A Multiple 'Omics' Approach to Study the Interaction between the Vitis Vinifera Transcriptome and Epigenome and the Barossa Valley Terroir

Pastor Jullian Fabres

A thesis submitted to The University of Adelaide in fulfilment of the requirements

for the degree of Doctor of Philosophy

Faculty of Sciences

School of Agriculture, Food and Wine

The University of Adelaide

THE UNIVERSITY of ADELAIDE

January 2020

Table of contents

Abstract 6

Declaration 8

Acknowledgement 9

Abbreviations 10

Publications arising from this thesis 12

Chapter 1: Introduction and literature review 14

Introduction 15

The Australian wine industry 15

Grapevine and its growth environment 20

Decoding *terroir* 26

Gene expression analyses in *V. vinifera* 26

Epigenetics in plants 29

Developmental and environmental epigenetics 30

Grapevine epigenetics 32

Mobile transcripts in grapevines 35

Research Aims 38

Chapter 2: A concise review on multi-omics data integration for *terroir* analysis in *Vitis vinifera* 39

Link between Chapters 1 and 2 41

Chapter 3: Planting year and local growing environment modify DNA methylation and gene expression of grapevine (*Vitis vinifer*a cv. Shiraz) 50

Abstract 54

Background 55

Results 58

Sub-region environmental differences 58

Genetic structure of Shiraz samples 58

Effect of provenance on DNA methylation 59

Effect of provenance on gene expression 60

Relationship between DNA methylation and gene expression 61

Identification of environmental variables that affect DNA methylation and gene expression 62

Discussion 63

Regional effect on DNA methylation and gene expression 63

Unraveling the Barossa environment and its consequences for DNA methylation and gene expression 68

Conclusions 70

Methods 70

Plant material 70

Methylation Sensitive Genotyping by Sequencing 71

msGBS data analysis 72

RNA isolation and sequencing 73

RNA-seq data analysis 74

Detection of environmental effect in gene expression and DNA methylation 74

List of abbreviations 75

Acknowledgements 76

Availability of data and materials 76

Table 77

Figure Legends 78

Figures 81

References 125

Chapter 4: Dicistronic tRNA-mRNA transcripts in grapevine (*Vitis vinifera*) display distinct, regional expression patterns that correlate with tRNA expression 133

Abstract 138

Introduction 139

Results 141

RNA-sequencing of *Vitis Vinifera* cv. Shiraz 141

Identification of putative dicistronic tRNA-mRNA transcripts 142

Characteristics of grapevine dicistronic tRNA-mRNA candidates 143

Regional patterns of dicistronic expression 145

Discussion 146

Material and methods 151
Sampling material 151
RNA extraction and RNA-seq library preparation 151
In silico detection of dicistronic tRNA-mRNA transcripts 152
RT-PCR confirmation 154
Regional effect on the expression of dicistronic tRNA:mRNAs 155
Gene analyses 156
Motif analysis 156
Acknowledgments 156
Table 158
Figure Legends 159
Figures 162
References 188
Chapter 5: General discussion 193
Appendices 200
References 205

Abstract

The local growing conditions have a significant impact in grapevine (*Vitis vinifera*) phenotype of as well as in its epigenome. Little is known about how grapevine responds to the local environment through epigenetic modifications and any effect on gene expression in field-grown plants. Here, we characterized the leaf methylome and transcriptome of 198 Shiraz vines cultivated across twenty-two vineyards, covering six sub-regions of the Barossa Wine Zone in Australia. Samples clustered for both DNA methylation and gene expression depending on the sub-region of their origin. In addition, we identified genes with DNA methylation and expression significantly associated to measured environmental variables. Planting year and annual rainfall were the variables showing a higher influence both on DNA methylation and gene expression.

Plants use long-distance signaling mechanisms to coordinate developmental and environmental cues between organs and tissues. Studies have begun to unravel the molecular nature of systemic signaling and RNA transcripts have emerged as key players by acting as mobile messengers. Along with DNA methylation and chromatin modifications, non-coding RNAs are a molecular mechanism for the epigenetic regulation of gene expression. Recent evidence in model plants suggests that dicistronic tRNA-like structures can act as mobile signals for mRNA transcripts to move between distant tissues, potentially acting as signals for development or changes in environmental conditions. However, it is not clear the extent to which dicistronic transcripts of tRNA and mRNAs are expressed in field-grown plants, or the factors that contribute to their expression. We identified and analysed the expression of tRNA-mRNA dicistronic transcripts in leaves and berries from the twenty-two vineyards, using a novel pipeline

developed to identify dicistronic transcripts from high-throughput short-read RNA sequencing data. Of the 124 tRNA genes that were expressed in both tissues, we detected 18 tRNA genes forming part of 19 dicistronic tRNA-mRNA molecules. We found that the presence and abundance of dicistronic molecules was tissue specific, and that each the sub-region displayed a distinct expression profile for dicistronic tRNA-mRNA transcripts. In this work we propose that DNA methylation can reflect the local growing conditions in grapevine, as a mechanism of response to its environment. We also detected the presence of dicistronic tRNA-mRNAs, which might act as signaling molecules between distant tissues. Both responses to the growing environment are especially relevant in long-lived woody plants, potentially giving them the necessary plasticity to adapt to local conditions. As grapevine is also an economically important fruit crop, understanding its response to the environment could give us a better inside of the possible mechanisms behind the expression of terroir in grapevine.

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 any 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 acknowledge 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 been granted by the University to restrict access for a period of time.

Signature Date

Acknowledgement

First, I would like to express my gratitude to my supervisors Dr. Penny Tricker, Dr. Carlos Rodriguez Lopez and Dr. Rakesh David for their excellent guidance, constant support and understanding. I would like also to acknowledge Dr. Cassandra Collins for her support at the beginning of my project.

I acknowledge the University of Adelaide and Wine Australia for offering me this study opportunity and financial support.

I thank Dr James Breen, Dr Steve Pederson and Dr Nathan Watson-Haigh for their support and guidance in bioinformatics.

I would like to extend my gratitude to all the people from the Plant Research Centre, particularly to Dr. Roberta De Bei, Dr Dylan Griggs, Wendy Sullivan, Na Sai and Konstantinos Bogias. I also would like to thank to the people from the Plant Genomics Centre, particularly Priyanka Kalambettu, Anzu Okada, Jessica Schmidt, Caterina Selva, and Dr Mateo Riboni. I especially thank the people from the University of Kentucky, Kiflu Tesfamicael and Lakshay Anand.

At last, I must thank my parents, and my brothers and sisters that always believed in me. I thank my father and mother in laws for their constant support and help. I am especially grateful of my wife, Thais Munoz, she gave me the strength to start and finish this journey. To my kids, Oliver and Matias, without them this whole project would be pointless.

Abbreviations

*C	Degree Celsius
5-AzaC	5-AzacCytidine
5mC	5-methylcytosine
ABA	Abscisic acid
ATP	Adenosine triphosphate
bp	Base pairs
CG	Central Grounds
CMT	Chromomethylase
Cnr	Colourless non-ripening
CPM	counts per million
CTP:	Cytidine triphosphate
DAPC	Descriptive analysis of principal components
DEGs	Differentially expressed genes
DMEGs	Differentially methylated and expressed genes
DMGs	Differentially methylated genes
DMMs	Differentially methylated markers
DNA	Deoxyribonucleic acid
DRM	Domain rearranged methyltransferase
EE	Eastern Edge
EV	Eden Valley
Gb	Giga base pairs
GFLV	Grapevine fanleaf virus
GTP	Guanosine triphosphate
GxE	Genotype by environment
ha.	Hectare
LDA	Linear discriminant analysis
m a.s.l.	Meters above the sea level
MADS-box	MCM1, Agamous, Deficiens and SRF box
Mb	Mega base pairs
MET1	Methyltransferase 1
miRNAs	MicroRNAs
ML/ha	Megaliter per hectare
mm	Millimeter
mRNA	Messenger RNA
MSAP	Methylation-sensitive amplified polymorphism
msGBS	Methylation sensitive genotyping by sequencing
NG	Northern Grounds
ORFs	Open reading frames

PC	Principal component
PCA	Principal components analysis
RIN	Ripening inhibitor
RNA	Ribonucleic acid
SBP-box	Squamosa promoter binding protein
SDG	SET DOMAIN GROUP
SG	Southern Grounds
siRNAs	Small interfering RNAs
SNPs	Single nucleotide polymorphisms
SPL	Squamosa promoter-binding protein-like
STS	Stilbene synthase
TE	Transposable elements
TLSs	tRNA-like structure
tRNA	Transfer RNA
UTP	Uridine triphosphate
UTR	Untranslated region
UV	Ultraviolet
WR	Western Ridge.

Publications arising from this thesis

1. Jullian Fabres P, Collins C, Cavagnaro TR, Rodríguez López CM. A concise review on multi-omics data integration for *terroir* analysis in *Vitis vinifera*. Front Plant Sci. 2017 Jun 20; 8:1065. Published.
2. Jullian Fabres P, Sai N, Pederson S, Zheng F, Breen J, Gilliham M, Tricker P, Rodríguez López C, David R. Dicistronic tRNA-mRNA transcripts in grapevine (*Vitis vinifera*) display distinct, regional expression patterns that correlate with tRNA expression. Submitted to RNA Journal.
3. Jullian Fabres P, Xie H, Sai N, Bogias K, David R, De Bei R, Gilliham M, Breen J, Tricker P and Rodríguez López C. Planting year and local growing environment modify DNA methylation and gene expression of grapevine (*Vitis vinifera* cv. Shiraz). Manuscript in preparation.

About this thesis

My thesis is presented as a series of five chapters that include three manuscripts. One manuscript has been published and two manuscripts are drafted for publication in scientific journals.

Chapter 1 is a general introduction that provides a background for the work presented in the thesis and establishes the research objectives. Chapter 2 is a published review about study of *terroir* in grapevine through the integration of multi-omics data. Chapters 3 and 4 are experimental papers, each with a statement of authorship. As each chapter is drafted as a separate journal article, there is inevitable repetition between chapters, but it has been kept to the minimum possible.

In Chapter 3, grapevine leaf samples from different growing regions were analysed to identify the effects of the plant age, environment and soil composition on gene expression and DNA methylation. The most significant variables were identified and described.

In Chapter 4, the expression of dicistronic transcripts that have the potential to act as systemic signaling messengers in grapevine is investigated. Dicistronic tRNA-mRNA transcripts showing regional specificity were identified from leaf and berry samples. We hypothesise that dicistronic tRNA-mRNA could have a functional role in grapevine transcriptomic adaptation to local environments.

Chapter 5 is a general discussion that covers the broader significance of the research conducted in this PhD project. Main conclusions are highlighted, and future research directions are also suggested.

Chapter 1: Introduction and literature review

Introduction

The Australian wine industry

In Australia, the wine industry is estimated to contribute more than $40 billion to the Australian economy each year, with an average yearly growth rate of 3% since 2015 (Wine Australia 2019). Much of this growth is due to an increased demand for premium Australian wines in overseas markets. The higher demand for wine has also increased the grape price and the revenue for grape-growing from $125/ha. in 2015 to $1,394/ha. in 2018. One of the most important wine zones in Australia is the Barossa in South Australia. The first Barossa vineyard dates from as early as 1842 and since then the industry has grown in the region with currently more than 170 wine companies and 550 grape growers (Barossa Grape and Wine Association). Within the Barossa Wine Zone there are two main regions: the Barossa Valley Region and the Eden Valley Region (Figure 1). The Barossa Valley Region is lower than 400 metres above sea level, whereas the Eden Valley Region situated in the hills to the east rises above 400 metres. This difference in altitude between regions has resulted in differences in climatic growing conditions (Figures 2, 3 and 4). The grape growers and the wine industry have characterized the Barossa Valley as a warmer region with lower rainfall that produces rich wines with depth of flavour. The Eden Valley is characterized as a cooler climate region that produce wines of greater elegance and acidity (Robinson and Sandercock 2014). This difference in climatic conditions has allowed growers to cultivate grapevines adapted for each regional environment; Shiraz from the Barossa Valley and Riesling from the Eden Valley have become the preferred varieties in each region. Shiraz is the most cultivated grape variety in Australia and, in the Barossa Wine Zone, old pre-phylloxera Shiraz vines can be found that are characterized by low yield but dark, rich and concentrated berries used for making rare wines (Barossa Grape and Wine Association). The Barossa Valley can be divided into different sub-regions depending on their altitude, climate and

soil types. These sub-regions are the Northern Grounds, Central Grounds and Southern Grounds (Figure 1) (Barossa Grape and Wine Association).

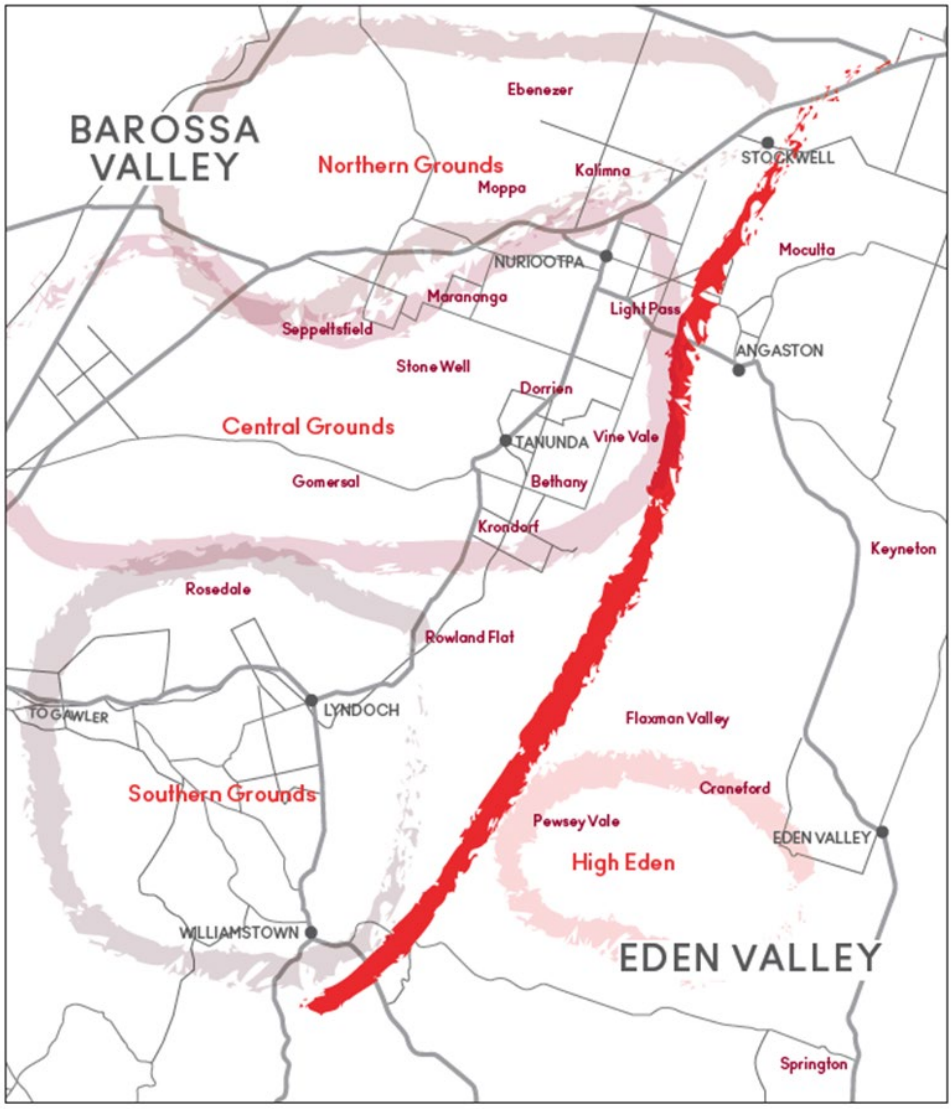
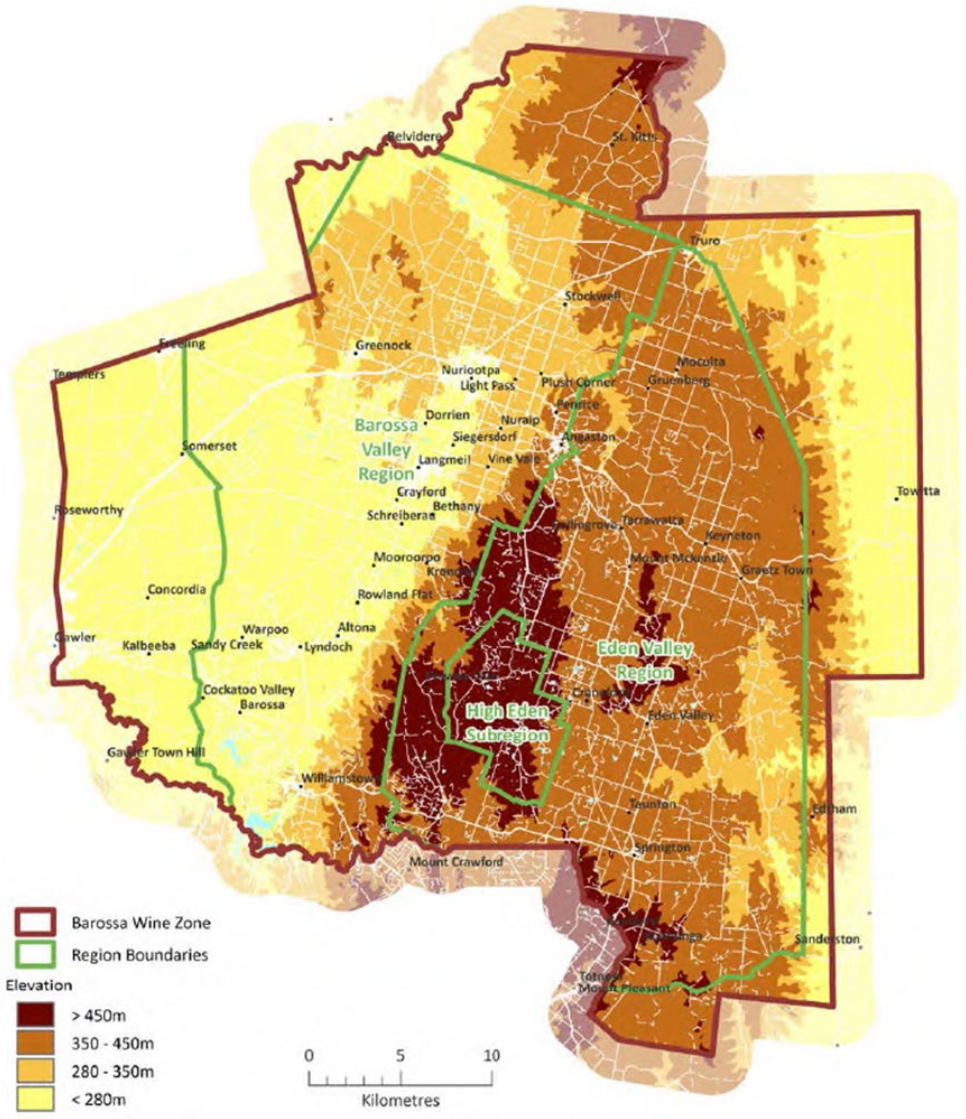


Figure 1. Map of Barossa Wine Zone. The Barossa Wine Zone can be divided into two regions, Barossa Valley Region, composed by the sub-regions Northern, Central and Southern Grounds, and the Eden Valley Region. Extracted from Barossa Grape and Wine Association, Barossa Chapters: Grounds, 2014 (https://www.barossawine.com/barossa/barossa-chapters/barossa-chapters-barossa-grounds/).


