of Food and Agriculture, **91**, 1712–1721.
Melino, V.J., Soole, K.L. and Ford, C.M., (2009) A method for determination of fruit-derived ascorbic, tartaric, oxalic and malic acids, and its application to the study of ascorbic acid catabolism in grapevines. Australian Journal of Grape and Wine Research, **15**, 293–302.
Mpelasoka, B.S., Schachtman, D.R., Treeby, M.T. and Thomas, M.R., (2003) A review of potassium nutrition in grapevines with special emphasis on berry accumulation. Australian Journal of Grape and Wine Research, **9**, 154–168.
Munoz-Robredo, P., Robledo, P., Manriquez, D., Molina, R. and Defilippi, B.G., (2011) Characterization of sugars and organic acids in commercial varieties of table grapes. Chilean Journal of Agricultural Research, **71**, 452–458.
Nosarzewski, M. and Archbold, D.D., (2007) Tissue-specific expression of Sorbitol Dehydrogenase in apple fruit during early development. Journal of Experimental Botany, **58**, 1863–1872.
Novagen, (2011) pET system manual 11th edition.
Ojeda, H., Deloire, A., Carbonneau, A., Ageorges, A. and Romieu, C., (1999) Berry development of grapevines: Relations between the growth of berries and their DNA content indicate cell multiplication and enlargement. Vitis **38**, 145–150.
Oraguzie, N.C. and Wilcox, P.L., (2007) An overview of association mapping, in: Association Mapping in Plants. Springer, New York.
De Orduna, R.M., (2010) Climate change associated effects on grape and wine quality and production. Food Research International, **43**, 1844–1855.
Pagay, V. and Cheng, L.L., (2010) Variability in Berry Maturation of Concord and Cabernet franc in a Cool Climate. Am. J. Enol. Vitic. **61**, 61–67.
Park, S.W., Song, K.J., Kim, M.Y., Hwang, J.H., Shin, Y.U., Kim, W.C. and Chung, W.I., (2002) Molecular cloning and characterization of four cDNAs encoding the isoforms of NAD-dependent sorbitol dehydrogenase from the Fuji apple. Plant Science, **162**, 513–519.
Patrick, J.W., Zhang, W.H., Tyerman, S.D., Offler, C.E. and Walker, N.A., (2001) Role of membrane transport in phloem translocation of assimilates and water. Australian Journal of Plant Physiology, **28**, 695–707.
Pavlousek, P. and Kumsta, M., (2011) Profiling of Primary Metabolites in Grapes of Interspecific Grapevine Varieties: Sugars and Organic Acids. Czech Journal of Food Science, **29**, 361–372.