Belvidere
Truro
Stockwell
Greenock
Nuriootpa
Light Pass
Plush Corner
Moculta
Gruenberg
Penrice
Dorrien
Nuraip
Siegersdorf
Angaston
Barossa Valley Region
Somerset
Langmeil
Vine Vale
Crayford
Bethany
Schreiberaa
Towitta
Roseworthy
Tarrawatta
Keyneton
Mooroorpo
Krondorf
Mount McKenzie
Greatz Town
Concordia
Rowland Flat
Warpoo
Altona
Crawler
Kalbeeba
Sandy Creek
Lyndoch
Eden Valley Region
Cockatoo Valley
Barossa
High Eden Subregion
Eden Valley
Gawler Town Hill
Williamstown
Tanunda
Tanunda
Tanunda

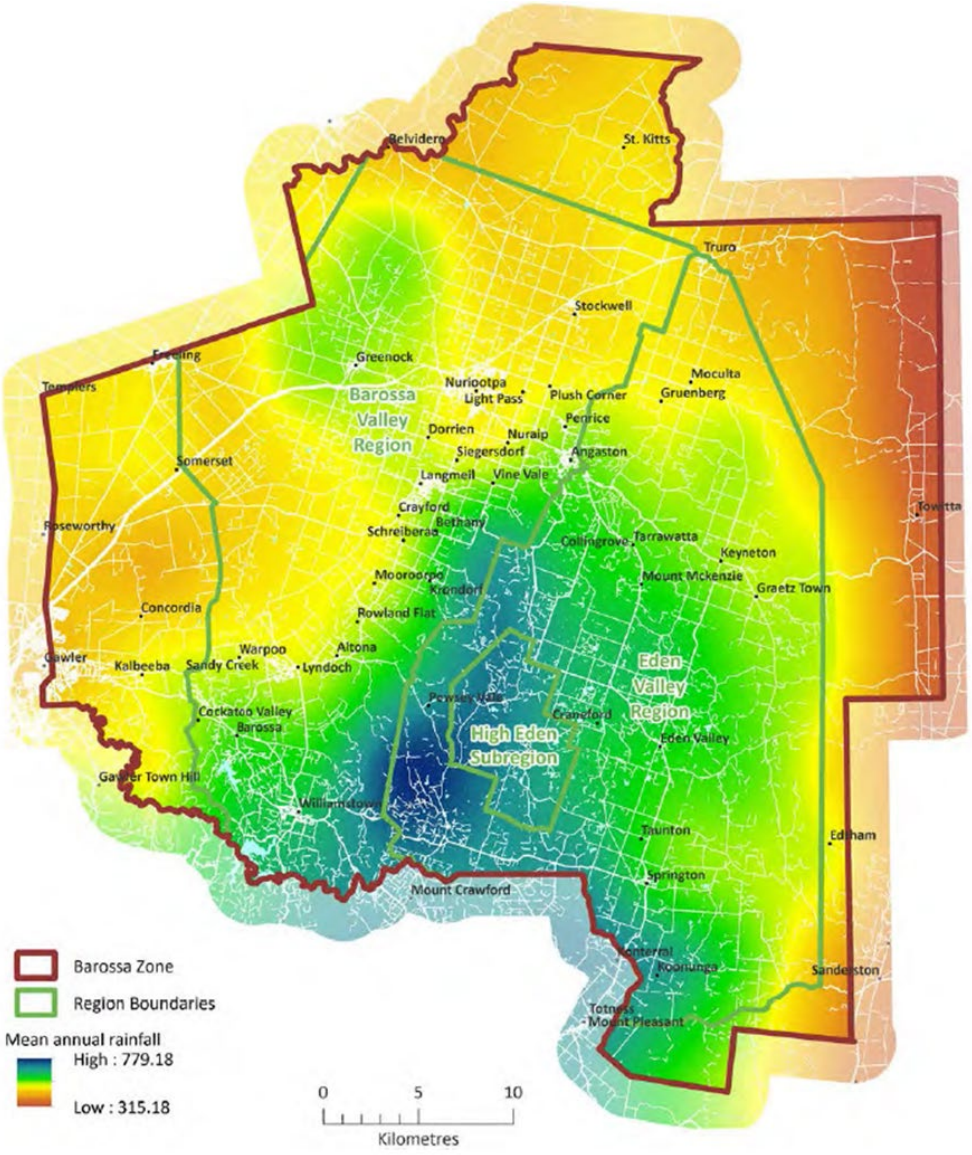


Figure 3. Mean annual rainfall in of the Barossa Wine Zone. "Mean values are calculated over at least 30 years. The data have been transformed to exactly five kilometre cell size and interpolated to produce estimates at a much finer resolution. The mean annual rainfall was divided into five classes across the following ranges; High; above 650 mm, Moderate: between 650 and 550 mm, Moderately Low: between 550 and 450 mm, Low: between 450 and 380 mm and Very Low: below 380 mm." Extracted from Robinson S. and Sandercock N., Primary Industries and Regions South Australia, 2014.

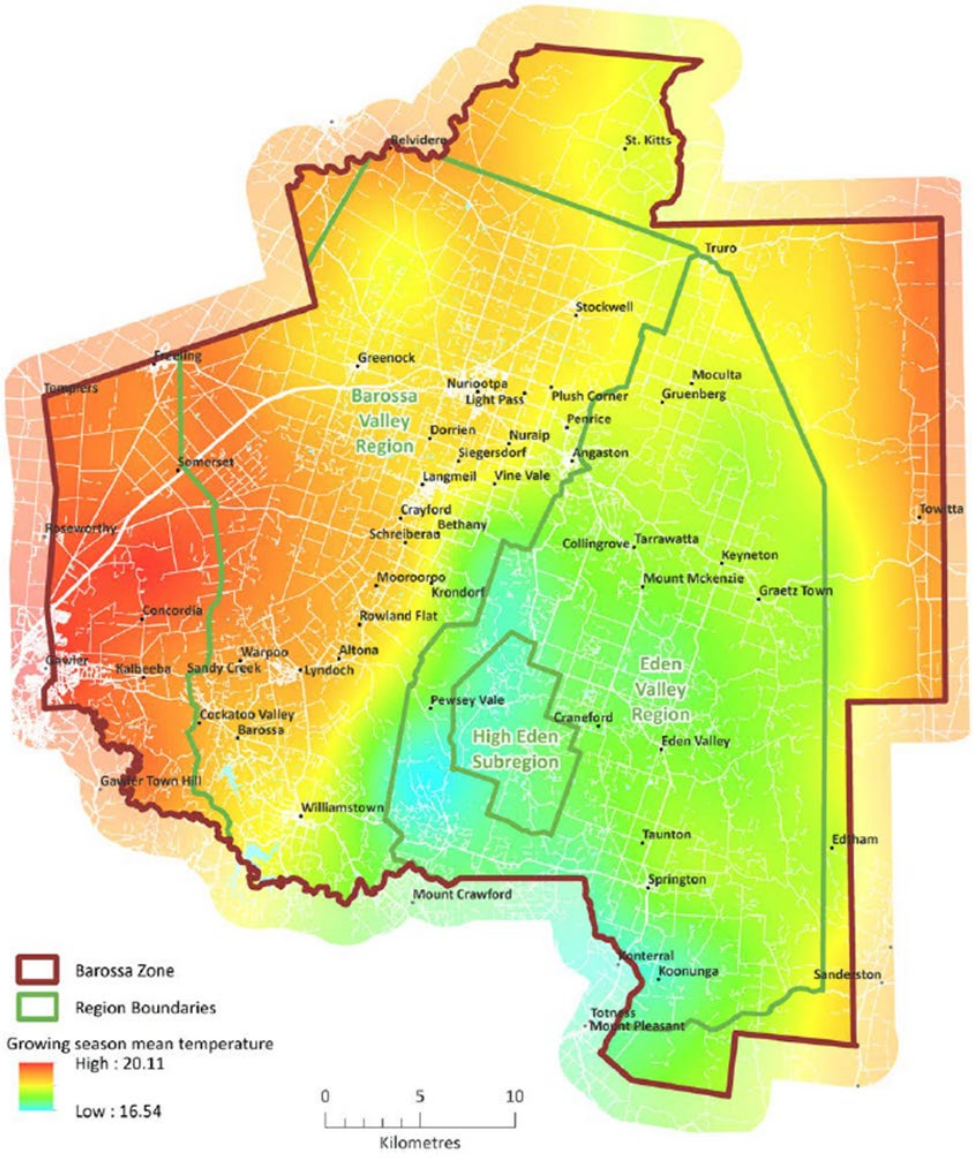


Figure 4. Mean temperature during growing season. "The growing season is regarded as 1 October through 30 April. Values are calculated by the mean of 0.5 x (daily maximum + daily minimum) over at least 30 years. The data have been transformed to exactly five kilometre cell size and interpolated to produce estimates at a much finer resolution." Extracted from Robinson S. and Sandercock N., Primary Industries and Regions South Australia, 2014.

Grapevine and its growth environment

Grapes (*Vitis* spp.) are one of the most economically important fruit crops and a large proportion of total production is used for wine production (Conde et al. 2007; Food Agriculture Organization 2012). The genus *Vitis* includes more than 70 species, but almost all commercial wines are made from *Vitis vinifera* berries due to their chemical composition that makes them ideal for wine production. Berries from *V. vinifera* accumulate high levels of sugars, organic acids and polyphenols - chemical compounds that give the individual colour, aroma and flavour to wine (Anesi et al. 2015; Conde et al. 2007).

The composition of grape berries is strongly influenced by the local growing conditions and this has generated a great interest in understanding all the factors that affect berry composition during ripening. The term *terroir* comes from the French word *terre*, meaning soil, and it is used to relate sensory wine attributes, like quality, taste and style, to the geographic origin in which the grapes were grown (van Leeuwen and Seguin 2006). *Terroir* can be defined as a product of the interaction between the climate, soil, human practices and cultivar (Gladstones 2011a). The effects of *terroir* on grape quality have been extensively studied and below the main known drivers of *terroir* are described.

Climate

Climatic conditions affect most of the key life cycle events in grapevine and have a strong influence on the chemical composition of the berries (Jones and Davis 2000; van Leeuwen and Seguin 2006; Zerihun et al. 2015). Temperature, radiation and rainfall have been identified as the most important climatic factors in fruit and wine quality (Romero et al. 2016).

Temperature

Temperature is one of the most important factors in viticulture. Grapevine phenology is tightly regulated by temperature and, along with light, temperature determines photosynthesis, dry matter production and potential yield (Gladstones 2011a). Although grapevines can survive at temperatures as low as -15˚C in winter, optimal growing conditions for grapevine range between 22 ºC to 25 ºC (Gladstones 2011a). Premium wines are usually made between the latitudes 35 to 50, both in the northern and southern hemispheres (van Leeuwen and Seguin 2006). Although grapevine can be grown in tropical areas, in these climatic conditions there is an increase in grapevine vigour affecting the quality of the fruit for winemaking purposes (van Leeuwen and Seguin 2006). At warmer temperatures, there is a negative effect during berry ripening, caused by reduction in berry weight (Greer and Weston 2010), reduction in total soluble solids (Carbonell-Bejerano et al. 2013) and reduction of flavonol content (Azuma et al. 2012). At high temperatures (30-35 °C) the total anthocyanin content is also reduced to less than half that in berries at control conditions (25 °C) (Mori et al. 2007). In contrast, berries from vines grown at low temperatures presented an increase in total soluble solids and anthocyanin content (Mori et al. 2005). In addition, cool day temperatures and warm night temperature increased berry weight and sugar content, and caused changes in flavonoid composition (Cohen et al. 2012).

Light

Sunlight and temperature have a similar effect on grapevine. Yield and soluble solids will increase with more radiation, either more intensity or duration of the radiation, due to an increase in the photosynthetic rate (Gladstones 2011a). In addition, during high radiation, temperature will also rise, increasing metabolic activity and photosynthesis (Bergqvist et al. 2001; Jackson and Lombard 1993). However, temperate plants like grapevines, only need a third (700 $E^{m-2\ s-1}$) of bright sunshine intensity (2500 $E^{m-2\ s-1}$) for maximum rates of photosynthesis (Jackson and

Lombard 1993). Photosynthesis can be inhibited when the sunlight intensity is too high, especially if it is accompanied by high temperatures (Iacono and Sommer 1996).

Water

The effects that rainfall has on berry composition largely depend on the phenological stage of the grapevine. Although plenty of water is required at the beginning of the growing season, moderate water stress by the end of the season is ideal, and regions that produce good quality wines are usually in areas where the annual precipitation is below 800 mm (Jackson and Lombard 1993). Rain after veraison can have serious effects on grapes, causing berry splitting and increasing the susceptibility to fungal diseases like botrytis (Jackson and Lombard 1993). Moreover, evidence suggests that mild water stress close to harvest increases the accumulation of precursors of volatile thiols that determine the grape aroma potential in white grapevine varieties (des Gachons et al. 2005), and the phenolic content in red grapevine varieties (Koundouras et al. 2006). A study that analysed 45 years of climate records and vintage quality from vineyards in Bordeaux, France, identified a positive correlation between berry and wine quality with wet winters followed by lower precipitation during flowering and harvest (Jones and Davis 2000). Similar results were obtained by Romero et al. (2016) in Murcia, Spain, corroborating this trend between rainfall and wine quality.

Soil

High quality wines are produced on a wide variety of soils, so there is not one definitive soil that will ensure great wines. Water-holding capacity, nutrient availability, depth and drainage are the most important soil factors to consider (Rankine et al. 1971). In addition, the structure of the soil will affect root growth depending on its penetrability, and may also affect the microclimate of the grapevine due to its heat-retaining and light-reflecting capacity (Jackson and Lombard 1993). Good soils are those that limit yield and vine vigour either by limiting the water supply or the

mineral uptake (van Leeuwen and Seguin 2006). Nitrogen strongly increases vine vigour, but negatively affects berry quality (van Leeuwen 2010). A study showed that when nitrogen uptake was limited, vine vigour, berry weight and the yield obtained were decreased, but at the same time the sugar accumulation, anthocyanin and tannin concentrations measured were increased, characteristics considered beneficial for producing high quality wines (Chone et al. 2001).

The use of grafting is a common practice in most viticultural areas. The main reason for grafting in grapevines is to gain immunity against phylloxera, a sap-sucking insect that devastated European vineyards during the XIX century (Granett et al. 2001). Most of the current rootstocks used in viticulture were generated from resistant American vines (Granett et al. 2001). Besides resistance to phylloxera, rootstocks have been selected for other characteristics, such as resistance to nematodes (Esmenjaud and Bouquet 2009), tolerance to limestone (Bavaresco et al. 2003), drought (Carbonneau 1985), salinity (Fisarakis et al. 2005), and improvements in grapevine vigour (Zhang et al. 2016b). There have been also efforts to develop transgenic rootstocks resistant to the *Grapevine fanleaf virus* (GFLV), one of the most important viral diseases of grapevines (Valat et al. 2006). As rootstocks are anchored in the soil and they differ in their ability to absorb minerals, this has an effect in fruit quality (Warschefsky et al. 2016). In addition the use of different combination of rootstocks can generate significant phenotypic effect in the scion, such as changes in the canopy structure and size which in turn affects the grapevine water use (Serra et al. 2014).

Human practices

According to van Leeuwen and Seguin (2006), human factors, such as history, socio-economics, as well as viticultural and oenological techniques are also part of *terroir*. Viticulture is a human activity, and the history of the socio-economic environment may be important in understanding why a given vineyard has emerged in a given site and why it has prospered. Mastery of

viticultural and oenological practices are necessary in order to optimise the wine to the potential of the natural environmental factors (van Leeuwen and Seguin 2006).

In viticulture, a set of general principles called the "cool climate paradigm" is often followed for obtaining premium wines. These are: (i) keep the fruit warm; (ii) keep the leaves exposed to sunlight; (iii) achieve vine balance between vegetative growth and crop size; and (iv) manage water stress (Reynolds 2010). These requirements can be met through agronomic practices such as canopy management and irrigation.

Canopy management

Canopies are managed to control the exposure of the fruit and leaves to sun and to maintain a consistent moderate to mild water stress between fruit set and harvest (Gladstones 2011b). The leaf area exposed to the sun strongly affects vine vigour, and, with proper management, canopy shading can be avoided. Pruning can also limit leaf area and, in this way, reduce water loss through transpiration. As fruits are exposed to the sun, their temperature increases and this will increase enzymatic activity (Reynolds 2010) affecting the metabolite composition of the berry (Cohen et al. 2008; Pereira et al. 2006). Fruit temperature can be managed by using the appropriate training system and by canopy management practices, such as basal leaf removal. Initial studies showed that basal leaf removal could significantly reduce titratable acidity by reducing the concentration of malic acid, reducing botrytis infection rate at the same time (Wolf et al. 1986; Smith et al. 1988). Another principal of the "cool climate paradigm" is that the vine must be balanced between vine vigour and crop size (Gladstones 2011b). A common practice is to restrict the crop size to optimise both yield and fruit maturity to obtain high quality fruit. Traditionally, pruning has been used to maintain this balance.

Irrigation

Although the traditional European view is that irrigation reduces wine quality, irrigation in dry regions offers several practical advantages, such as better control over the vine's water economy (Gladstones 2011b). While traditional sprinkler and furrow irrigation requires replenishment of the upper soil layer, drip irrigation offers better water use efficiency (Hepner and Bravdo 1985). Partial Rootzone Drying is an irrigation technique that consists in drying half of a vine root system while the other half is frequently irrigated for alternate "dry" and "wet" zones (Dry and Loveys 1998). This technique aims to reduce vine vigour by imposing drought stress. In these conditions, abscisic acid (ABA), a hormone that regulates transpiration and shoot growth, is synthesised without affecting crop yield and fruit quality (Antolín et al. 2006).

Grapevine varieties

The timing of grape ripening initiation, length of maturation period, berry size and seed content can significantly differ between different grapevine varieties (Costantini et al. 2008; Duchêne et al. 2010). The availability of early and late ripening varieties is desirable for growers who take advantage of this feature to select varieties best fitted to the local environmental conditions and for staggering harvest along the growing season (Costantini et al. 2008). Early ripening varieties, such as Pinot noir, Chardonnay and Gewürztraminer, are usually grown in cooler regions at high latitudes. Meanwhile, the late-ripening varieties, like Grenache and Mourvèdre, are preferred in warmer weathers at lower latitudes (van Leeuwen 2010). The effect of *terroir* on grapevine growth and fruit composition reflects the interaction between the vine and its surroundings (Anesi et al. 2015). Understanding how and to what extent these elements affect a specific grapevine genotype will allow us the opportunity to implement viticultural practices in order to obtain the desired fruit characteristics for each climate-cultivar combination (Jones and Davis 2000).

Decoding *terroir*

Many analytical studies have been carried out to understand *terroir* from the viticultural point of view, mainly through characterisation of the different environmental factors that affect grape berry composition and wine quality (Robinson et al. 2012). In addition, studies have investigated the molecular mechanisms involved in grapevine's response to its growing environment and the potential effect that this could have on berry ripening. Research into the molecular basis of *terroir* has increased with the development of high throughput sequencing technologies. More information about decoding *terroir* using 'omics can be found in chapter 2 in the review Jullian Fabres et al. (2017).

Gene expression analyses in *V. vinifera*

Grape berries are non-climacteric fruits that undergo a series of complex biochemical changes during development and ripening that can be divided into three major phases (Coombe 1992). In phase I there is a sigmoidal growth curve in berry size due to cell division and cell expansion. There is also accumulation of organic acids (mainly malate and tartrate), tannins and hydroxycinnamates (Coombe 1992). Phase II is characterized as a lag phase in which cell expansion stops and sugars start to accumulate. Phase III starts with veraison, when there is a second sigmoidal growth of the berries due to mesocarp cell expansion. During this phase anthocyanin pigments for berry color and volatile compounds for aroma and sugars (mainly glucose and fructose) accumulate and organic acids concentrations decline (Coombe 1992). Deluc et al. (2007) assembled one the first transcriptional networks during grape berry development (cv. Cabernet Sauvignon), showing that ripening is a dynamic process and a considerable number of genes are expressed. They also revealed the expression patterns of genes known to play a key role during ripening (for ABA synthesis, cell wall metabolism, flavonoid metabolism, organic acid metabolism and sugars metabolism). Similar results were obtained by

transcriptomic analysis in grape berry from the cv. Shiraz in four developmental stages (Sweetman et al. 2012). The study identified 4,185 upregulated genes with specific genes profiles for each developmental stage and coordination in the transcription of secondary metabolites such as organic acids, stilbene and terpenoids. Since then, many studies have used transcriptomic analysis to analyse grape berry softening during ripening (Glissant et al. 2008), the identification of stage specific biomarkers for berry development (Zamboni et al. 2010), the differences in gene expression between different varieties during berry development (Degu et al. 2014) and the ethylene signaling pathway during the late stage of grape berry ripening (Cramer et al. 2014). In addition, transcriptomic analyses have been used to analyse grapevine response to fungal infection (Agudelo-Romero et al. 2015; Camps et al. 2010), heat stress (Lecourieux et al. 2017; Liu et al. 2012), exposure to light (du Plessis et al. 2017), water stress (Berdeja et al. 2015; Dal Santo et al. 2016b) and UV radiation (Carbonell-Bejerano et al. 2014; Martínez-Lüscher et al. 2014; Pontin et al. 2010; Suzuki et al. 2015). More recently, Fasoli et al (2018) analysed the transcriptome and metabolome of two grapevine varieties by sampling every week from fruit set to maturity over three consecutive vintages revealing the more up to date gene expression atlas of berry development in grapevine. Researchers were able to identify coordinated waves on the gene expression associated with the different developmental stages of grape berries. They were also able to identify a core network of transcripts involved in the reprograming of berry development.