Payne, R.., Murray, D.., Harding, S.., Baird, D.. and Soutar, D.., (2013) GenStat for windows (15th Edition) introduction. VSN International, Hemel Hempstead.
Pereira, G.E., Gaudillere, J.P., Pieri, P., Hilbert, G., Maucourt, M., Deborde, C., Moing, A. and Rolin, D., (2006) Microclimate influence on mineral and metabolic profiles of grape berries. Journal of Agriculture and Food Chemistry, **54**, 6765–6775.
Pilati, S., Perazzolli, M., Malossini, A., Cestaro, A., Dematte, L., Fontana, P., Dal Ri, A., Viola, R., Velasco, R. and Moser, C., (2007) Genome-wide transcriptional analysis of grapevine berry ripening reveals a set of genes similarly modulated during three seasons and the occurrence of an oxidative burst at veraison. BMC Genomics, **8**.
Pollard, K.S., Dudoit, S. and van der Laan, M.J., (2005) Multiple Testing Procedures: R multtest Package and Applications to Genomics, in Bioinformatics and Computational Biology Solutions Using R and Bioconductor., Statistics for biology and health. Springer.
Preiner, D., Tupajic, P., Kontic, J.K., Andabaka, Z., Markovic, Z. and Maletic, E., (2013) Organic acids profiles of the most important Dalmatian native grapevine (V. vinifera L.) cultivars. Journal of Food Composition and Analysis, **32**, 162–168.
Rezaian, M.A. and Krake, L.R., (1987) Nucleic acid extraction and virus detection in grapevine. Journal of Virological Methods, **17**, 277–285.
Rienth, M., Torregrosa, L., Kelly, M.T., Luchaire, N., Pellegrino, A., Grimplet, J. and Romieu, C., (2014) Is transcriptomic regulation of berry development more important at night than during the day? PLoS One, **9**, e88844.
Rienth, M., Torregrosa, L., Luchaire, N., Chatbanyong, R., Lecourieux, D., Kelly, M.T. and Romieu, C., (2014) Day and night heat stress trigger different transcriptomic responses in green and ripening grapevine (vitis vinifera) fruit. BMC Plant Biology, **14**, 18.
Ronaghi, M., (2001) Pyrosequencing sheds light on DNA sequencing. Genome Research, **11**, 3–11.
Ruffner, H.P., (1982) Metabolism of tartaric and malic acids in Vitis: a review part A. Vitis **21**, 247–259.
Ruffner, H.P. and Hawker, J.S., (1977) Control of glycolysis in ripening berries of Vitis vinifera. Phytochemistry, **16**, 1171–1175.
Ruffner, H.P., Hawker, J.S. and Hale, C.R., (1976) Temperature and enzymic control of malate metabolism in berries of Vitis vinifera. Phytochemistry, **15**, 1877–1880.
Sabir, A., Kafkas, E. and Tangolar, S., (2010) Distribution of major sugars, acids and total phenols in juice of five grapevine (Vitis spp.) cultivars at different stages of berry development. Spanish Journal of Agricultural Research,. **8**, 425–433.
Sadras, V.O., Moran, M.A. and Bonada, M., (2013) Effects of elevated temperature in grapevine. I Berry sensory traits. Australian Journal of Grape and Wine Research, **19**, 95–106.
Sadras, V.O., Petrie, P.R. and Moran, M.A., (2013) Effects of elevated temperature in grapevine. II juice pH, titratable acidity and wine sensory attributes. Australian Journal of Grape and Wine Research, **19**, 107–115.
Saito, K., (1994) Specific transfer of 3H from D-[3-3H] Gluconic acid into L-tartaric acid in vitaceous plants. Phytochemistry, **37**, 1017–1022.
Saito, K. and Kasai, Z., (1969) Tartaric acid synthesis from L-Ascorbic acid-1-14C in grape berries. Phytochemistry, **8**, 2177–2182.
Saito, K. and Kasai, Z., (1982) Conversion of L-ascorbic acid to L-idonic acid, L-idono-gamma-lactone and 2-keto-L-Idonic acid in slices of immature grapes. Plant Cell Physiology, **23**, 499–507.
Saito, K. and Kasai, Z., (1984) Synthesis of L-(+)-tartaric acid from L-ascorbic acid via 5-