Grape berries are characterized by considerable phenotypic plasticity (Keller 2010). Here, phenotypic plasticity refers to the range of phenotypes a single genotype can express as a function of its environment (Dal Santo et al. 2013). Through transcriptome analysis, Dal Santo (2013), analysed the phenotypic plasticity of grape berries form the variety Corvina grown in eleven different vineyards during three consecutive years. The study revealed that the year of growth had the strongest effect on the transcriptome, while transcripts related to secondary

metabolites were more sensitive to the different climates. Among the plastic transcripts, common transcripts were identified between vineyards sharing similar viticultural practices or environmental conditions. In another study, metabolomic and transcriptomic analyses were used to examine the phenotypic plasticity of grape berries from the white variety Garganega grown at four sites with different growing conditions (altitude, soil texture, and composition) (Dal Santo et al. 2016a). Clustering analysis revealed that the growing sites had a strong effect on the transcriptome and environmentally-modulated genes were identified. Additionally, researchers observed that genes involved in the biosynthesis pathway of phenylpropanoids showed a plastic response to the environment, reflecting the accumulation of the corresponding metabolites. More recently Dal Santo et al. (2018) analysed the grape berry transcriptome plasticity of the varieties Sangiovese and Cabernet Sauvignon cultivated in three different environments at four developmental stages over two growing years. By analyzing the gene expression profile, they identified genes unaffected by genotype or environment, genes that were genotype-dependent, genes that were environment-dependent and genes influenced by the genotype by environment (GxE) interaction. Notably, the GxE related genes were enriched for stress response, signaling and secondary metabolism categories (Dal Santo et al. 2018).

A recent study analysed the transcriptome of Cabernet Sauvignon berries at the late stages of berry development grown in two different locations (Reno, USA and Bordeaux, France) identifying a set of common ripening genes between the two locations. Additionally, the authors observed that most of the differences in genes expression could be associated to the environmental conditions between locations (Cramer et al. 2019).

Epigenetics in plants

Many studies have shown the impact of the environment on the epigenetic regulation of genes in plants (Baulcombe and Dean 2014; Bräutigam et al. 2013). Notably, epigenetic mechanisms can

have a phenotypic effect independent of the genotype (Herrera and Bazaga 2013; Zhang et al. 2013). Epigenetics refers to modifications of the genome that do not change the nucleotide sequence (Berger et al. 2009; Eichten et al. 2014; Giovannoni et al. 2017) that are usually mediated by DNA methylation and chromatin modifications (Feil and Fraga 2012; Gallusci et al. 2016; Rodríguez López and Wilkinson 2015). DNA methylation refers to the addition of a methyl group to the carbon 5 of cytosine [5-methylcytosine (5mC)] (Gallusci et al. 2016). In plants, DNA methylation occurs at the sequence contexts CG, CHG and CHH (where H = A, T or C) (Law and Jacobsen 2010). In *Arabidopsis* spp., cytosine methylation is catalyzed and maintained by DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM 1/2), DNA METHYLTRANSFERASE 1 (MET1) in the CG context, CHROMOMETHYLASE 3 (CMT3) for the CHG context (Law and Jacobsen 2010) and CHROMOMETHYLASE 2 (CMT2) for the context CHH (Stroud et al. 2014). Less understood are the processes that dictate where and when DNA methylation is established and it is here that environmental signals could have a considerable impact (Feil and Fraga 2012). Most of the studies about the impact of the environment on the epigenome have been on DNA methylation, which is essential in plant development and for silencing of transposable elements (TE) in the genome (Zhang et al. 2006). In *A. thaliana*, high levels of DNA methylation or hypermethylation in gene promoter regions have been correlated with gene silencing; this contrasts with hypermethylation in gene bodies which can enhance gene expression (Suzuki and Bird 2008; Jones 2012). Genome-wide methylation studies in *Arabidopsis thaliana* have shown that over one-third of expressed genes present methyl groups within the coding region, while only 5% presented methylation in their promoter regions (Zhang et al. 2006; Vaughn et al. 2007). Notably, promoter-methylated genes had a high degree of tissue specific expression (Zhang et al. 2006). In another genome-wide study in *A. thaliana*, gene-body methylated genes exhibited a moderate level of expression (Zilberman et al. 2007). Many of these genes could be classified as 'housekeeping genes',

necessary for basic cellular processes. By contrast, highly expressed genes and low expressed genes, presented similar low levels of methylation.

Developmental and environmental epigenetics

The impact of DNA methylation in plant development has been shown through the cloning of the *Colourless non-ripening (Cnr)* locus and its effect in tomato ripening (Manning et al. 2006). The *Colourless non-ripening (Cnr)* locus plays a key role in the regulatory network controlling fruit ripening, and the *Cnr* mutant is a stably inherited natural epigenetic mutation that produces tomatoes fruits with abnormal ripening. Manning et al. (2006) identified a gene with high homology to the SBP-box transcription factor family (named *LeSPL-CNR*) in the *Cnr* locus. They found that, in the *Cnr* mutant, the promoter of the *LeSPL-CNR* gene was highly methylated (Manning et al. 2006). In another study, the effect of DNA demethylation in fruit ripening was assessed. Tomatoes treated with the demethylating agent 5-AzacCytidine (5-AzaC) ripened prematurely (Zhong et al. 2013). Whole-genome bisulfite sequencing revealed that 1% of the tomato genome was differentially methylated during ripening (Zhong et al. 2013) and researchers noticed a loss of 5mC in the promoter of more than 200 ripening-related genes. The differentially methylated regions in these genes were typically adjacent to binding sites for RIN (Zhong et al. 2013), a MADS-box transcription factor that acts as a master regulator of ripening in tomato (Vrebalov et al. 2002). These results suggested that DNA demethylation played a role in the ripening process. These results were then confirmed by Liu et al. (Liu et al. 2015) who identified a relationship between active demethylation and the induction of fruit ripening, demonstrating epigenetic control of fruit ripening.

Epigenetic modifications have also been identified as playing a key role in seed development. Amoah et al. (2012) treated *Brassica rapa* seeds with 5-AzaC in order to generate heritable epialleles. The resultant hypomethylated progeny showed diverse effects on seed size and

composition, as well as others yield components, demonstrating that epigenetic changes could potentially modify the composition and traits of economically important crops. Together these results show that plant development can be modified through environmentally-induced DNA methylation.

Several studies have shown that an altered DNA methylation pattern, caused by the environment, can affect plant phenotype. One of the best well-known examples in which the environment affects the phenotype through epigenetic modifications is vernalisation (Burn et al. 1993; Feil and Fraga 2012; Finnegan et al. 1998). Through this process, plants in temperate regions mitigate the deleterious effects of low winter temperatures on flower and fruit development by breaking dormancy only after having been exposed to a cold period (Kumar et al. 2016). Unusual environmental conditions during dormancy, such as high winter temperatures, have been shown to exert a negative effect on fruit quality and the yield of perennial crops requiring a vernalisation period (Sugiura et al. 2012). A study in apple showed that the methylation and expression levels of key genes involved in flowering and fruit set were modified by the level of chill received during bud dormancy (Kumar et al. 2016). Additionally, reduced levels of flavour volatiles have been associated with cold-induced changes in DNA methylation and expression levels of genes encoding volatile synthesis enzymes in tomatoes (Zhang et al. 2016a). Although the biological implications of the changes in the levels of such volatiles are not fully understood, it is clear that environmentally-induced methylation changes affected the fidelity of gene expression required to respond to environmental challenges, without disrupting broader fruit developmental processes (Zhang et al. 2016a).

Differences in growing conditions have been shown to modify the phenotype of plants through epigenetic variations. Fonseca Lira-Madeiros et al. (2010) analysed mangrove plants from two, nearby habitats with different nutrition. They saw clear morphological changes between the

plants of the different populations and these were strongly correlated with differences in the DNA methylation levels (a difference of 17.5% of methylated loci between populations) (Fonseca Lira-Medeiros et al. 2010). Furthermore, they observed that epigenetic diversity was greater than the genetic diversity in this species, suggesting that epigenetic variations in plants populations could help them to overcome environmental stresses. Verhoeven et al. (2010) studied this possibility by analysing DNA methylation in asexual dandelions under several stresses. Apomictic dandelions reproduce through unfertilized seeds, and the offspring are genetic copies of the mother plant; so by limiting the genetic variation, they were able to analyse only the epigenetic effect. They showed that changes in the DNA methylation in dandelions under stressed conditions were common and heritable. They also suggested that the variation in the methylation pattern was independent of the DNA sequence, although some loci were more susceptible to methylation.

Grapevine epigenetics

Grapevine combines several features that could make it an appealing model to study epigenetic regulations in woody, perennial plants. It is used as one of the main models for non-climacteric fruits, and also its flower development is programmed one year in advance; the impact of environmental conditions on flower and subsequently fruit development seem to be, in part, determined by the environmental conditions the year before (Guilpart et al. 2014).

Commercially cultivated grapevines are normally propagated vegetatively. Long-living, vegetatively propagated plants accumulate somatic variation with time (Baali-Cherif and Besnard 2005). This variation is considered the source of monozygotic clones that give rise to one variety, as opposed to polyzygotic clones (those originated from multiple plants with similar phenotypic characteristics) (Ocaña et al. 2013). Clonal variability has been shown to be genetic and epigenetic in origin (Zhang et al. 2006). Epifingerprinting studies (Rodríguez López and

Wilkinson 2015) have been performed in *V. vinifera* for the identification of clones that were otherwise genetically identical (Imazio et al. 2002; Ocaña et al. 2013; Schellenbaum et al. 2008).

Another common practice in grapevine breeding programs is to use *in vitro* propagation to clonally propagate a desired genotype. It is well documented that *in vitro* propagation can induce molecular and phenotypic variation, referred to as somaclonal variation (Miguel and Marum 2011; Schellenbaum et al. 2008). While genetic variation has been found to be induced at a low rate, changes in DNA methylation profiles occur at higher rates and have been linked to the observed phenotypic variation (Kitimu et al. 2015; Miguel and Marum 2011). Baránek et al., (2015) reported that grapevine plants stressed by *in vitro* cultivation had altered DNA methylation patterns. Moreover, after the stress was ended, 40% of the acquired aberrant methylation gradually reverted to a similar state as found in the donor plant. However, the remaining 60% of the observed DNA methylation diversity was still maintained one year after plant ceased to be exposed to the *in vitro* conditions.

Epigenetic mechanisms have also been associated with the regulation of the biosynthesis of metabolites in *Vitis*. One of the most well-known secondary metabolites in plants is resveratrol, a phenolic compound that has been related with a range of human health benefits (Aziz et al. 2003; Shankar et al. 2007). Resveratrol is synthesised via the phenylpropanoid pathway, and stilbene synthase (STS) directly catalysed resveratrol formation. Using cell cultures of *Vitis amurensis,* a close relative of *V. vinifera*, Tyunin et al. (2013) assessed the synthesis of resveratrol. Plant cell cultures treated with 5-AzaC showed a significant increase in the gene expression of *VaSTS10*, and also an increase in the synthesis of resveratrol (Tyunin et al. 2013). Additionally, they found that 5-AzaC treatment significantly reduced the methylation level of the *VaSTS10* gene, correlating with its increased gene expression.

In another study, UV radiation affected the accumulation of phenolic compounds in the berry skin of grapevine (cv. Malbec) (Marfil et al. 2019). Epigenetic analysis through methylation sensitive amplified polymorphisms (MSAPs) revealed that treatments of UV-B and water deficit induced significant changes in the DNA methylation of samples. UV-B treatment was also associated with flavonol accumulation in berries, and hydroxycinnamic acid in the next season's fruit shoots. The researchers suggested that DNA methylation could regulate the accumulation of low molecular weight phenols in grapevine under UV light exposure (Marfil et al. 2019).

Recently, it was reported that the global DNA methylation patterns could explain *terroir* effect in grapevine. Epigenetic diversity in samples from 22 vineyards from six sub-regions of the Barossa Wine Zone was assessed through MSAPs (Xie et al. 2017). DNA methylation profiles showed a high level of differentiation that grouped vineyards by their area of provenance. In addition, researchers observed that the main driver of DNA methylation diversity between vineyards was the distribution of the vineyards on a north to south axis. Notably, vineyards pruned with the same method, showed a higher correlation between geographic distance and epigenetic distance, suggesting that the pruning system affected inter-vineyard epigenetic differentiation. Methylation sensitive genotyping by sequencing (msGBS) analysis from eleven vineyards from three sub-regions of the Barossa Wine Zone also revealed 3,598 differentially methylated genes in grapevine leaves, where 8.6% were genes associated with response to an environmental stimulus (Xie et al. 2017). These data suggested that DNA methylation differences between vineyards and sub-regions within the Barossa were influenced both by the geographic location and, to a lesser extent, by vineyard altitude and the pruning system. However, in a recent study investigating the interaction between genotype (Sangiovese and Cabernet sauvignon) and environment (three different location over two vintages) in grapevine berries, DNA methylation remained stable regardless of variation in external cues and in gene expression and plant genotype was the main factor in the variance in methylation between samples (Dal Santo et al. 2018). A study analysing

the effect of genetic and epigenetic properties on gene expression in ten grape varieties, revealed that gene body methylation was negatively correlated with gene expression variation between grapevine varieties (Magris et al. 2019). The authors suggested that the negative association between gene body methylation and expression variation may have resulted from the higher density of transposable elements present within grapevine genes.

Mobile transcripts in plants

Non-coding RNAs are part of the molecular epigenetic machinery regulating gene expression along with DNA methylation and histone modifications (Choudhuri 2011). Plants can use non-coding RNAs, such as small interfering RNAs (siRNAs) and microRNAs (miRNAs) as a mechanism of defense against viral infections and in regulating plant development respectively (Lewsey et al. 2007). Additionally, siRNAs and miRNA are able to mobilize long distances through the phloem (Kalantidis et al. 2008). Mobile miRNAs have been identified in three *V. vinifera* cultivars (Chardonnay, Chenin blanc, Cabernet Sauvignon) from the phloem tissue (Bester et al. 2017). There are also reports that messenger RNAs (mRNA) are able move through the phloem to distant tissues. This long-distance mobility of mRNA was reported in tomatoes (Kim et al. 2001) and potatoes (Banerjee et al. 2006) in a KNOTTED1-like homeobox gene, an important family of transcription factors that regulate development of the apical meristem. More recently, in *A. thaliana* 2,006 genes that produced mobile mRNA were identified and some of these were shown to transcribe to proteins in their destination sites (Thieme et al. 2015).

A study looking at the effect of grafting in grapevine, identified more than 3000 genes transporting mRNA across graph junctions, including genes for response to abiotic stress and signal transduction (Yang et al. 2015). Recent evidence suggests that mobile transcripts may impact the epigenetic state of both rootstocks and scion (Fortes and Gallusci 2017). In grafted *A. thaliana* mobile smalls RNAs can regulate the DNA methylation status of thousands of sites

associated with TEs (Lewsey et al. 2016). In grapevine, grafting induced the up-regulation of genes associated with DNA methylation and chromatin modification in the shoot apical meristem (Cookson and Ollat 2013). Expression of mobile transcripts could potentially provide a mechanism of communication between tissues in response to environmental cues.

The mechanisms involved in enabling intercellular mRNA transport are still are poorly understood. Recent evidence demonstrated that polycistronic transcripts were systemically transported via the plant phloem vasculature (Zhang et al. 2016c). Polycistronic mRNAs are RNA molecules that contain two or more open reading frames (ORFs). They are common in bacteria and archaea, where genes are often arranged into clusters and co-transcribed into a single polycistronic mRNA (Sugita and Sugiura 1996). However, polycistronic transcripts are less common in eukaryotes and most nuclear encoded genes in plants are monocistronic. An analysis revealed that 7.5 % (249 of 3333) of the mobile mRNA in grapevine (Yang et al. 2015) had a tRNA-like structure (TLS) motifs in their coding sequences or 3`UTR (Zhang et al. 2016c). TLSs are found in various groups of plant viruses at the end of the mRNA instead of 3` poly(A) tail, and have a key role during viral infection (Barends et al. 2004).

A recent work in *A. thaliana* and tobacco showed that TLSs co-transcribed with mRNA transcripts could act as mobility signals, triggering the systemic movement of the mRNA between roots and shoots (Zhang et al. 2016c). Notably, the mRNA components of the dicistronic transcripts were also shown to be translated into functional proteins. Endogenously produced tRNA-mRNA dicistronic transcripts have also been detected in *A. thaliana,* suggesting that functional tRNA and TLSs could act as non-autonomous signals in plants, being able to deliver functional mRNA to distantly located tissues (Zhang et al. 2016c). This study suggested that tRNAs and TLSs could act as signalling molecules in plants mobilising functional mRNA

between distantly located tissues and, potentially, influencing environmental adaptation between tissues in long-lived plants.

Research Aims

The aims of this PhD project were to characterize the methylome and transcriptome of 198 Shiraz plants cultivated across twenty-two vineyards, covering six sub-regions from the Barossa Wine Zone, Australia. We hypothesized that DNA methylation can play a role in defining *terroir*. To test this, we integrated vineyard information (climate, soil and management) with genomic, epigenomic and transcriptomic data from leaf tissue, to identify molecular drivers of *terroir* in *V. vinifera* from the Barossa Wine Zone. Leaf tissue can acquire environmentally-inducible epigenetic markers that reflect the local growing conditions. In this work we present a large scale attempt to analyse the effect of the environment on DNA methylation and on gene expression in field-grown grapevines. We also analysed the synthesis of dicistronic tRNA-mRNA transcripts in grapevine from leaf and berry gene expression data. We hypothesized that dicistronic transcripts could work as long-distance signaling molecules to coordinate grapevine response to environmental cues. In order to explore this, we identified and analysed the expression of tRNA-mRNA dicistronic transcripts in grapevine from leaf and berry samples from the twenty-two vineyards from the Barossa Wine Zone. We also assessed if the environment and the tissues had a differential effect on the expression of dicistronic transcripts.

Chapter 2: A concise review on multi-omics data integration for *terroir* analysis in *Vitis vinifera*

Statement of Authorship

Title of Paper	A Concise Review on Multi-Omics Data Integration for Terroir Analysis *in Vitis vinifera*
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Fabres PJ, Collins C, Cavagnaro TR, Rodríguez López CM. A Concise Review on Multi-Omics Data Integration for Terroir Analysis in Vitis vinifera. Front Plant Sci. 2017 Jun 20;8:1065. This is a review article that summarizes the current understanding of the terroir from a multi-omics point of view and the different strategies used to combine different high throughput data to better understand the terroir effect.

Principal Author

Name of Principal Author (Candidate)	Pastor Jullian Fabres
Contribution to the Paper	Collected, analysed and interpreted literature, wrote manuscript.
Overall percentage (%)	65%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 5/11/19

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Cassandra Collins
Contribution to the Paper	Evaluate and edited the manuscript, I hereby certify that the statement of the contribution is accurate.
Signature	Date 5/11/19

Name of Co-Author	Timothy R. Cavagnaro
Contribution to the Paper	Evaluate and edited the manuscript, I hereby certify that the statement of the contribution is accurate.
Signature	Date 6th Nov., 2019.