keto-D-gluconic acid in grapes. Plant Physiology, **76**, 170–174.
Saito, K. and Loewus, F.A., (1989) Occurrence of L(+)-tartaric acid and its formation from D-gluconate or D-xylo-5-hexulosonate in bean leaf (Phaseolus-vulgaris L. Plant Science, **62**, 175–180.
Sato, A., Yamada, M., Iwanami, H. and Hirakawa, N., (2000) Optimal spatial and temporal measurement repetition for reducing environmental variation of berry traits in grape breeding. Scientia Horticulturae, **85**, 75–83.
Sauvage, C., Segura, V., Bauchet, G., Stevens, R., Do, P.T., Nikoloski, Z., Fernie, A.R. and Causse, M., (2014) Genome-Wide Association in Tomato Reveals 44 Candidate Loci for Fruit Metabolic Traits. Plant Physiology, **165**, 1120–1132.
Schaart, J.G., Mehli, L. and Schouten, H.J., (2005) Quantification of allele-specific expression of a gene encoding strawberry polygalacturonase-inhibiting protein (PGIP) using Pyrosequencing((TM)). Plant Journal, **41**, 493–500.
Schneider, A., Gerbi, V. and Redoglia, M., (1987) A rapid HPLC method for separation and determination of major organic-acids in grape musts and wines. American Journal of Enology and Viticulture, **38**, 151–155.
Soyer, Y., Koca, N. and Karadeniz, F., (2003) Organic acid profile of Turkish white grapes and grape juices. Journal of Food Composition and Analysis, **16**, 629–636.
Stafford, H.A., (1959) Distribution of tartaric acid in the leaves of certain angiosperms. American Journal of Botany, **46**, 347–352.
Sun, A.J., Ge, J.B., Siffert, W. and Frey, U.H., (2005) Quantification of allele-specific G-protein beta 3 subunit mRNA transcripts in different human cells and tissues by pyrosequencing. Eur. J. Hum. Genetics, **13**, 361–369.
Swanepoel, M., du Toit, M. and Nieuwoudt, H.H., (2007) Optimisation of the quantification of total soluble solids, pH and titratable acidity in South African grape must using Fourier transform mid-infrared spectroscopy. South African Journal of Enology and Viticulture, **28**, 140–149.
Sweetman, C., Deluc, L., Cramer, G.R., Ford, C.M. and Soole, K.L., (2009) Regulation of malate metabolism in grape berry and other developing fruits. Phytochemistry, **70**, 1329–1344.
Sweetman, C., Sadras, V.O., Hancock, R.D., Soole, K.L. and Ford, C.M., (2014) Metabolic effects of elevated temperature on organic acid degradation in ripening Vitis vinifera fruit. Journal of Experimental Botany, **65**, 5975–5988.
Sweetman, C., Wong, D.C.J., Ford, C.M. and Drew, D.P., (2012) Transcriptome analysis at four developmental stages of grape berry (Vitis vinifera cv. Shiraz) provides insights into regulated and coordinated gene expression. BMC Genomics, **13**, 25.
Turner, S.D., (2014) qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv.
Vargas, A.M., Le Cunff, L., This, P., Ibanez, J. and Teresa de Andres, M., (2013) VvGAI1 polymorphisms associate with variation for berry traits in grapevine. Euphytica, **191**, 85–98.
Velasco, R., Zharkikh, A., Troggio, M., Cartwright, D.A., Cestaro, A., Pruss, D., Pindo, M., FitzGerald, L.M., Vezzulli, S., Reid, J., Malacarne, G., Iliev, D., Coppola, G., Wardell, B., Micheletti, D., Macalma, T., Facci, M., Mitchell, J.T., Perazzolli, M., Eldredge, G., Gatto, P., Oyzerski, R., Moretto, M., Gutin, N., Stefanini, M., Chen, Y., Segala, C., Davenport, C., Dematte, L., Mraz, A., Battilana, J., Stormo, K., Costa, F., Tao, Q., Si-Ammour, A., Harkins, T., Lackey, A., Perbost, C., Taillon, B., Stella, A., Solovyev, V., Fawcett, J.A., Sterck, L., Vandepoele, K., Grando, S.M., Toppo, S., Moser, C., Lanchbury, J., Bogden, R., Skolnick, M., Sgaramella, V., Bhatnagar, S.K., Fontana, P., Gutin, A., Van de Peer, Y., Salamini, F. and Viola, R.,

(2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. Plos One, **2**.

Vouillamoz, J.F., McGovern, P.E., Ergul, A., Soylemezoglu, G., Tevzadze, G., Meredith, C.P. and Grando, M.S., (2006) Genetic characterization and relationships of traditional grape cultivars from Transcaucasia and Anatolia. Plant Genetic Resources Characterization and Utilization, **4**, 144–158.

Webb, M., Cavaletto, J., Carpita, N., Lopez, L. and Arnott, H., (1995) The intravacuolar organic matrix associated with calcium-oxalate crystals in the leaves of Vitis. Plant Journal, **7**, 633–648.

Wen, Y.-Q., Cui, J., Zhang, Y., Duan, C.-Q. and Pan, Q.-H., (2014) Comparison of organic acid levels and L-IdnDH expression in Chinese-type and European-type grapes. Euphytica, **196**, 63–76.

Wen, Y.-Q., Li, J.-M., Zhang, Z.-Z., Zhang, Y.-F. and Pan, Q.-H., (2010) Antibody Preparation, Gene Expression and Subcellular Localization of L-Idonate Dehydrogenase in Grape Berry. Bioscience, Biotechnology and Biochemistry, **74**, 2413–2417.