Name of Co-Author	Carlos M. Rodríguez López
Contribution to the Paper	Wrote, evaluate and edited the manuscript. Acted as the corresponding author. I hereby certify that the statement of the contribution is accurate.
Signature	Date 6/11/2019

Link between Chapters 1 and 2

This review (chapter 2) focuses on the study of terroir from a multi-omics point of view and the different experiment and integration strategies that could be used. This paper provides an overview of the current literature on how to unravel the molecular basis of environmental effects in plants, especially in grapevines. We also discuss the possible function of epigenetic mechanisms in plant adaptation to the local growing conditions. This chapter has been published as follow: Fabres PJ, Collins C, Cavagnaro TR, Rodríguez López CM. A Concise Review on Multi-Omics Data Integration for Terroir Analysis in *Vitis vinifera*. Front Plant Sci. 2017 Jun 20; 8: 1065. doi: 10.3389/fpls.2017.01065.

frontiers in Plant Science

MINI REVIEW
published: 20 June 2017
doi: 10.3389/fpls.2017.01065

A Concise Review on Multi-Omics Data Integration for Terroir Analysis in *Vitis vinifera*

*Pastor Jullian Fabres[1], Cassandra Collins[2], Timothy R. Cavagnaro[2] and Carlos M. Rodríguez López[1]**

[1] Environmental Epigenetics and Genetics Group, Plant Research Centre, School of Agriculture, Food and Wine, University of Adelaide, Glen Osmond, SA, Australia, [2] The Waite Research Institute, The School of Agriculture, Food and Wine, The University of Adelaide, Glen Osmond, SA, Australia

Vitis vinifera (grapevine) is one of the most important fruit crops, both for fresh consumption and wine and spirit production. The term *terroir* is frequently used in viticulture and the wine industry to relate wine sensory attributes to its geographic origin. Although, it can be cultivated in a wide range of environments, differences in growing conditions have a significant impact on fruit traits that ultimately affect wine quality. Understanding how fruit quality and yield are controlled at a molecular level in grapevine in response to environmental cues has been a major driver of research. Advances in the area of genomics, epigenomics, transcriptomics, proteomics and metabolomics, have significantly increased our knowledge on the abiotic regulation of yield and quality in many crop species, including *V. vinifera*. The integrated analysis of multiple 'omics' can give us the opportunity to better understand how plants modulate their response to different environments. However, 'omics' technologies provide a large amount of biological data and its interpretation is not always straightforward, especially when different 'omic' results are combined. Here we examine the current strategies used to integrate multi-omics, and how these have been used in *V. vinifera*. In addition, we also discuss the importance of including epigenomics data when integrating omics data as epigenetic mechanisms could play a major role as an intermediary between the environment and the genome.

Keywords: multi-omics, environment, *Vitis vinifera*, data integration, epigenetics, transcriptomics, metabolomics

OPEN ACCESS

Edited by:
Giovanni Battista Tornielli, University of Verona, Italy

Reviewed by:
Pablo Carbonell-Bejerano, Instituto de las Ciencias de la Vid y del Vino (ICVV), CSIC, Spain
Stefania Pilati, Fondazione Edmund Mach, Italy

****Correspondence:***
Carlos M. Rodríguez López
carlos.rodriguezlopez@adelaide.edu.au

Specialty section:
This article was submitted to Crop Science and Horticulture, a section of the journal Frontiers in Plant Science

Received: *08 February 2017*
Accepted: *02 June 2017*
Published: *20 June 2017*

Citation:
Fabres PJ, Collins C, Cavagnaro TR and Rodríguez López CM (2017) A Concise Review on Multi-Omics Data Integration for Terroir Analysis in Vitis vinifera. Front. Plant Sci. 8:1065. doi: 10.3389/fpls.2017.01065

INTRODUCTION

Grapevine is one of the most economically important fruit crops, and it is largely used for wine production (FAO, 2012). Most the chemical compounds that give its unique characteristics to wine are synthesized during berry development (Conde et al., 2007). However, fruit/wine composition is strongly influenced by the interactions between the plant's genome and the local growing conditions (including the vine management system), and the oenological practices of each winery (**Figure 1**), which could explain why it is so difficult to replicate a wine from a region outside that area.

Terroir is defined as the interactions between the plants, the environment and human factors (Gladstones, 2011) and it is frequently used to relate wine sensory attributes to its geographic origin (Van Leeuwen and Seguin, 2006). Although the relevance of *terroir* is still under debate (Anesi et al., 2015), a better understanding of how the environment affects grape berry composition can have a

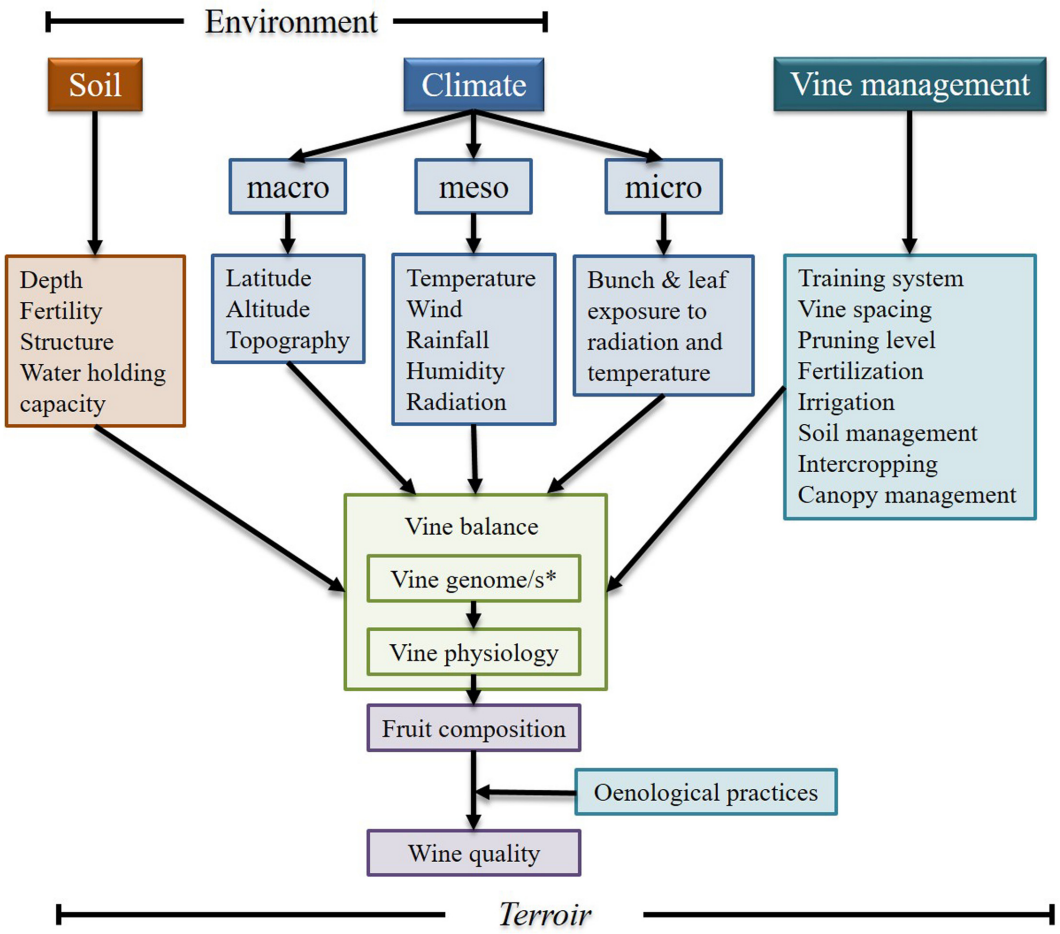


FIGURE 1 | A conceptual view of some of the *terroir* factors that contribute to vine balance, fruit composition and wine quality. *Vine genome/s refers to the possibility of the scion and rootstock used in a vineyard could be from different varieties/species (Modified from: Smart et al., 1985 and reproduced with permission from the copyright holder).

significant effect on viticulture. To achieve such an understanding, it is necessary to identify the elements that drive *terroir* and analyze the interaction between them and the grapevine.

DECODING *Terroir*

Terroir has been long studied, through the characterization of the different environmental factors affecting berry composition and wine quality, and climate exerts the strongest effect on berry composition (Robinson et al., 2012). Soil physicochemical properties as well have been identified as an influential factor defining the uniqueness of berry composition by vines grown in a specific climate (Cheng et al., 2014; Zerihun et al., 2015). Grapevine microbiome community may play an important role determining wine quality (Burns et al., 2015; Bokulich et al., 2016). Efforts have been made to study the grapevine microbiome landscape in relation to the vegetative growth cycle of the plant (Pinto et al., 2014), the post-harvest treatment of berries (Salvetti et al., 2016) and provenance (Bokulich et al., 2016) (For a review on microbiome analysis see Ibrahim and Kumar, 2017). Less work has been done to elucidate the molecular mechanisms involved in the plant response to *terroir*. A strategy to better understand the genome and environment interaction is to use 'omics' technologies. Omics refers to high throughput technologies that generate a large amount of data for each sample, allowing a deeper insight of the mechanisms regulating biological systems.

ANALYSIS OF *Terroir* EFFECT ON GRAPE COMPOSITION USING TRANSCRIPTOMICS

Using transcriptomics is possible to study the grape's complete set of RNA transcripts encoded by the genome using high throughput methods (Hale et al., 2005). Dal Santo et al. (2013) performed gene expression analysis in a single Corvina clone cultivated in 11 different vineyards for three consecutive years.

Samples strongly clustered by season, known as a vintage effect, rather than common environmental conditions. However, the genes that showed more variation in expression between years were those involved in secondary metabolism, (mainly the biosynthesis of phenylpropanoids). Only when samples from a single vintage (i.e., 2008) were analyzed, it was observed that 5% of the studied annotated coding genes were differentially regulated under different growing conditions and agronomical practices. Anesi et al. (2015) complemented this study by analyzing the transcriptome and metabolome of the same cultivar. They identified metabolites that could describe a *terroir* signature for each vineyard. Moreover, it was possible to correlate terroir-sensitive metabolites with changes in the transcript level of genes involved in the biosynthesis of these metabolites. Similar results were obtained by Dal Santo et al. (2016) as they identified a clear correlation between gene expression and accumulation of phenylpropanoids and flavonoids in the variety Garganega grown at four different vineyards. Small RNA profiles have been analyzed to understand the interaction between genotype and environment in the varieties Sangiovese and Cabernet Sauvignon. *In silico* analysis suggests that microRNAs may be involved in berry development and the accumulation of secondary metabolites (Paim Pinto et al., 2016). Transcriptional analysis of berries from different regions has also shown that transcripts from the abscisic acid (ABA) biosynthesis pathway are among the most *terroir* sensitive genes (Sun et al., 2015). ABA is a plant hormone that regulates important steps in plant growth and development as well as play a key role in plant biotic and abiotic stress response (Cutler et al., 2010). ABA concentrations affect anthocyanin and flavonol accumulation (Koyama et al., 2010), suggesting a possible mechanism through which the environment affects grape berry composition and wine flavor and aroma.

ANALYSIS OF *Terroir* EFFECT ON GRAPE COMPOSITION USING METABOLOMICS

Metabolomics is defined as the identification and quantification of metabolites using high-throughput techniques (Cevallos-Cevallos et al., 2009). This technology can screen higher numbers of products than more traditional approaches (Pereira et al., 2006; Atanassov et al., 2009; Hong, 2011), while the use of non-targeted metabolomics approaches allows the identification of un characterized metabolites (Panighel et al., 2015). *Terroir* can be explored by analyzing berry metabolite composition through different analytical methods (For a review in grape and wine metabolomics see Cozzolino, 2016). Son et al. (2009) identified that differences in berry metabolomes associated to environmental regional differences (radiation and rainfall) could explain the observed differences in wine composition. Similar results were obtained by Tarr et al. (2013) who distinguished the metabolic signatures of different grapevine varieties. Metabolomic analysis has also been performed to identify chemical compounds that can be associated to regional wine quality traits (Gambetta et al., 2016, 2017). Roullier-Gall et al. (2014) assessed the metabolomics profiles of two different *terroirs,* which were just 2 km apart, over three vintages. Although vintage had the greatest effect in the berry's metabolite composition, differences in fruit chemical composition associated to nearby *terroirs* could be detected when vintages were individually analyzed. This suggests that subtle geographical differences have a significant effect on grape/wine composition even when variability within vineyards can be relatively high (Mulas et al., 2011).

MULTI-OMICS INTEGRATION

The aim of integrating multi-omic data is to reduce the gap between data generation and the ability to analyze and understand the biological mechanisms behind an organism's response to environmental cues. The objective of multi-omic data integration is to combine different types of data to construct a model that can be used to predict complex traits and phenotypes (**Figure 2**). This approach also allows the identification of biomarkers and of previously unknown relationships between the datasets (Rajasundaram and Selbig, 2016). Through the integration of environmental information with genomic, epigenomic, transcriptomic, and metabolomic data, we hypothesize that it will be possible to better understand the effect of terroir at a molecular level. The use of a multi-omic approach will also help reduce the incidence of false positives generated from single source data sets (Aho, 2013; Ritchie et al., 2015). However, integration of multi-omics data is not a trivial task, because the diversity of characteristics of the data generated from the different high throughput technologies (machine sensitivity, error rate, data structure) makes its combination challenging.

APPROACHES TO DATASETS INTEGRATION

Analysis of large data sets from different origins has been done using two main approaches: network models (NMs) and pathway analysis (PA). Both share the basic idea of storing the data in a clear and meaningful way. NMs use concepts from mathematical graph theory, to represent biological components (e.g., genes) as nodes and their interactions (physical, genetic or functional) as their links (For a review on NM applied to plant biology see Fukushima et al., 2014). NMs are classified as homogeneous or heterogeneous depending on the number of different levels of information integrated (Gligorijević and Pržulj, 2015). Homogenous approaches integrate datasets with the same type of nodes and therefore cannot analyze the connectivity between multiple datasets simultaneously. However, complex biological questions such as the molecular regulation of fruit composition in grapevine are increasingly being addressed through the integration of multiple layers of cellular information (Wong and Matus, 2017), including but not restricted to genomics, transcriptomics, proteomics

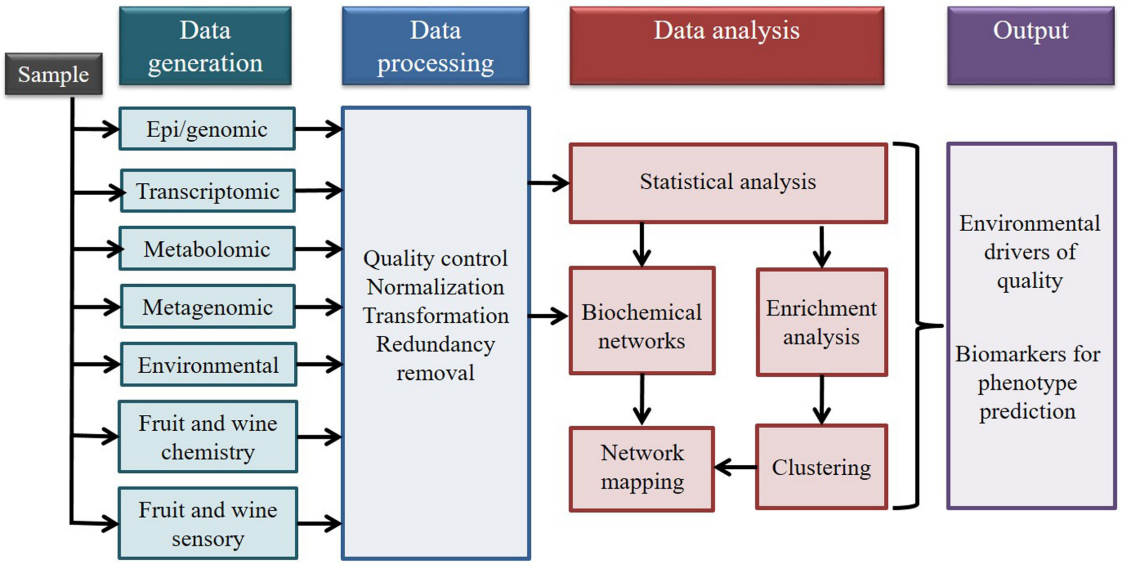


FIGURE 2 | Data integration workflow for four omics technologies in addition to environmental data, and fruit and wine quality data (Modified from Wanichthanarak et al., 2015 and reproduced with permission from the copyright holder).

and metabolomics, using heterogeneous methods. Bayesian networks (BNs) and Kernel-based methods (KBMs) are heterogeneous approaches commonly used for data integration (Zhang, 2009; Gligorijević and Pržulj, 2015). BNs are efficient detecting relationships hidden in noisy datasets but they are computationally demanding (Gligorijević and Pržulj, 2015) and are therefore, better suited for the interrogation of small datasets in hypothesis driven questions (Gligorijević and Pržulj, 2015) (i.e., the analysis of terroir on defined pathways such as those leading to the biosynthesis of metabolites related to fruit quality). KBMs are not as computationally demanding and so can integrate large molecular, structural and phenotypic datasets (Mizrachi et al., 2017), making them ideal for data driven terroir exploratory studies, biomarker discovery or for the reclassification of previously identified drivers of quality (Qi et al., 2008).

On the other hand, pathway analysis requires well documented biochemical pathways where omics data is combined to seek overrepresented groups (Wanichthanarak et al., 2015). For example, multiple co-inertia analysis (MCIA) can detect explanatory omic features even when they are not present in all datasets (Meng et al., 2014), which makes it attractive for the integration of terroir data from different studies. Random Forest implemented for pathway analysis (Pang et al., 2006), can be used to predict fruit/wine quality traits associated to terroir integrating multi-omic and phenotypic data as shown recently for potato (Acharjee et al., 2016).

Most of these multi-omics analysis approaches are pipelines that perform task sequences which share statistical methods (Bersanelli et al., 2016). Correlation analyses are the most common approaches performed to find relationships between the omics data. Simple correlation analyses, like Pearson or Spearman correlation, are widely used for multi-omics data integration (Rajasundaram et al., 2014; Rajasundaram and Selbig, 2016). Partial least square/projections to latent structures (PLS) and its extension, orthogonal partial least square (OPLS) (Tobias, 1995) have also been used for data integration from multi-omics results. Even though their predictive power is similar, OPLS results are much easier to interpret and outliers are quickly detected. OPLS can be used as a discriminate analysis (OPLS-DA), to identify differences between the overall data properties while removing systematic variation (Kirwan et al., 2012). However, these methods provided little insight when they are used in complex biological systems (highly multicollinear systems) (Wanichthanarak et al., 2015).

Modifications to these methods have been implemented to facilitate the interpretation of the data, for example, sparse PLS (sPLS) (Chun and Keleş, 2010) can better predict phenotypes through multi-omics data integration than previous methods (Rajasundaram et al., 2014). Orthogonal 2PLS (O2PLS), capable of dealing with unrelated systematic variation between datasets (Bouhaddani et al., 2016), has been successfully used for data integration of transcriptomics and metabolomics results from aspen under different light treatments (Bylesjö et al., 2007). Srivastava et al. (2013) used orthogonal projections to latent structures (OnPLS), an extension of O2PLS, to integrate transcriptomics, proteomics, and metabolomics data to construct a model that could identify biological relevant events in the oxidative stress response in poplar.

DATA INTEGRATION IN *V. vinifera*

In plant science, most of data integration of omics results comes from model plants; however, there is an increase in publications on multi-omics data integration in *V. vinifera*.

One of the first publications in multi-omics data integration in *V. vinifera* was the work of Zamboni et al. (2010). Integrating transcriptome, proteome and metabolome data, they identified stage specific biomarkers for berry development. Data integration was performed using two strategies, one hypothesis driven (i.e., a hypothesis was tested) and the other hypothesis free (i.e., discovery driven), in both cases principal component analysis (PCA), O2PLS and O2PLS-DA were used.

Using five different omics technologies and correlation analysis (PCA and Pearson correlation) together with biochemical pathway analysis (KEGG, PlantCyC and VitisCyC), Ghan et al. (2015) could differentiate biochemical characteristics from five different cultivars. Moreover, Anesi et al. (2015) studied the *terroir* effect in *V. vinifera* cultivar Corvina in seven different sites over a 3 years period using metabolome and transcriptome data. Using correlation analyses (PCA, PLS-DA and O2PLS-DA) they could identify a *terroir* signature in the berry metabolome composition for each growing site. Network analyses have been recently adopted to integrate grapevine multi omics results (Wong and Matus, 2017). For example, Palumbo et al. (2014) using network-based methods, identified "fight-club" nodes (genes with negatively correlated profiles) that may be relevant for the control of berry transition between development and ripening.

There are also online resources available that can help analyze omic data from *V. vinifera*. For example, VitisNet (Grimplet et al., 2009, 2012) offers manually annotated molecular networks (16,000 genes and 247 networks) where omics data can be loaded to visualize changes in the transcriptome, proteome and metabolome for a given experiment. VTCdb (Wong et al., 2013) is a gene co-expression database for *V. vinifera* that allows exploring transcription regulation. With more than 29,000 genes (95% of the predicted grapevine transcriptome) to query co-expression networks, VTCdb offers the possibility to analyze the transcriptional network of grapevine development, metabolism and stress response. VitisCyc (Naithani et al., 2014) is a grapevine metabolic pathway database that also allows omics data to be uploaded (transcriptome, proteome and metabolome) and to analyze changes in metabolic networks in each experiment. VESPUCCI (Moretto et al., 2016) is a manually annotated gene expression compendium exploratory tool that can be used to investigate grapevine's gene expression patterns.

PHENOTYPIC PLASTICITY THROUGH EPIGENETIC MODIFICATIONS

Epigenetics is the study of heritable phenotypes that occur through modifications that alter DNA activity without modifying its basic nucleotide structure (Feil and Fraga, 2012). Many epigenetic mechanisms, acting in an interactive and redundant fashion (Grant-Downton and Dickinson, 2005; Berger et al., 2009), have been described to date, with DNA methylation probably being the best-studied of all (Rapp and Wendel, 2005). DNA methylation affects chromatin condensation in a rapid and reversible manner (Grativol et al., 2012). In turn, the regional level of chromatin condensation affects the transcriptional state of nearby genomic features such as genes and transposable elements (Zhang et al., 2006). Global changes in DNA methylation associated to local environments can be analyze using a myriad of methods (Kurdyukov and Bullock, 2016). Bisulfite modification of genomic DNA combined with whole genome sequencing (BS-Seq) is the gold standard for methylation analysis because it can assess an entire methylome with single base resolution (Krueger et al., 2012). However, due to their lower cost, other approaches such as next generation sequencing following the capture of the methylated fraction of the genome or its fragmentation using methylation sensitive restriction enzymes (Bock et al., 2010; Li et al., 2010; Kitimu et al., 2015) are better suited to study large number of samples. Both generate quantitative and qualitative information of the methylation status of a reduced but significant representation of the total genome.

Environmental signals are one of the elements that can have a major effect in modifying the DNA methylation patterns leading to gene expression changes that ultimately affect the plant phenotype (Feil and Fraga, 2012). The idea that the environment could modify the epigenetic status, and these modifications passed to the offspring (Tricker et al., 2013) or maintained as epigenetic memory on long lived organisms (Latzel et al., 2016), has attracted attention from scientists studying mechanisms involved in adaptation to local environments (Consuegra and Rodríguez López, 2016) and how these could be used to enhance crop performance (Rodríguez López and Wilkinson, 2015). There are many reported examples of how the environment affects the epigenome in natural environments and how epigenetic variations in plant populations could help to overcome the lack of genetic diversity (Fonseca Lira-Medeiros et al., 2010; Verhoeven et al., 2010).

One of the most well-known examples in which the environment affects the phenotype through epigenetic modifications is vernalization (Feil and Fraga, 2012). Through this process, plants in temperate regions mitigate the deleterious effects of low winter temperatures on flower and fruit development by breaking dormancy only after having been exposed to a cold period (Kumar et al., 2016). Unusual environmental conditions during dormancy such as high winter temperatures have been shown to exert a negative effect on fruit quality and yield on perennial crops requiring a vernalization period (Sugiura et al., 2012). Recent work in apple shows how methylation and expression levels of key genes involved in flowering and fruit set are modified by the level of chill received during bud dormancy (Kumar et al., 2016), indicating that the environmentally induced changes observed in fruit quality could be regulated by DNA methylation.

Together these studies suggest that the environment can have a long lasting phenotypic effect in plants through epigenetic changes without the need for genetic variation, and that epigenetic mechanisms could be working as intermediaries between environmental variation and the plant genome, and in this way, potentially contributing to plant phenotypic plasticity. Moreover, this mechanism could give plant populations a way of adapting to the local growing conditions (Platt et al., 2015; González et al., 2016). However, to our knowledge, almost

all epigenetic studies done in *V. vinifera* have focused on the identification of commercial clones (Imazio et al., 2002; Schellenbaum et al., 2008; Ocaña et al., 2013) and on the assessment of *in vitro* culture on the epigenome (Baránek et al., 2015), there is, therefore, a lack of information of how the environment affects a grapevine's epigenome and to what extent this interaction affects fruit quality. Until now, there are no studies looking at the epigenome to understand the control of gene expression in *V. vinifera* and how environmental signals can change the regulation of metabolic pathways through epigenetic modifications. In our view, the inclusion of epigenomic data on the analysis of the *terroir* effect will not only increase the resolution of analysis but will also help us to understand the regulatory mechanisms behind the observed differences.

CONCLUSION

There is no doubt that the elements affecting grapevine growth and fruit composition are complex and multifarious. While the concept of *terroir* is widely discussed, the underlying mechanisms remain somewhat enigmatic. However, with the recent parallel development of omics technologies and of statistical approaches for their integration, we are reaching a point where it may be possible to overcome this challenge. The geographic delimitation of a terroir is the first challenge to overcome before its molecular characterization. This delimitation could be achieved 1. Empirically, based in the number of significantly different environmental subregions present in the study or/and 2. based on the traditionally defined wine regions. Moreover, the masking effect that environmental inter-annual variations can have over single year measurements demands the incorporation of data from multiple seasons to be able to determine terroirs with enough confidence. Ideally such seasons should be, from a weather perspective, variable within the range characteristic for the region of study to be able to capture its "normal terroir."

Understanding how the genome, environment and viticulture practices interact to affect fruit quality will allow us the opportunity to implement agricultural practices aimed to obtain the desired fruit characteristics for every climate/cultivar combination (Jones and Davis, 2000), leading to more efficient use of resources and better management of vineyards. In addition, grape growers can maximize the *terroir* effect on the grapevine to highlight the uniqueness of their vineyards ultimately increasing their industrial competitiveness. We propose that the integration of multi-omic and environmental datasets will contribute to a better understanding of the drivers of the *terroir* effect in grapevine. Moreover, multiple dataset data integration will increase our understanding of the molecular mechanisms involved in the regulation of multifactorial genome by environment interactions. Finally, it is increasingly recognized that plants are involved in complex interactions their soil and epiphytic microbiomes, which can affect their phenotype (Mueller and Sachs, 2015). The 'omics' era gives us the ability to explore the nature and consequences of biotic/abiotic interactions and so, a future challenge will be to bring the concept of the holobiont (the plant host plus its microbiomes) into the analysis of terroir and its effect on grapevine growth and fruit composition.

AUTHOR CONTRIBUTIONS

PF, TC, CC, and CR designed and wrote the manuscript.

ACKNOWLEDGMENTS

PF was supported by Australian Grape and Wine Authority (AGWA) and University of Adelaide Graduate Research Scholarships. CR is supported by a University of Adelaide Research Fellowship. This study was funded through a Pilot Program in Genomic applications in Agriculture and Environment Sectors jointly supported by the University of Adelaide and the Australian Genome Research Facility Ltd.

REFERENCES

Acharjee, A., Kloosterman, B., Visser, R. G. F., and Maliepaard, C. (2016). Integration of multi-omics data for prediction of phenotypic traits using random forest. *BMC Bioinformatics* 17(Suppl. 5), 180. doi: 10.1186/s12859-016-1043-4

Aho, K. A. (2013). *Foundational and Applied Statistics for Biologists Using R. Illustrated*. Boca Raton, FL: CRC Press.

Anesi, A., Stocchero, M., Dal Santo, S., Commisso, M., Zenoni, S., Ceoldo, S., et al. (2015). Towards a scientific interpretation of the terroir concept: plasticity of the grape berry metabolome. *BMC Plant Biol.* 15, 191. doi: 10.1186/s12870-015-0584-4

Atanassov, I., Hvarleva, T., Rusanov, K., Tsvetkov, I., and Atanassov, A. (2009). Wine metabolite profiling: possible application in winemaking and grapevine breading in Bulgaria. *Biotechnol. Biotechnol. Equip.* 23, 1449–1452. doi: 10.2478/V10133-009-0011-9

Baránek, M., Èechová, J., Raddová, J., Holleinová, V., Ondrušíková, E., and Pidra, M. (2015). Dynamics and reversibility of the DNA methylation landscape of grapevine plants (*Vitis vinifera*) stressed by *in vitro* cultivation and thermotherapy. *PLoS ONE* 10:e0126638. doi: 10.1371/journal.pone.0126638

Berger, S. L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. *Genes Dev.* 23, 781–783. doi: 10.1101/gad.1787609

Bersanelli, M., Mosca, E., Remondini, D., Giampieri, E., Sala, C., Castellani, G., et al. (2016). Methods for the integration of multi-omics data: mathematical aspects. *BMC Bioinformatics* 17(Suppl. 2), S15. doi: 10.1186/s12859-015-0857-9

Bock, C., Tomazou, E. M., Brinkman, A. B., Müller, F., Simmer, F., Gu, H., et al. (2010). Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat. Biotechnol.* 28, 1106–1114. doi: 10.1038/nbt.1681

Bokulich, N. A., Collins, T. S., Masarweh, C., Allen, G., Heymann, H., Ebeler, S. E., et al. (2016). Associations among wine grape microbiome, metabolome, and fermentation behavior suggest microbial contribution to regional wine characteristics. *MBio* 7:e00631-16. doi: 10.1128/mBio.00631-16

Bouhaddani, S. E., Houwing-Duistermaat, J., Salo, P., Perola, M., Jongbloed, G., and Uh, H.-W. (2016). Evaluation of O2PLS in Omics data integration. *BMC Bioinformatics* 17(Suppl. 2), S11. doi: 10.1186/s12859-015-0854-z

Burns, K. N., Kluepfel, D. A., Strauss, S. L., Bokulich, N. A., Cantu, D., and Steenwerth, K. L. (2015). Vineyard soil bacterial diversity and composition revealed by 16S rRNA genes: differentiation by geographic features. *Soil Biol. Biochem.* 91, 232–247. doi: 10.1016/j.soilbio.2015.09.002

Bylesjö, M., Eriksson, D., Kusano, M., Moritz, T., and Trygg, J. (2007). Data integration in plant biology: the O2PLS method for combined modeling of transcript and metabolite data. *Plant J.* 52, 1181–1191. doi: 10.1111/j.1365-313X.2007.03293.x

Cevallos-Cevallos, J. M., Reyes-De-Corcuera, J. I., Etxeberria, E., Danyluk, M. D., and Rodrick, G. E. (2009). Metabolomic analysis in food science: a review. *Trends Food Sci. Technol.* 20, 557–566. doi: 10.1016/j.tifs.2009.07.002

Cheng, G., He, Y.-N., Yue, T.-X., Wang, J., and Zhang, Z.-W. (2014). Effects of climatic conditions and soil properties on Cabernet Sauvignon berry growth and anthocyanin profiles. *Molecules* 19, 13683–13703. doi: 10.3390/molecules190913683

Chun, H., and Keleş, S. (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* 72, 3–25. doi: 10.1111/j.1467-9868.2009.00723.x

Conde, C., Silva, P., Fontes, N., Dias, A. C. P., Tavares, R. M., Sousa, M. J., et al. (2007). Biochemical changes throughout grape berry development and fruit and wine quality. *Food* 1, 1–22. doi: 10.1186/s12864-016-2660-z

Consuegra, S., and Rodríguez López, C. M. (2016). Epigenetic-induced alterations in sex-ratios in response to climate change: an epigenetic trap? *Bioessays* 38, 950–958. doi: 10.1002/bies.201600058

Cozzolino, D. (2016). Metabolomics in grape and wine: definition, current status and future prospects. *Food Anal. Method* 9, 2986–2997. doi: 10.1007/s12161-016-0502-x

Cutler, S. R., Rodriguez, P. L., Finkelstein, R. R., and Abrams, S. R. (2010). Abscisic acid: emergence of a core signaling network. *Annu. Rev. Plant Biol.* 61, 651–679. doi: 10.1146/annurev-arplant-042809-112122

Dal Santo, S., Fasoli, M., Negri, S., D'Incà, E., Vicenzi, N., Guzzo, F., et al. (2016). Plasticity of the berry ripening program in a white grape variety. *Front. Plant Sci.* 7:970. doi: 10.3389/fpls.2016.00970

Dal Santo, S., Tornielli, G. B., Zenoni, S., Fasoli, M., Farina, L., Anesi, A., et al. (2013). The plasticity of the grapevine berry transcriptome. *Genome Biol.* 14:r54. doi: 10.1186/gb-2013-14-6-r54

FAO (2012). *FAOSTAT*. Available at: http://faostat.fao.org/site/339/default.aspx [accessed August 3, 2016].

Feil, R., and Fraga, M. F. (2012). Epigenetics and the environment: emerging patterns and implications. *Nat. Rev. Genet.* 13, 97–109. doi: 10.1038/nrg3142

Fonseca Lira-Medeiros, C., Parisod, C., Fernandes, R. A., Mata, C. S., Cardoso, M. A., and Ferreira, P. C. (2010). Epigenetic variation in mangrove plants occurring in contrasting natural environment. *PLoS ONE* 5:e10326. doi: 10.1371/journal.pone.0010326

Fukushima, A., Kanaya, S., and Nishida, K. (2014). Integrated network analysis and effective tools in plant systems biology. *Front. Plant Sci.* 5:598. doi: 10.3389/fpls.2014.00598

Gambetta, J. M., Cozzolino, D., Bastian, S. E. P., and Jeffery, D. W. (2016). Towards the creation of a wine quality prediction index: correlation of Chardonnay juice and wine compositions from different regions and quality levels. *Food Anal. Method* 9, 2842–2855. doi: 10.1007/s12161-016-0467-9

Gambetta, J. M., Schmidtke, L. M., Wang, J., Cozzolino, D., Bastian, S. E. P., and Jeffery, D. W. (2017). Relating expert quality ratings of Australian chardonnay wines to volatile composition and production method. *Am. J. Enol. Vitic.* 68, 39–48. doi: 10.5344/ajev.2016.16058

Ghan, R., Van Sluyter, S. C., Hochberg, U., Degu, A., Hopper, D. W., Tillet, R. L., et al. (2015). Five omic technologies are concordant in differentiating the biochemical characteristics of the berries of five grapevine (*Vitis vinifera* L.) cultivars. *BMC Genomics* 16, 946. doi: 10.1186/s12864-015-2115-y

Gladstones, J. (2011). "Introduction and definition of terroir," in *Wine, Terroir and Climate Change*, ed. M. Deves (Kent Town, SA: Wakefield Press), 1–4.

Gligorijević, V., and Pržulj, N. (2015). Methods for biological data integration: perspectives and challenges. *J. R. Soc. Interface* 12:112. doi: 10.1098/rsif.2015.0571

González, A. P., Chrtek, J., Dobrev, P. I., Dumalasová, V., Fehrer, J., Mráz, P., et al. (2016). Stress-induced memory alters growth of clonal offspring of white clover (*Trifolium repens*). *Am. J. Bot.* 103, 1567–1574. doi: 10.3732/ajb.1500526

Grant-Downton, R. T., and Dickinson, H. G. (2005). Epigenetics and its implications for plant biology. 1. The epigenetic network in plants. *Ann. Bot.* 96, 1143–1164. doi: 10.1093/aob/mci273

Grativol, C., Hemerly, A. S., and Ferreira, P. C. G. (2012). Genetic and epigenetic regulation of stress responses in natural plant populations. *Biochim. Biophys. Acta* 1819, 176–185. doi: 10.1016/j.bbagrm.2011.08.010

Grimplet, J., Cramer, G. R., Dickerson, J. A., Mathiason, K., Van Hemert, J., and Fennell, A. Y. (2009). VitisNet: "Omics" integration through grapevine molecular networks. *PLoS ONE* 4:e8365. doi: 10.1371/journal.pone.0008365

Grimplet, J., Van Hemert, J., Carbonell-Bejerano, P., Díaz-Riquelme, J., Dickerson, J., Fennell, A., et al. (2012). Comparative analysis of grapevine whole-genome gene predictions, functional annotation, categorization and integration of the predicted gene sequences. *BMC Res. Notes* 5:213. doi: 10.1186/1756-0500-5-213

Hale, W. G., Saunders, V. A., and Margham, J. P. (2005). *Collins Dictionary of Biology*, 3rd Edn. London: Collins.

Hong, Y.-S. (2011). NMR-based metabolomics in wine science. *Magn. Reson. Chem.* 49(Suppl. 1), S13–S21. doi: 10.1002/mrc.2832

Ibrahim, K. S., and Kumar, N. S. (2017). "Methods for microbiome analysis," in *Translational Bioinformatics and Its Application*, eds D.-Q. Wei, Y. Ma, W. C. S. Cho, Q. Xu, and F. Zhou (Dordrecht: Springer), 269–298. doi: 10.1007/978-94-024-1045-7_12

Imazio, S., Labra, M., Grassi, F., Winfield, M., Bardini, M., and Scienza, A. (2002). Molecular tools for clone identification: the case of the grapevine cultivar 'Traminer.' *Plant Breed.* 121, 531–535. doi: 10.1046/j.1439-0523.2002.00762.x

Jones, G. V., and Davis, R. E. (2000). Climate influences on grapevine phenology, grape composition, and wine production and quality for Bordeaux, France. *Am. J. Enol. Vitic.* 51, 249–261.

Kirwan, G. M., Johansson, E., Kleemann, R., Verheij, E. R., Wheelock, ÅM., Goto, S., et al. (2012). Building multivariate systems biology models. *Anal. Chem.* 84, 7064–7071. doi: 10.1021/ac301269r

Kitimu, S. R., Taylor, J., March, T. J., Tairo, F., Wilkinson, M. J., and Rodríguez López, C. M. (2015). Meristem micropropagation of cassava (*Manihot esculenta*) evokes genome-wide changes in DNA methylation. *Front. Plant Sci.* 6:590. doi: 10.3389/fpls.2015.00590

Koyama, K., Sadamatsu, K., and Goto-Yamamoto, N. (2010). Abscisic acid stimulated ripening and gene expression in berry skins of the Cabernet Sauvignon grape. *Funct. Integr. Genomics* 10, 367–381. doi: 10.1007/s10142-009-0145-8

Krueger, F., Kreck, B., Franke, A., and Andrews, S. R. (2012). DNA methylome analysis using short bisulfite sequencing data. *Nat. Methods* 9, 145–151. doi: 10.1038/nmeth.1828

Kumar, G., Rattan, U. K., and Singh, A. K. (2016). Chilling-Mediated DNA methylation changes during dormancy and its release reveal the importance of epigenetic regulation during winter dormancy in apple (*Malus* × *domestica* Borkh.). *PLoS ONE* 11:e0149934. doi: 10.1371/journal.pone.0149934

Kurdyukov, S., and Bullock, M. (2016). DNA methylation analysis: choosing the right method. *Biology* 5:e3. doi: 10.3390/biology5010003

Latzel, V., Rendina González, A. P., and Rosenthal, J. (2016). Epigenetic memory as a basis for intelligent behavior in clonal plants. *Front. Plant Sci.* 7:1354. doi: 10.3389/fpls.2016.01354

Li, N., Ye, M., Li, Y., Yan, Z., Butcher, L. M., Sun, J., et al. (2010). Whole genome DNA methylation analysis based on high throughput sequencing technology. *Methods* 52, 203–212. doi: 10.1016/j.ymeth.2010.04.009

Meng, C., Kuster, B., Culhane, A. C., and Gholami, A. M. (2014). A multivariate approach to the integration of multi-omics datasets. *BMC Bioinformatics* 15:162. doi: 10.1186/1471-2105-15-162

Mizrachi, E., Verbeke, L., Christie, N., Fierro, A. C., Mansfield, S. D., Davis, M. F., et al. (2017). Network-based integration of systems genetics data reveals pathways associated with lignocellulosic biomass accumulation and processing. *Proc. Natl. Acad. Sci. U.S.A.* 114, 1195–1200. doi: 10.1073/pnas.1620119114

Moretto, M., Sonego, P., Pilati, S., Malacarne, G., Costantini, L., Grzeskowiak, L., et al. (2016). VESPUCCI: exploring patterns of gene expression in grapevine. *Front. Plant Sci.* 7:633. doi: 10.3389/fpls.2016.00633

Mueller, U. G., and Sachs, J. L. (2015). Engineering microbiomes to improve plant and animal health. *Trends Microbiol.* 23, 606–617. doi: 10.1016/j.tim.2015.07.009

Mulas, G., Galaffu, M. G., Pretti, L., Nieddu, G., Mercenaro, L., Tonelli, R., et al. (2011). NMR analysis of seven selections of Vermentino grape berry: metabolites composition and development. *J. Agric. Food Chem.* 59, 793–802. doi: 10.1021/jf103285f