Williams, L. and Biscay, P., (1991) Partitioning of dry-weight, nitrogen, and potassium in Cabernet-Sauvignon grapevines from anthesis until harvest. American Journal of Enology and Viticuture, **42**, 113–117.

Williams, M. and Loewus, F.A., (1978) Synthesis of (+)-Tartaric acid from L-[4-14C]Ascorbic acid in grape and geranium. Plant Physiology, **61**, 672–674.

Wolpert, J.A. and Vilas, E.P., (1992) Estimating vineyard yields: introduction to a simple, two-step method. American Journal of Enology and Viticulture, **43**, 384–388.

Wu, B.-H., Li, S.-H., Nosarzewski, M. and Archbold, D.D., (2010) Sorbitol Dehydrogenase gene expression and enzyme activity in apple: tissue specificity during bud development and response to rootstock vigor and growth manipulation. Journal of the American Society of Horticultural Science, **135**, 379–387.

Wu, R., Ma, C.-X., Painter, I. and Zeng, Z.-B., (2002) Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. Theoretical population biology. **61**, 349–363.

Zacinas, B., Cartwright, B. and Spouncer, L., (1987) Nitric-acid digestion and multielement analysis of plant-material by inductively coupled plasma spectrometry. Communications in Soil Science and Plant Analysis, **18**, 131–146.

Zelleke, A. and Kliewer, W.M., (1985) Influence of root temperature, root stock and fertilization on the composition of Cabernet Sauvignon grapevines. Acta Horticulturae, 255–263.

Zheng, Y.J., Duan, Y.T., Zhang, Y.F., Pan, Q.H., Li, J.M. and Huang, W.D., (2009) Determination of organic acids in red wine and must on only one RP-LC-column directly after sample dilution and filtration. Chromatographia, **69**, 1391–1395.

Journal, **7**, 633–648.

Wen, Y.-Q., Cui, J., Zhang, Y., Duan, C.-Q. and Pan, Q.-H., (2014) Comparison of organic acid levels and L-IdnDH expression in Chinese-type and European-type grapes. Euphytica, **196**, 63–76.

Wen, Y.-Q., Li, J.-M., Zhang, Z.-Z., Zhang, Y.-F. and Pan, Q.-H., (2010) Antibody Preparation, Gene Expression and Subcellular Localization of L-Idonate Dehydrogenase in Grape Berry. Bioscience, Biotechnology and Biochemistry, **74**, 2413–2417.

Williams, L. and Biscay, P., (1991) Partitioning of dry-weight, nitrogen, and potassium in Cabernet-Sauvignon grapevines from anthesis until harvest. American Journal of Enology and Viticuture, **42**, 113–117.

Williams, M. and Loewus, F.A., (1978) Synthesis of (+)-Tartaric acid from L-[4-14C]Ascorbic acid in grape and geranium. Plant Physiology, **61**, 672–674.

Wolpert, J.A. and Vilas, E.P., (1992) Estimating vineyard yields: introduction to a simple, two-step method. American Journal of Enology and Viticulture, **43**, 384–388.

Wu, B.-H., Li, S.-H., Nosarzewski, M. and Archbold, D.D., (2010) Sorbitol Dehydrogenase gene expression and enzyme activity in apple: tissue specificity during bud development and response to rootstock vigor and growth manipulation. Journal of the American Society of Horticultural Science, **135**, 379–387.

Wu, R., Ma, C.-X., Painter, I. and Zeng, Z.-B., (2002) Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. Theoretical population biology. **61**, 349–363.

Zacinas, B., Cartwright, B. and Spouncer, L., (1987) Nitric-acid digestion and multielement analysis of plant-material by inductively coupled plasma spectrometry. Communications in Soil Science and Plant Analysis, **18**, 131–146.

Zelleke, A. and Kliewer, W.M., (1985) Influence of root temperature, root stock and fertilization on the composition of Cabernet Sauvignon grapevines. Acta Horticulturae, 255–263.

Zheng, Y.J., Duan, Y.T., Zhang, Y.F., Pan, Q.H., Li, J.M. and Huang, W.D., (2009) Determination of organic acids in red wine and must on only one RP-LC-column directly after sample dilution and filtration. Chromatographia, **69**, 1391–1395.