Naithani, S., Raja, R., Waddell, E. N., Elser, J., Gouthu, S., Deluc, L. G., et al. (2014). VitisCyc: a metabolic pathway knowledgebase for grapevine (*Vitis vinifera*). *Front. Plant Sci.* 5:644. doi: 10.3389/fpls.2014.00644

Ocaña, J., Walter, B., and Schellenbaum, P. (2013). Stable MSAP markers for the distinction of *Vitis vinifera* cv Pinot noir clones. *Mol. Biotechnol.* 55, 236–248. doi: 10.1007/s12033-013-9675-3

Paim Pinto, D. L., Brancadoro, L., Dal Santo, S., De Lorenzis, G., Pezzotti, M., Meyers, B. C., et al. (2016). The influence of genotype and environment on small RNA profiles in grapevine berry. *Front. Plant Sci.* 7:1459. doi: 10.3389/fpls.2016.01459

Palumbo, M. C., Zenoni, S., Fasoli, M., Massonnet, M., Farina, L., Castiglione, F., et al. (2014). Integrated network analysis identifies fight-club nodes as a class of hubs encompassing key putative switch genes that induce major transcriptome reprogramming during grapevine development. *Plant Cell* 26, 4617–4635. doi: 10.1105/tpc.114.133710

Pang, H., Lin, A., Holford, M., Enerson, B. E., Lu, B., Lawton, M. P., et al. (2006). Pathway analysis using random forests classification and regression. *Bioinformatics* 22, 2028–2036. doi: 10.1093/bioinformatics/btl344

Panighel, A., De Rosso, M., Dalla Vedova, A., and Flamini, R. (2015). Putative identification of new p-coumaroyl glycoside flavonoids in grape by ultra-high performance liquid chromatography/high-resolution mass spectrometry. *Rapid Commun. Mass Spectrom.* 29, 357–366. doi: 10.1002/rcm.7115

Pereira, G. E., Gaudillere, J. P., van Leeuwen, C., Hilbert, G., Maucourt, M., Deborde, C., et al. (2006). 1H NMR metabolite fingerprints of grape berry: comparison of vintage and soil effects in Bordeaux grapevine growing areas. *Anal. Chim. Acta* 563, 346–352. doi: 10.1016/j.aca.2005.11.007

Pinto, C., Pinho, D., Sousa, S., Pinheiro, M., Egas, C., and Gomes, A. C. (2014). Unravelling the diversity of grapevine microbiome. *PLoS ONE* 9:e85622. doi: 10.1371/journal.pone.0085622

Platt, A., Gugger, P. F., Pellegrini, M., and Sork, V. L. (2015). Genome-wide signature of local adaptation linked to variable CpG methylation in oak populations. *Mol. Ecol.* 24, 3823–3830. doi: 10.1111/mec.13230

Qi, Y., Suhail, Y., Lin, Y.-Y., Boeke, J. D., and Bader, J. S. (2008). Finding friends and enemies in an enemies-only network: a graph diffusion kernel for predicting novel genetic interactions and co-complex membership from yeast genetic interactions. *Genome Res.* 18, 1991–2004. doi: 10.1101/gr.077693.108

Rajasundaram, D., Runavot, J.-L., Guo, X., Willats, W. G. T., Meulewaeter, F., and Selbig, J. (2014). Understanding the relationship between cotton fiber properties and non-cellulosic cell wall polysaccharides. *PLoS ONE* 9:e112168. doi: 10.1371/journal.pone.0112168

Rajasundaram, D., and Selbig, J. (2016). More effort - more results: recent advances in integrative "omics" data analysis. *Curr. Opin. Plant Biol.* 30, 57–61. doi: 10.1016/j.pbi.2015.12.010

Rapp, R. A., and Wendel, J. F. (2005). Epigenetics and plant evolution. *New Phytol.* 168, 81–91. doi: 10.1111/j.1469-8137.2005.01491.x

Ritchie, M. D., Holzinger, E. R., Li, R., Pendergrass, S. A., and Kim, D. (2015). Methods of integrating data to uncover genotype-phenotype interactions. *Nat. Rev. Genet.* 16, 85–97. doi: 10.1038/nrg3868

Robinson, A. L., Adams, D. O., Boss, P. K., Heymann, H., Solomon, P. S., and Trengove, R. D. (2012). Influence of geographic origin on the sensory characteristics and wine composition of *Vitis vinifera* cv. Cabernet Sauvignon wines from Australia. *Am. J. Enol. Vitic.* 63, 467–476. doi: 10.5344/ajev.2012.12023

Rodríguez López, C. M., and Wilkinson, M. J. (2015). Epi-fingerprinting and epi-interventions for improved crop production and food quality. *Front. Plant Sci.* 6:397. doi: 10.3389/fpls.2015.00397

Roullier-Gall, C., Lucio, M., Noret, L., Schmitt-Kopplin, P., and Gougeon, R. D. (2014). How subtle is the "terroir" effect? Chemistry-related signatures of two "climats de Bourgogne". *PLoS ONE* 9:e97615. doi: 10.1371/journal.pone.0097615

Salvetti, E., Campanaro, S., Campedelli, I., Fracchetti, F., Gobbi, A., Tornielli, G. B., et al. (2016). Whole-Metagenome-Sequencing-Based community profiles of *Vitis vinifera* L. cv. Corvina berries withered in two post-harvest conditions. *Front. Microbiol.* 7:937. doi: 10.3389/fmicb.2016.00937

Schellenbaum, P., Mohler, V., Wenzel, G., and Walter, B. (2008). Variation in DNA methylation patterns of grapevine somaclones (*Vitis vinifera* L.). *BMC Plant Biol.* 8:78. doi: 10.1186/1471-2229-8-78

Smart, R. E., Robinson, J. B., Due, G. R., and Brien, C. J. (1985). Canopy microclimate modification for the cultivar Shiraz. I. Definition of canopy microclimate. *Vitis* 24, 17–24.

Son, H.-S., Hwang, G.-S., Kim, K. M., Ahn, H.-J., Park, W.-M., Van Den Berg, F., et al. (2009). Metabolomic studies on geographical grapes and their wines using 1H NMR analysis coupled with multivariate statistics. *J. Agric. Food Chem.* 57, 1481–1490. doi: 10.1021/jf803388w

Srivastava, V., Obudulu, O., Bygdell, J., Löfstedt, T., Rydén, P., Nilsson, R., et al. (2013). OnPLS integration of transcriptomic, proteomic and metabolomic data shows multi-level oxidative stress responses in the cambium of transgenic hipI-superoxide dismutase Populus plants. *BMC Genomics* 14:893. doi: 10.1186/1471-2164-14-893

Sugiura, T., Sumida, H., Yokoyama, S., and Ono, H. (2012). Overview of recent effects of global warming on agricultural production in Japan. *Jpn. Agr. Res. Q.* 46, 7–13. doi: 10.6090/jarq.46.7

Sun, R., He, F., Lan, Y., Xing, R., Liu, R., Pan, Q., et al. (2015). Transcriptome comparison of Cabernet Sauvignon grape berries from two regions with distinct climate. *J. Plant Physiol.* 178, 43–54. doi: 10.1016/j.jplph.2015.01.012

Tarr, P. T., Dreyer, M. L., Athanas, M., Shahgholi, M., Saarloos, K., and Second, T. P. (2013). A metabolomics based approach for understanding the influence of terroir in *Vitis Vinifera* L. *Metabolomics* 9, 170–177. doi: 10.1007/s11306-013-0497-x

Tobias, R. D. (1995). "An introduction to partial least squares regression," in *Proceedings of the Twentieth Annual SAS Users Group International Conference*, (Cary, NC: SAS Institute Inc.), 1250–1257.

Tricker, P. J., López, C. M. R., Gibbings, G., Hadley, P., and Wilkinson, M. J. (2013). Transgenerational, dynamic methylation of stomata genes in response to low relative humidity. *Int. J. Mol. Sci.* 14, 6674–6689. doi: 10.3390/ijms14046674

Van Leeuwen, C., and Seguin, G. (2006). The concept of terroir in viticulture. *J. Wine Res.* 17, 1–10. doi: 10.1080/09571260600633135

Verhoeven, K. J. F., Jansen, J. J., van Dijk, P. J., and Biere, A. (2010). Stress-induced DNA methylation changes and their heritability in asexual dandelions. *New Phytol.* 185, 1108–1118. doi: 10.1111/j.1469-8137.2009.03121.x

Wanichthanarak, K., Fahrmann, J. F., and Grapov, D. (2015). Genomic, proteomic, and metabolomic data integration strategies. *Biomark Insights* 10(Suppl. 4), 1–6. doi: 10.4137/BMI.S29511

Wong, D. C. J., and Matus, J. T. (2017). Constructing integrated networks for identifying new secondary metabolic pathway regulators in grapevine: recent applications and future opportunities. *Front. Plant Sci.* 8:505. doi: 10.3389/fpls.2017.00505

Wong, D. C. J., Sweetman, C., Drew, D. P., and Ford, C. M. (2013). VTCdb: a gene co-expression database for the crop species *Vitis vinifera* (grapevine). *BMC Genomics* 14:882. doi: 10.1186/1471-2164-14-882

Zamboni, A., Di Carli, M., Guzzo, F., Stocchero, M., Zenoni, S., Ferrarini, A., et al. (2010). Identification of putative stage-specific grapevine berry biomarkers and omics data integration into networks. *Plant Physiol.* 154, 1439–1459. doi: 10.1104/pp.110.160275

Zerihun, A., McClymont, L., Lanyon, D., Goodwin, I., and Gibberd, M. (2015). Deconvoluting effects of vine and soil properties on grape berry composition. *J. Sci. Food Agric.* 95, 193–203. doi: 10.1002/jsfa.6705

Zhang, A. (2009). "Data fusion in the analysis of protein interaction networks," in *Protein Interaction Networks: Computational Analysis* (New York, NY: Cambridge University Press), 243–257. Available at: http://assets.cambridge.org/97805218/88950/frontmatter/9780521888950_frontmatter.pdf

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W.-L., Chen, H., et al. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. *Cell* 126, 1189–1201. doi: 10.1016/j.cell.2006.08.003

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Fabres, Collins, Cavagnaro and Rodríguez López. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Chapter 3: Planting year and local growing environment modify DNA methylation and gene expression of grapevine (*Vitis vinifer*a cv. Shiraz)

Statement of Authorship

Title of Paper	**Local growing environment can modify DNA methylation and gene expression of grapevine (*Vitis vinifera* cv. Shiraz)**
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☒ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Pastor Jullian Fabres, Rakesh David, Konstantinos Bogias, Huahan Xie, Na Sai, Roberta De Bei, Matthew Gilliham, James Breen, Cassandra Collins, Penny Tricker and Carlos M. Rodriguez López This is an experimental manuscript about the environmental effect has in the gene expression and DNA methylation from grapevine leaf samples from 22 commercial vineyards covering six sub-regions of the Barossa Wine growing region, Australia. Findings of our study suggest that DNA methylation shows signed of local adaptation and vineyards altitude is a major contributor of the grapevine DNA methylation diversity.

Principal Author

Name of Principal Author (Candidate)	Pastor Jullian Fabres		
Contribution to the Paper	Data generation and analysis, interpretation and wrote manuscript.		
Overall percentage (%)	60%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	26 Nov 2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	*Rakesh David*		
Contribution to the Paper	Supervised development of the project, review and edited the manuscript. I hereby certify that the statement of the contribution is accurate.		
Signature		Date	26/11/2019

Name of Co-Author	Konstantinos Bogias
Contribution to the Paper	Development of the bioinformatics script to analyse RNA-seq and msGBS data, review and edited the manuscript. I hereby certify that the statement of the contribution is accurate.
Signature	Date 3/12/2019

Name of Co-Author	Huahan Xie
Contribution to the Paper	msGBS data generation, review and edited the manuscript. I hereby certify that the statement of the contribution is accurate.
Signature	Date 2/12/2019

Name of Co-Author	Na Sai
Contribution to the Paper	RNA-seq data generation for a subset of the vineyards samples, review and edited the manuscript. I hereby certify that the statement of the contribution is accurate.
Signature	Date 11-12-19

Name of Co-Author	Roberta De Bei
Contribution to the Paper	Conceived the experimental design, selected vineyards for sampling and performed sampling of the leaf tissue. Review and edited the manuscript. I hereby certify that the statement of the contribution is accurate.
Signature	Date 02/12/19

Name of Co-Author	Matthew Gilliham
Contribution to the Paper	Conceived the experimental design, review and edited the manuscript. I hereby certify that the statement of the contribution is accurate
Signature	Date 6/12/19

Name of Co-Author	James Breen
Contribution to the Paper	Conceived the experimental design, review and edited the manuscript. I hereby certify that the statement of the contribution is accurate
Signature	Date 04/12/2019

Name of Co-Author	Cassandra Collins
Contribution to the Paper	Conceived the experimental design, supervised development of the project and performed sampling of the leaf tissue. Review and edited the manuscript. I hereby certify that the statement of the contribution is accurate.
Signature	Date 6/12/19.

Name of Co-Author	Penny Tricker
Contribution to the Paper	Supervised development of the project, review and edited the manuscript. I hereby certify that the statement of the contribution is accurate.
Signature	Date 26/11/19.

Name of Co-Author	and Carlos M. Rodriguez López
Contribution to the Paper	Conceived and supervise the experimental design. Supervised development of the project. Reviewed and edited the manuscript and acted as the corresponding author. I hereby certify that the statement of the contribution is accurate.
Signature	Date 12/11/2019

Planting year and local growing environment modify DNA methylation and gene expression of grapevine (*Vitis vinifer*a cv. Shiraz)

Pastor Jullian Fabres[1], Konstantinos Bogias[1,3], Huahan Xie[1], Na Sai[1,2], Roberta De Bei[1], Matthew Gilliham[1,2], James Breen[3,4], Rakesh David[1,2], Penny Tricker[1] and Carlos M. Rodríguez López[1,5]

[1] School of Agriculture, Food and Wine, University of Adelaide, Adelaide, Australia

[2] ARC Centre of Excellence in Plant Energy Biology, University of Adelaide, Adelaide, Australia

[3] Bioinformatics Hub, School of Biological Sciences, University of Adelaide, Adelaide, SA, Australia.

[4] Robinson Research Institute, University of Adelaide, Adelaide, SA, Australia

[5] Environmental Epigenetics and Genetics Group, Department of Horticulture, College of Agriculture, Food and Environment, University of Kentucky, Lexington KY, U.S.A

Abstract

Background: Local environment has a strong effect on grapevine (*Vitis vinifera*) phenotype. Through changes in DNA methylation, plants can modulate their gene expression depending on environmental conditions. Studies in *A. thaliana* have shown that environmental signals can induce epigenetic modifications in the genome. Epigenetic mechanisms such as histone modifications and DNA methylation can modulate plant's gene expression depending on environmental conditions. However, little is known about how grapevine responds to the local environment through epigenetic modifications or the effect on gene expression in field grown plants. Here, we characterized the leaf methylome and transcriptome of 198 Shiraz plants cultivated across twenty-two vineyards, covering six sub-regions of the Barossa Wine Zone, Australia.

Results: Both global DNA methylation and gene expression clustered with the sub-region of origin of the samples. Genes that were differentially methylated and expressed were identified for all sub-regions. We did not find a correlation between changes in DNA methylation and changes in expression of individual genes. We did, however, identify genes in which variation in DNA methylation and expression was significantly associated with the measured environmental variables across the six sub-regions. Planting year and annual rainfall were the main drivers for modifications of gene expression and DNA methylation for grapevine cultivated in the Barossa Wine Zone.

Conclusion: We showed that DNA methylation and gene expression of Shiraz grapevine leaves reflected the environmental differences from the different sub-regions of the Barossa Wine Zone. Although both, DNA methylation and gene expression were sensitive to environmental factors, a direct relationship between individual gene methylation and expression was not observed probably due to the reduced representation approach used to identify differentially methylated loci.

Keywords: Gene expression, DNA methylation, grapevine, *Vitis vinifera*, regionality, planting year, environmental effect.

Background

Grapevine (*Vitis vinifera*) is one of the most economically important fruit crops with much of its production destined for wine making (Conde et al. 2007). In order to maintain desirable phenotypic attributes, most grapevine varieties are from cultivars that have been established for centuries through clonal propagation (Adam-Blondon et al. 2016) and elite grapevines cultivars are closely related (Laucou et al. 2018). Grapevine can be cultivated in a wide variety of environments, showing considerable phenotypic plasticity. It has been proposed that plants may

achieve phenotypic plasticity through epigenetic modifications of the genome without changes in the nucleotide sequence (Eichten et al. 2014; Giovannoni et al. 2017), through methylation of the DNA, modifications to histone proteins (i.e. methylation, acetylation, phosphorylation and ubiquitination) (Gallusci et al. 2016; Rodrigues and Zilberman 2015), and RNA interference (Pikaard and Mittelsten Scheid 2014). Epigenetic mechanisms have been shown to be involved in many plant processes, such as transposon silencing (Rigal et al. 2016), control of flowering time (Sheldon et al. 1999) and fruit development (Daccord et al. 2017). DNA methylation refers to the addition of a methyl group to the carbon 5 of cytosine [5-methylcytosine (5mC)] (Gallusci et al. 2016). In plants, DNA methylation occurs in the sequence context CG, CHG and CHH (H = A, T or C). There is increasing evidence showing that the environment can induce changes in DNA methylation of plants (Yaish et al. 2011; Kumar et al. 2016) and that these changes may regulate gene expression to adapt to local environment, as well as to overcome stressful growing conditions (Fonseca Lira-Medeiros et al. 2010; Raj et al. 2011). It has also been proposed that such changes could act as a molecular memory of the environment experienced (Tricker et al. 2013; Verhoeven et al. 2010).

As a perennial, woody plant with established genotypes, grapevine offers an opportunity to study epigenetic adaptation to the environment. The *Vitis vinifera* genome is diploid, highly heterozygous (This et al. 2006) and relatively small (~500 Mb) compared to other plants (e.g. ~950 Mb tomato, ~17 Gb wheat). In recent years there has been an increase in the number of epigenetic studies in *V. vinifera*. Studies focused of describing the crop's epigenetic diversity with methylation-sensitive amplified polymorphism (MSAP) analysis have been more successful in clonal identification than techniques that rely on genetic polymorphisms (Imazio et al. 2002; Ocaña et al. 2013). More recent studies have focused on the importance of DNA methylation for plant response to different environments. Magris et al. (2019) compared ten grapevine varieties

to analyse the effects of genetic and epigenetic polymorphisms on gene expression, revealing that gene body methylation was negative correlated with gene expression in grapevine berries. In another study, gene expression analysis in grape berries revealed that SET DOMAIN GROUP (SDG) proteins, a family of conserved epigenetic modulators, could be involved during grape berry development (Aquea et al. 2011). Additionally, environmental signals have been shown to influence the DNA methylation status of grapevine. Grapevine plants treated with UV radiation had modified DNA methylation status; this treatment also produced a change in the accumulation of polyphenols in the berries (Marfil et al. 2019). Recent studies have suggested that DNA methylation may partially explain the variation between genotypes in gene expression in different environments. Analysis of global patterns of DNA methylation revealed highly differentiated epigenetic profiles by area of origin, altitude of the vineyard and agronomical practices such as pruning system (Xie et al. 2017). Recently, Dal Santo et al. (2018) explored the interaction between genotype x environment in grapevine berries in two different cultivars, Sangiovese and Cabernet sauvignon, showing that plant genotype was an important factor in the variance in methylation between samples. However, their results suggested that DNA methylation remained stable regardless of variation in external cues or gene expression (Dal Santo et al. 2018).

Although, some evidence suggests that DNA methylation could be involved in the regulation of berry development and acclimation to the environment, our understanding is, so far, limited. The intertwined nature of genetic and epigenetic factors makes it intrinsically difficult to determine and isolate their specific contributions to plant adaptation to the environment. This is especially relevant in multi-genotype studies with a low number of replicates and locations. The current work constitutes a large-scale attempt to explore the effect of environmentally induced DNA methylation on gene expression in grapevine. The objective of this study was to characterise the leaf genome, methylome and transcriptome of vines (cv. Shiraz), planted in twenty-two

vineyards across the Barossa Wine Zone. We also aimed to identify the most significant variables influencing DNA methylation and gene expression of *V. vinifera*. Genetic and epigenetic analyses showed that the origin of the sample was reflected in DNA methylation status and gene expression, suggesting that grapevine plants adapt to the local environment through DNA methylation.

Results

Sub-region environmental differences

Fully expanded leaves from 198 vines were collected from twenty-two vineyards (nine samples per vineyard) located in six sub-regions from the Barossa Wine Zone, Australia, during the 2016 growing season. The numbers of vineyards per sub-regions were: four in Northern Grounds (NG), four in Central Grounds (CG), three in Southern Grounds (SG), four in Western Ridge (WR), four in Eastern Edge (EE) and three in Even Valley (EV). Vineyards details, soil properties and environmental data are reported in Additional file 1. Principal component analysis (PCA) was undertaken on the environmental data (Additional file 1: Table S1) to estimate key environmental measurements and their relationship with each sub-region (Figure 1). The first two principal components explain 60.39% of the variation in the dataset. PC1 explained 38.91% of the variation and separated Central Grounds and Southern Grounds from Eden Valley, mainly by measurements of temperature and elevation. PC2 explained 21.48% of the variation in the data and separated Western Ridge from Northern Grounds by overall rainfall in each region.

Genetic structure of Shiraz samples

In order to investigate the genetic and epigenetic bases of grapevine's adaptation to its growing environment, we used methylation sensitive genotyping by sequencing (msGBS) in the 198 leaf samples. Sequencing of the msGBS data delivered, in total, 484,776,478 paired-end (2x75) reads,

with an average of 2,486,366 reads per sample (Additional file 2: Table S1). On average, 23,135,111 reads per sample aligned to the *V. vinifera* reference genome (IGGP_12X.V1) obtained from the Ensembl Plants website (plants.ensembl.org/Vitis_vinifera/), with an average alignment rate of 63%. Samples with fewer than 100,000 mapped reads were discarded. This selection resulted in a DNA methylation data set for 179 grapevine samples (NG =32/36, SG = 25/27, CG = 35/36, EE = 34/36, EV = 20/27, WR= 33/36), which were used for downstream analyses.

A total of 8,139 genetic single nucleotide polymorphisms (SNPs) were identified from mapped msGBS sequencing reads using TASSEL3. Of these SNPs, only those present within loci sequenced in at least 80% of the samples were selected for further analysis (i.e. 4,893 SNPs). Genetic variability between samples inferred from the selected SNPs was visualized through PCA. This PCA revealed no significant genetic structure in the studied samples (Figure 2A). Descriptive analysis of principal components (DAPC) (Jombart et al. 2010) indicated that the most provable number of genetic clusters was 1 (Additional file 3), validating the lack of genetic structure observed in the PCA.

Effect of provenance on DNA methylation

We used the methylation sensitive restriction enzyme *MspI*, which cleaves at the recognitions site CCGG. After alignment and filtering (see Methods) 450,501 *MspI* sites were recognized across all 179 samples. DNA methylation was inferred via read coverage. The diversity in the DNA methylation was estimated through Principal Components-Linear Discriminant Analysis (PC-LDA) that revealed a clear clustering of samples by their sub-region of origin (Figure 2B). To assess the effect of the environment on DNA methylation, differentially methylated markers (DMMs) were determined through pairwise comparison between sub-regions (Figure 2C, Additional file 4: Table S1). The pairwise comparison Central Grounds vs Western Ridge

identified the highest number of DMMs (20,530), from with 3671 DMMs were exclusive on these two sub-regions. Pairwise comparison between Central Grounds vs Southern Grounds identified 18,780 DMMs, and 3147 exclusive DMMs. On the other hand, there were only two DMMs in the pairwise comparison between Eden Valley and Eastern Edge, followed by Southern Grounds and Western Ridge with six DMMs. A comparison of the median fold-change in methylation across all DMMs between sub-regions revealed that SG and WR presented higher methylation levels when compared to each of the other regions, while relative methylation levels were comparison dependent for the rest of the pair-wise region assessments (Figure 3).

In order to identify differentially methylated genes (DMGs), we selected genes that had one or more DMM within 5 kb upstream or downstream of the gene including the gene-body model (Additional file 4: Table S1 and S2). In total, samples from the sub-region Central Grounds had the largest number of DMGs (3,349). In contrast, grapevine samples from sub-region Eden valley had the lowest number of DMGs (250). In most cases (86%) DMGs comprised a single differentially methylated marker and the majority of these (70%) localized in the gene body (Additional file 4: Table S2). In total, we identified 87,790 DMMs between all sub-regions and from those, only 10% of the DMMs (8,924) were assigned to one fifth (3,985 of 18,952) of total genes expressed in the RNA-seq data.

The largest number of DMGs (1,604) for the pairwise comparison was between Central Grounds and Southern Grounds, whereas there were no DMGs between Southern Grounds and Western Ridge and only one gene was differentially methylated between Eden Valley and Eastern Edge.

Effect of provenance on gene expression

The transcriptomes of 198 leaf samples collected at budburst from the same vineyards were sequenced using next-generation sequencing (RNA-seq) to identify differentially expressed genes. Independently extracted total RNA samples were combined in equimolar proportion to

create 66 RNA pools (three pools per vineyard; each pool contained RNA from three plants). A total of 1,526,917,359 paired end reads (2x75 bp) were generated, with an average of 23,135,111 paired-end reads per pool (Additional file 2: Table S2). Sequencing reads were aligned against the *V. vinifera* reference genome (IGGP_12X.V1) with an average mapping percentage of 90%. Differences in global gene expression were analyzed using PC-LDA of the RNA-seq data. Similar to the methylome data, this analysis also showed clustering of the grapevine samples according to the sub-region of origin (Figure 4A). Pairwise comparison of the gene expression revealed the differentially expressed genes (DEGs) between sub-regions (Figure 4B, Figure 5, Additional file 4: Table S1). Central Grounds versus Northern Ground presented the highest number of DEGs (8,405) from which 678 genes were unique for this pairwise comparison, while Northern Ground versus Southern Grounds had the lowest number of DEGs (380).

Relationship between DNA methylation and gene expression

To assess if there was a relationship between DNA methylation and gene expression, we selected genes that were both differentially methylated (DMGs) and differentially expressed (DEGs) and called them differentially methylated and expressed genes (DMEGs). We analyzed the data from the DNA methylation and gene expression of the DMEGs identified in each pairwise comparison between sub-regions. To assess if the genomic location of DMM had an effect on the direction of gene expression, we separated the methylation data depending on the position of the DMM in the differentially expressed gene (gene body, 5 kb upstream of the gene, or 5 kb downstream of the gene) (Figure 6). There was no clear relationship between DNA methylation and gene expression in any combination between sub-regions, neither in the genomic position of the methylation marker (upstream, downstream or the gene body) or in the direction of gene expression.

Identification of environmental variables that affect DNA methylation and gene expression

We used a generalized linear model (GLM) analysis to test whether there was an association of DNA methylation and gene expression with the following variables: altitude (meters above the sea level [m a.s.l.]), growing season mean temperature, mean annual rainfall (mm), mean season rainfall (mm), and planting year. In addition, we tested for soil measurements such as plant available phosphorus (µg P/g soil), salinity (g/l) and pH. We consider the normalized read counts (logCPM) as a dependent variable and the environmental and soil data as the independent variables. All variables tested showed genes with significant association (FDR-adjusted p-value < 0.05) with gene expression or DNA methylation (Table 1). The variables with the most significant effect on gene expression and DNA methylation were planting year (97 genes), followed by mean annual rainfall (16 genes), soil pH (5 genes) and mean season rainfall (2 genes) (Table 1, Additional file 5: Tables S1-S4). Singular enrichment analysis (SEA) performed in AgriGO (Du et al. 2010; Tian et al. 2017) identified 42 enriched gene ontology (GO) terms (31 Molecular functions and 17 Biological processes) for planting year (Additional file 6: Table S1). The molecular function for the enriched GO terms in planting year were zinc ion binding, DNA polymerase activity and aspartic-type endopeptidase activity (Additional file 7).

Vineyard altitude and soil phosphorus content were significantly associated to changes in expression for a large number of genes (3,914 and 913 DEGs respectively), however, both variables seemed to have a relatively smaller effect on gene DNA methylation (4 and 12 DMGs respectively). To explore the change in gene expression as a particular environmental variable increased, we plotted gene expression through heatmaps by sorting the samples as planting year and annual rainfall increased. Although there was a significant association between planting year and annual rainfall with changes in the DNA methylation and gene expression (Additional file 8,

Figures S1-S2), graphical representation of the gene expression did not reveal a clear pattern as planting year (from 1893 to 2004) (Figure 7A) or annual rainfall (from 431.2 to 663.2 mm) (Figure 7B) increased between vineyards. To further explore this, we divided the genes that correlated with changes with planting year into four groups depending of the direction in gene expression (upregulated/downregulated) and DNA methylation (hypo-methylated/hyper-methylated) (Figure. 8). We observed a slight increase in gene expression as the elevation increased. However, there was no significant association between changes in gene expression and DNA methylation (Additional file 9).

Discussion

Local environment can significantly affect grapevine growth (van Leeuwen et al. 2004), gene expression (Anesi et al. 2015; Dal Santo et al. 2013), metabolite accumulation (Roullier-Gall et al. 2014; Son et al. 2009; Tarr et al. 2013) and wine composition (Robinson et al. 2012). Adaptation to the growing environment might be achieved through epigenetic modification of gene expression (Feil and Fraga 2012; Jaenisch and Bird 2003). In this work, we hypothesized that leaf DNA methylation status and gene expression could reflect the grapevine growing environment. We observed that both DNA methylation and gene expression of grapevine were affected by the local growing environment and this was not related to genetic variance. We identified genes that were both differentially methylated and expressed between twenty-two vineyards covering six sub-regions of the Barossa Wine Zone. We also identified that planting year and mean annual rainfall were the main modifiers of DNA methylation and gene expression in Shiraz grapevine leaves from the Barossa Wine Zone.

Regional effect on DNA methylation and gene expression

We performed msGBS analysis to induvial samples (198) and for RNA analysis we performed pooling of three samples (66 samples). This strategy was chose to obtain minimum of 2 million

reads per sample for the msGBS analysis and a minimum of 20 million RNAseq reads per pool. It has been shown that increasing the number of samples sequenced even at cost of lower depth coverage achieves higher power than sequencing a small number of samples at higher depth coverage (Kim et al. 2010). In addition, low-coverage (0.1-0.5x) sequencing captures as much variation across the genome as common (>5%) coverage (Pasaniuc et al. 2012). As correlation analysis between DNA methylation and RNA expression was performed at vineyard level, the pooling strategy used should not significantly affect the results, as previously seen by Futschik and Schlötterer (2010) where pooled samples were as effective in SNP discovery as sequencing individual samples.

PCA was performed to explore how the overall trends in the environmental data related to each sub-region (Figure 1). This showed that differences in temperature and rainfall between regions could be used to classify the different sub-regions. PCA of SNPs (Figure 2A) obtained from the msGBS data showed that there was no genetic structure with samples within sub-region and DNA methylation explained more of the variance than genetic polymorphisms. Previous studies on grapevine genetic diversity have shown that cultivated grapevines are closely related (Laucou et al. 2018) and consist of a single complex gene pool where much of its diversity has been reduced by artificial selection and the method of clonal propagation (Aradhya et al. 2003). The DNA methylation profile of the samples (Figure 2B) clustered clearly depending on the sub-region of origin. Together, these data showed that the DNA methylation profiles were similar between grapevines samples from the same sub-region of origin and that this diversity was not explained by genetic diversity. Gene expression profiles also clustered by sub-region, but the discrimination between sub-regions was greater for methylation than for gene expression profiles, suggesting that environmental conditions of each sub-region particularly influenced DNA methylation of grapevine leaves. Although a recent study analyzing the genome

methylation landscape of grape berries from two varieties at three locations over two years observed that DNA methylation was not influenced by environmental conditions (Dal Santo et al. 2018), our results were similar to those of Xie et al. (Xie et al. 2017), who analyzed just three sub-regions of the Barossa (NG, CG and SG).

Adaptation to local growing environment through epigenetic modifications has been observed in long-lived organisms. Raj et al. (2011) reported that transcript abundance and genome-wide DNA methylation profiles of clonally propagated trees from the genus *Populus* were based on differences in geographic origin of the clones. Differences in gene expression due to environmental factors have been observed previously in grape berries, where plastic transcripts were identified between vineyards with common viticulture practices and environmental conditions (Dal Santo et al. 2013). Transcriptomic analysis of cv. Cabernet Sauvignon berries grown in different regions also revealed differences in gene expression, particularly genes involved in ABA biosynthesis pathway (Sun et al. 2015). We identified two differentially methylated and expressed genes, members of the C2-doimain ABA-related (CAR) protein family. This protein family is involved in ABA signaling pathway and recruits ABA receptors to the membrane (Diaz et al. 2016). Interestingly, it has been shown that ABA can mediate the response of a plant under stress or during plant development through epigenetic mechanisms (Chinnusamy et al. 2008).

Pairwise comparison between sub-regions revealed that Central Grounds had the higher abundance of DMMs and DMGs. Central Grounds is characterized by higher temperatures during the growing season (19.2 °C) and lower mean annual rainfall (576.2 mm) than other sub-regions. Eden Valley presented the lowest number of DMMs and DMGs between all the sub-regions. Vineyards in the Eden Valley have lower temperatures during the growing season (17.4 °C) and higher annual rainfall (663.2 mm), suggesting that grapevines grown in warmer

conditions accumulated more differences in their DNA methylation with respect to other sub-regions. Pairwise comparison of DMMs also revealed a shift in the methylation status between sub-regions (Figure 4). Samples from the sub-regions Southern Grounds and Western Ridge were hypermethylated, whereas samples from sub-regions Northern Grounds and Central Grounds were comparatively hypo-methylated. Although growing season temperatures between these four sub-regions were similar (NG: 18.1°C; CG: 19.2°C; SG: 19.2°C and WR: 18.8°C), Northern Grounds and Central Grounds have lower mean annual rainfall (547.2 mm) than Southern Grounds and Western Ridge (630.0 mm). Together, these results suggested that temperature and rainfall might play a key role in the DNA methylation status of grapevine leaves. However, we observed that only a small number of DMMs (14) and genes (170) were associated with increase in temperature and either changes in DNA methylation or gene expression. Growing season mean temperature was associated with methylation but, within the range of temperature (1.8°C) sampled here, a linear pattern of positive or negative correlation was not clear. However, within the range of annual rainfall data (232 mm) for the sampled vineyards, we observed a significant association between increased rainfall and changes in DNA methylation (6140 DMMs and 4088 DMGs), and from these genes, 16 were both differentially methylated and expressed. It is important to notice that vineyards have different irrigation regimes, from no irrigation to 1.2 ML / ha (Additional file 1: Table S2), therefore, rainfall does not indicate the soil and grapevine water status; however, rainfall can be used as another pedoclimatic measurement that characterize each vineyard. Even though irrigation has been shown to significantly affect plant water status, yield, and grape and wine composition (Intrigliolo and Castel 2010, 2011; Trigo-Córdoba et al. 2015), vintage effect still exerts the greatest influence on chemical composition in grapevine (Bouzas-Cid et al. 2018; Ortega-Heras et al. 2014).

Although the effect of DNA methylation on gene expression has been studied extensively, it is still not completely understood (Zilberman 2017). Studies based on genomic location have proven that DNA methylation in the promoter region and gene body can significantly influence gene expression (Li et al. 2012; Zhang et al. 2006; Zilberman 2017). Usually, in plants, DNA methylation in the promoter regions leads to reduced expression (Zhang et al. 2018) and DNA methylation in the gene body leads to increases in gene expression (Zhang et al. 2018). However, a recent study analyzing the gene expression variation between ten grape varieties during berry development showed that gene body methylation negatively correlated with expression (Magris et al. 2019). The authors suggested that the negative correlation might have been due to the high number of long interspersed elements (LINEs) within introns of genes with high body methylation. In our study, we observed that 10% of all DMMs were within or in close proximity to a gene (differentially methylated genes). Additionally, we did not observe a correlation between changes in DNA methylation and changes in gene expression (Figure 6). Our results are in agreement with previous studies in *A. thaliana*, showing that gene methylation rarely corresponds to changes in gene expression. In one study, only 6% of methylated genes were differentially expressed after loss of DNA methylation (Vaughn et al. 2007). Similarly, in another study 5% of methylated genes were differentially expressed in *A. thaliana met1-3* mutants that almost completely lack of genic methylation (Zhang et al. 2006). The environment might still affect the gene expression through DNA methylation, but as suggested by Vaughn et al. (2007) methylation can influence gene expression in only a few cells, so the difference between sub-regions might be too small to detect using a genome-wide approach.

Unravelling the Barossa environment and its consequences for DNA methylation and gene expression

Using molecular genetic techniques, our results confirmed the classification of the different sub-regions of the Barossa Wine Zone that had been proposed by the wine industry of the Barossa (Barossa Grape and Wine Association), from which different wines with particular characteristics can be made. Sample origin had a clear effect on DNA methylation and gene expression. We observed a clear difference between sub-regions even when they were in close proximity. We identified which of the complex factors of *terroir* most affected gene expression and DNA methylation. This suggests that *terroir* can be explored through 'omics analysis (Anesi et al. 2015; Dal Santo et al. 2016; Roullier-Gall et al. 2014). To identify genes responding to a changing environmental variable we fitted a generalized linear model to the transcriptome and methylome data. Planting year and annual rainfall were identified as the environmental variables that significantly associated to changes in gene expression and DNA methylation. As grapevines are vegetatively propagated, it is difficult to establish when the ancestral seedling first germinated to establish true biological age. Therefore, planting date is used to estimate the age of grapevine (Grigg et al. 2017). Aging has been shown to influence the methylome of plants. In the annual plant *A. thaliana*, Ruiz-García et al. (2005) observed an increase in DNA methylation during plant development through amplified fragment length polymorphisms (AFLP), whereas Ogneva et al. (2016) using methylation-sensitive DNA fragmentation assay observed a reduction in DNA methylation in *A. thaliana* with aging. In perennial species such as bamboo (*Phyllostachys heterocycla*) (Yuan et al. 2014) and in long-lived trees like *Sequoiadendron giganteum* (Monteuuis et al. 2008) and *Acacia mangium* (Monteuuis et al. 2009) age can also affect their DNA methylation status. In grapevine, plant age has been shown to have an impact on both fruit yield and wine quality (Grigg 2017; Reynolds et al. 2008) and DNA methylation

analysis between older and younger grapevines has revealed differences in the global methylation levels between grapevines of different ages (Grigg 2017).

We identified that vineyard altitude influenced the gene expression of grapevine. The effect of altitude on grapevine can be divided into a combination of other environmental factors such as rainfall, temperature and sunlight radiation; at higher altitudes, there is a decrease in temperature, an increase in rainfall and an increase in sunlight intensity. Sunlight and temperature have a similar effect on grapevine. Both yield and soluble solids will increase with radiation, either more intensity or duration (Gladstones 2011) until reaching maximum photosynthesis rate at 700 E^{m-2} ^{s1} (Jackson and Lombard 1993). As a component of sunlight, vines are also exposed to ultraviolet (UV) radiation. Transcriptomic analysis of grape berry skin exposed to UV radiation in mid-altitude vineyards revealed that expression of secondary metabolism transcripts was influenced by UV radiation (Carbonell-Bejerano et al. 2014). Interestingly, we identified ten differentially expressed genes that correlated with altitude that were reported to be affected by UV light in grapevine at high altitudes (Carbonell-Bejerano et al. 2014) (Additional file 5: Table S5), where five of the genes are involved in the primary and secondary metabolism.

Thus, the local growing conditions can affect the epigenetic status of grapevine, priming it for adaptation to the local environment. Plants could use epigenetic modification as mechanism to record previous experience and even to pass the memory to their progeny. Several reports have suggested that the progeny of previously exposed plants have an advantage while growing in the same environment as their parents (Agrawal et al. 1999; Boyko and Kovalchuk 2010; Sultan 2000). This is known as transgenerational memory (Lämke and Bäurle 2017) and it could be used potentially in an agronomic context by priming the parental plants in order to obtain better adapted plants for local environments.

Conclusions

The local environment affected both DNA methylation and gene expression of grapevine leaf samples. The major contributing environmental factors for DNA methylation and gene expression variability were planting year and annual rainfall. Finally, our work showed that DNA methylation profiles could be used to build provenance predictive models, highlighting the potential application of DNA methylation as a tool against provenance fraud.

Methods

Plant material

The sampling strategy and plant material is described in Xie et al. (2017). In brief, 198 *Vitis vinifera* cv. Shiraz vines from twenty-two commercial vineyards located in the Barossa Wine Zone (South Australia, Australia) were selected for this study. Vineyards were chosen to be representative of the climate, soil and management practices that are used in the different Barossa sub-regions (Robinson and Sandercock 2014). These sub-regions are the Eden Valley (EV), Northern Grounds (NG), Central Grounds (CG), Southern Grounds (SG), Eastern Edge (EE) and Western Ridge (WR). All vineyards were planted with own-rooted vines of Shiraz. Ten vineyards were planted with clone SA 1654 (Whiting 2003), one with clone BVRC30 (Whiting 2003), one with clone PT15 Griffith (Farquhar 2005) and the remaining 10 vineyards the clonal status was classified as 'unknown' (Additional file 1: Table S1). Nine plants from three rows for each vineyard were sampled during the 2016 growing season. Leaf (first fully expanded leaf at bud burst, E-L 7) (Coombe 1995) samples were collected from three nodes per plant and pooled into a single sample per plant. All samples were taken before dawn (between 10:00 PM and sunrise) to minimize variability associated with differences in plant water status (Williams and Araujo 2002). Samples were snap-frozen in liquid nitrogen in the vineyards and stored at -80 °C until later use. Topographic and climatic data was obtained from the Barossa Grounds project

(Robinson and Sandercock 2014). Vineyard management information was collected from participating growers. Soil texture was determined following (Dpi.nsw.gov.au 2016). Soil pH and electrical conductivity were determined on a 1:5 soil/water mixture and then measured using pH/salinity meter (WP-81 Conductivity-Salinity-pH-mV Meter, v6.0, TPS Pty Ltd). Plant-available phosphorus was extracted and measured using Colwell P method. Principal component analysis (PCA) of environmental data was generated using XLSTAT version 2015.04.01.20116 (Addinsoft SARL, Paris, France).

Methylation Sensitive Genotyping by Sequencing

Genomic DNA was obtained from all 198 leaves samples powdered using an automatic mill grinder (Genogrinder). The DNA extraction was performed in the Oktopure automated DNA extraction platform (LGC Genomics GmbH) according to manufacturer instructions. DNA was quantified using the Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, United States). Methylation sensitive genotyping by sequencing (msGBS) was performed as described by Kitimu et al. (2015). Genomic DNA (200 ng) was digested with 8 U of HF-*EcoRI* (recognition site G^AATTC, methylation insensitive) and 8 U of *MspI* (recognition site C^CGG, methylation sensitive) (New England Biolabds Inc, Ipswich, MA, United States) for two hours at 37°C followed by 10 minutes of enzyme inactivation at 65°C. Sequencing adapters were ligated by adding 0.1 pmol *MspI* adapters (uniquely barcoded for each of the 198 samples), 15 pmol of the *EcoRI* Y adapter, 200 U of T4 ligase and T4 ligase buffer (New England BioLabs Inc., Ipswich, MA, United States) to each individual restriction product. Ligation reactions were incubated for 2 hours at 24°C followed by 10 minutes at 65°C for enzyme inactivation. Excess adapters were removed from ligation products using Agentcourt AMPure XP beads (Beckman Coulter, Australia) at a ratio of 0.85x following manufacturer's instructions. Two sequencing libraries were generated by pooling 25 ng of ligation product from 99 samples. The library was

amplify with eight PCR reactions containing 10 µl of the library, 0.25 µl Q5 HF polymerase, 5 µl 5x Q5 HF buffer and 1 µl of both forward and reverse primers at 10 µM. PCR was performed in a Bio-Rad T100 thermocycler (98°C for 30 seconds, 10 cycles of 98°C for 30seconds, 62°C for 20 seconds and 72°C for 30 seconds followed by 72°C for 5 minutes). PCR products were pooled again. DNA Fragments ranging between 200 and 350 bp in size were captured using the AMPure XP beads following manufacturer's instructions. Libraries were sequenced using Illumina NextSeq High Output (Illumina Inc., San Diego, CA, United States) 75 bp pair-end run at the Australian Genome Research Facility (AGRF, Adelaide, SA, Australia).

msGBS data analysis

Raw sequencing data were demultiplexed using GBSX (Herten et al. 2015) and filtered for reads matching the barcodes used during the library construction. Pair-end reads were merged using bbmerge (Bushnell et al. 2017). Paired-reads were mapped to the reference genome of *Vitis vinifera* (IGGP_12X.V1 from EnsemblPlants) obtained from the Ensembl Plants website (plants.ensembl.org/Vitis_vinifera/) using Bowtie2 v2.2.3 (Langmead and Salzberg 2012). BAM files were sorted and indexed with Samtools v1.2 (Li et al. 2009), and used as source data to perform subsequent analysis in the R environment (R Core Team 2014). The genetic diversity between the sub-regions was assessed by SNP calling using TASSEL3 (Bradbury et al. 2007) from the msGBS sequencing data. SNPs were only selected for further analysis if they were present in at least 80% of the samples. TASSEL3 was used to obtain the first three principal components from the SNPs data and clustering analysis for DNA methylation and gene expression data was performed through the supervised method Principal Components-Linear Discriminant Analysis (PC-LDA) using the R package FIEmspro (Enot et al. 2008).

DNA methylation was inferred via read coverage. Differentially methylated markers (DMM) between sub-regions were identified using the R package msgbsR (Mayne et al. 2018). In brief,

the *rawCounts* function imports the sorted and indexed BAM files and extracts the genomic location of the cut site for each read. Incorrectly mapped reads are filtered with the *checkCuts* function. The function *diffMeth* normalizes the sequencing data based on library size and identifies differentially methylated sites. The methylation status between samples was determined by read counts from the sequenced *MspI* loci with at least 1 CPM and present in at least 9 samples per sub-region. Differentially methylated markers were detected by pairwise comparison between sub-regions, setting the false discovery rate (FDR-adjusted p-value < 0.05) for differences in read CPM. To identify protein coding genes associated to DMMs, we used bedtools (Quinlan and Hall 2010) to find DMMs located within 5 kb of gene upstream regions, gene bodies and 5 kb of gene downstream regions in the reference genome. Genes that presented DMMs were referred to as differentially methylated genes (DMGs).

RNA isolation and sequencing

Messenger RNA was extracted from 198 frozen leaves samples grounded using mortar and pestle. The RNA extraction was performed using the Spectrum Plant Total RNA kit following manufacturer instructions (Sigma-Aldrich), including the DNAase step to digest genomic DNA. RNA was quantified using the Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, United States). RNA integrity and quality was obtained using Experion RNA StdSens Analysis Kit (Bio-Rad Laboratiories, Inc.). Three samples per vineyard were generated by pooling 2 µg of total RNA from three plants from the same row in the vineyard generating 66 pools (3 pools per vineyard). Ribosome RNA depletion was performed to 6 µg of RNA using the Dynabeads mRNA purification kit (Ambion, Invitrogen) following manufacturer instructions. RNA was quantified using the Nanodrop 2000 spectrophotometer. The libraries were made with 25 ng of depleted RNA using the NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs Inc.) following manufacturer instructions. The concentration of the libraries was

determined using QubitR 2.0 fluorometer (ThermoFisher Scientific, USA) and fragment sizes were assessed on the Bioanalyzer High Sensitive Chip (Agilent Technologies, Inc). Libraries were sequenced using Illumina NextSeq High Output (Illumina Inc., San Diego, CA, United States) 75 bp pair-end run at the Australian Genome Research Facility (AGRF, Adelaide, SA, Australia).

RNA-seq data analysis

Pair-ends reads were filtered with AdapterRemoval v2.2.1 (Schubert et al. 2016) using default settings. Reads were aligned to the reference genome (IGGP_12X.V1 from EnsemblPlants) with Hisat2 v2.1.0 (Kim et al. 2015). Reads counts were obtained from FeatureCounts v1.5.2 (Liao et al. 2014) using default settings. Analysis of differentially expressed genes was performed in R environment with the R package edgeR (Robinson et al. 2010) with a p-value set to < 0.05. In order to identify both differentially methylated and differentially expressed genes we combined the list of DMGs and DEGs and selected the genes that were in both lists. The expression (CPM) of the differentially methylated and expressed genes was plotted using heatmap.2 from gplot v3.0.1 (Warnes et al. 2009). Plotting of the DNA methylation (-log10 of the fold change) versus gene expression (log10 of the fold change) was used to find a correlation between localization/position of the DNA methylation and direction of the gene expression.

Detection of environmental effect in gene expression and DNA methylation

Expression and methylation data were filtered by removing genes and methylation sites containing fewer than 9 samples with at least 1 read CPM within them. Normalising factors were calculated using trimmed mean of M-values (TMM) method (Robinson et al. 2010). To identify genes whose expression changed with a specific environmental variable, we analysed the gene expression data with the *voom* method (Law et al. 2014) that estimates the mean-variance relationships of the log-counts of expression and generates a precision weight for each

observation (Law et al. 2014). The design matrix was constructed from the environmental variable. A linear model was fitted to the gene counts using *lmFit* function including empirical Bayes adjustment for moderated statistics (Robinson and Smyth 2007). Genes that significantly (FDR-adjusted p-value < 0.05) fit with the linear model were selected. Differential methylation was analyzed using generalized linear model methods in order to assess the changes in methylation in response to changing environmental conditions. We used the *glmFit* function within the edgeR package (McCarthy et al. 2012) to analyze count data from the msGBS results using an FDR-adjusted p-value < 0.05. The result of differential gene expression and differential methylation for each environmental variable were combined to identify genes that were both differentially expressed and differentially methylated. A heatmap of genes show their expression and methylation correlate with change in altitude and was visualised using the gplot package (Warnes et al. 2009) split in four quadrants depending on the direction of the fold change. Gene annotation and identification of gene ontology (GO) terms were obtained from Cramer et al. (2020). Singular enrichment analysis (SEA) of GO terms was performed in the web-based tool agriGO (Du et al. 2010; Tian et al. 2017) with default settings. The output of the SEA was loaded into REVIGO (Supek et al. 2011) to visualize the enriched GO terms from our data.

List of abbreviations

ABA: Abscisic acid, ATP: Adenosine triphosphate, bp: Base pairs, CG: Central Grounds, CPM: counts per million, CTP: Cytidine triphosphate, DAPC: Descriptive analysis of principal components, DEGs: Differentially expressed genes, DMEGs: Differentially methylated and expressed genes, DMGs: Differentially methylated genes, DMMs: Differentially methylated markers, DNA: Deoxyribonucleic acid, EE: Eastern Edge, EV: Eden Valley, Gb: Giga bases, GTP: Guanosine triphosphate, m a.s.l.: Meters above the sea level, Mb: Mega bases, ML/ha: Megaliter per hectare, MSAP: Methylation-sensitive amplified polymorphism, msGBS:

Methylation sensitive genotyping by sequencing, NG: Northern Grounds, PC1: Principal component 1, PC2: Principal component 2, PCA: Principal component analysis, PC-LDA: Principal components-Linear discriminant analysis RNA: Ribonucleic acid, SDG: SET DOMAIN GROUP, SG: Southern Grounds, SNPs: Single nucleotide polymorphisms, UTP: Uridine triphosphate, UV: Ultraviolet and WR: Western Ridge.

Acknowledgements

This study was funded through a Pilot Program in Genomic Applications in Agriculture and Environment Sectors jointly supported by the University of Adelaide and the Australian Genome Research Facility Ltd. PJF was supported by Graduate Research Scholarships from Wine Australia (PH1503) and the University of Adelaide. We thank the Barossa Grounds Project and the growers who allowed us to collect samples and supplied information about their vineyards. We thank Kendall R. Corbin for performing the DNA extraction of the leaf samples. We thank Dr. Cassandra Collins for the experimental design and collection of the plant material. We thank Dr. Sandra Milena Mantilla, Annette James, and Valentin Olek who helped with the sample collection. We thank Dr. Timothy Cavagnaro and Dr. Andrew Metcalfe for their contribution in the experimental design. This work was supported with supercomputing resources provided by the Phoenix HPC service at the University of Adelaide.

Availability of data and materials

The full sequencing data are available on SRA database under the accession number PRJNA591273.

Table

Table 1: Total numbers of DMMs and genes identified through generalized linear model that were significantly associated (FDR-adjusted p-value < 0.05) to environmental variables and soil properties.

Environmental variable	Number of DMMs that are significantly associated with the measured variables	Number of DMGs (genes at +/- 5Kb from a DMM) significantly associated with the measured variables	Number of genes whose expression is significantly associated with the measured variables	Number of genes whose its methylation and expression are significantly associated with the measured variables (intersection between previous two columns)
Planting year	845	170	6,072	97
Mean annual rainfall (mm)	6140	4088	81	16
Mean season rainfall (mm)	424	221	417	2
Soil pH	186	112	1225	5
Mean growing season mean temperature (°C)	14	10	170	0
Soil Phosphorus (µg P/g soil)	11	12	913	0
Salinity (g/l)	5	6	1	0
Altitude (m a.s.l.)	3	4	3,914	0

Figure Legends

Figure 1. Combined PCA biplot of values for environmental and soil variables in the factors space (F1 and F2) from 22 vineyards over the season 2016 from Shiraz vines in the Barossa Wine Zone. Vineyards are also represented in the PCA space. Shared colours indicate common sub-regions.

Figure 2. Visualization of genetic and epigenetic diversity between grapevine samples from six sub-regions of the Barossa Wine Zone. A) Three-dimensional plot of PCA of 4983 SNPs presenting the genetic diversity of leaves samples. All the samples are mapped on the first three major components of variability (PC1, PC2 and PC3). B) Three-dimensional plot of PC-LDA of 450,501 msGBS markers showing the DNA methylation diversity of 179 samples on the three axis between-group variation (discriminant factor, DF). C) UpSet plot summarizing the number of DMMs between sub-regions. The horizontal bar graphs show the total number DMMs for each sub-region. Connected black circles in each figure represent the overlapping combination between sub-regions. The top bar graphs summarize the number of DMMs between connected sub- regions.

Figure 3. Boxplots of the DNA methylation status in sub-regions. Differentially methylated markers (DMMs), generated from the msGBS, were identified in pairwise comparison between sub-regions from the Barossa Wine Zone. Boxplots show the distribution of the DNA methylation changes (-log10 fold change) of the DMMs between one sub-region against the other five, A) Central Grounds, B) Eastern Edge, C) Eden Valley, D) Northern Grounds, E) Southern Grounds and F) Western Ridge. Median represents the direction of the DNA methylation at whole genome level (positive median = increase global in DNA methylation, negative media = decrease in global DNA methylation), horizontal blue line no change in methylation.

Figure 4. Visualization of gene expression diversity between grapevine samples from six sub-regions of the Barossa Wine Zone. A) Three-dimensional plot of PC-LDA of the gene expression profile of 18,952 genes of 66 samples on the three axis between-group variation (discriminant function, DF). B) UpSet plot summarizing the intersection of DEGs between sub-regions. The horizontal bar graphs show the total number of DEGs per sub-region. Connected black circles in each figure represent the overlapping combination between sub-regions. The top bar graphs summarize the number of DEGs between connected sub- regions.

Figure 5. Total numbers of the differentially methylated markers (blue bar), differentially expressed genes (purple bar) and the number of genes differentially methylated and expressed (green bar) for each sub-region of the Barossa Wine Zone.

Figure 6. Relationship between DNA methylation versus gene expression of the differentially methylated and expressed genes by each sub-region combination. Scatter plot of the gene expression (logFC) against changes in the DNA methylation (-logFC) depending on the position of the DMMs with respect to the gene (blue: up to 5 kb downstream the gene, green: up to 5 kb upstream the gene and red: gene body. Heatmap of the gene expression (logCPM) of differentially methylated and expressed genes for each combination of sub-regions. Hierarchical clustering was assembled for the vineyards (columns) depending on their expression similarities. Z-score represents the gene expression standard deviation from the mean, with red for low expression and white for high expression levels.

Figure 7. Effects of environmental factor on gene expression. Heatmaps of genes (logCPM) where expression and methylation significantly correlate (FDR-adjusted p-value < 0.05) with A) Planting year and B) Mean annual rainfall (mm). Hierarchical clustering groups vineyards by

their gene expression. Z-score represents the gene expression standard deviation from the mean, with red for low expression and white for high expression levels.

Figure 8. Effects of plant age on gene expression. Heatmap of genes (logCPM) where expression and methylation correlate with planting year of the grapevine. Genes are separated by the direction of the change in gene expression and methylation in relation to planting year. Positive changes in DNA methylation indicate decrease in DNA methylation while negative changes indicate increase in DNA methylation. Vineyards are sorted from left to right from older to younger grapevines. Z-score represents the gene expression standard deviation from the mean, with red for low expression and white for high expression levels. (Pannels on upper quadrants represent genes which expression has a significantly positive correlation with planting year; panels on lower quadrants represent genes which expression has a significantly negative correlation with planting year; panels on right quadrants represent genes which DNA methylation has a significantly positive correlation with planting year; panels on left quadrants represent genes which DNA methylation has a significantly negative correlation with planting year)

Figures

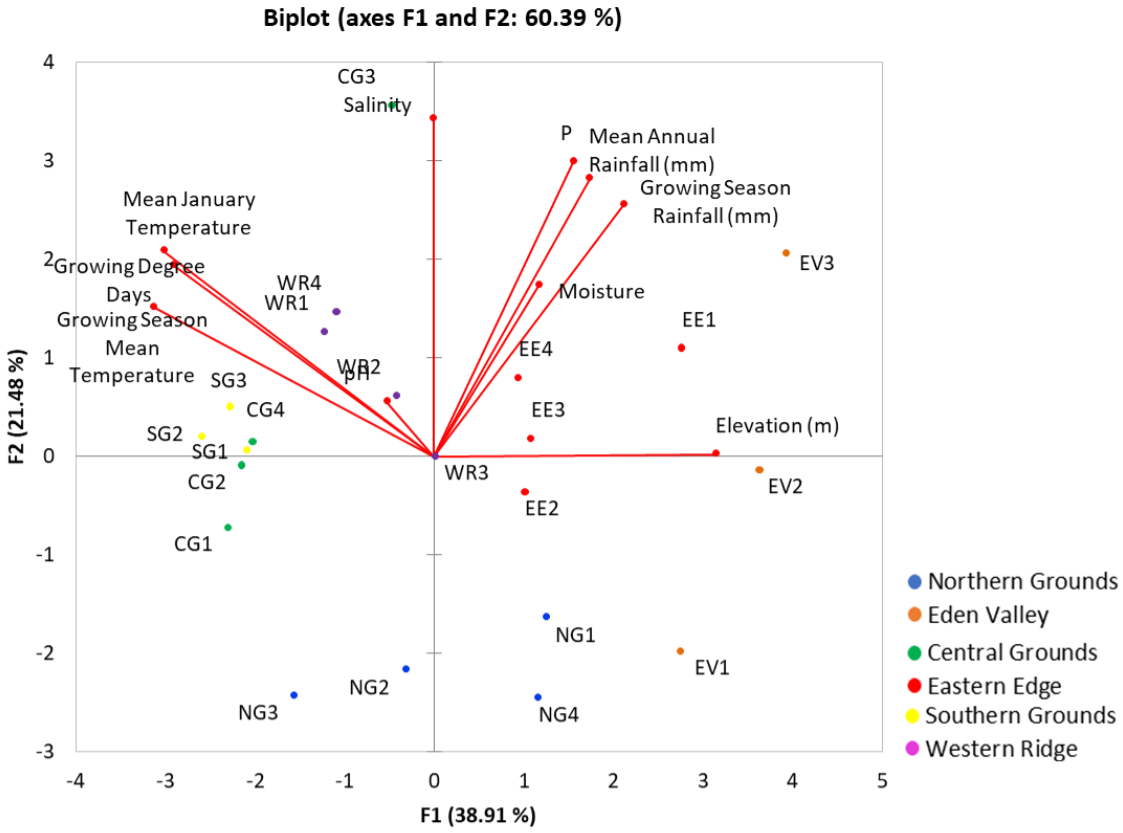


Figure 1. Combined PCA biplot of values for environmental and soil variables in the factors space (F1 and F2) from 22 vineyards over the season 2016 from Shiraz vines in the Barossa Wine Zone. Vineyards are also represented in the PCA space. Shared colours indicate vineyards from common sub-regions.

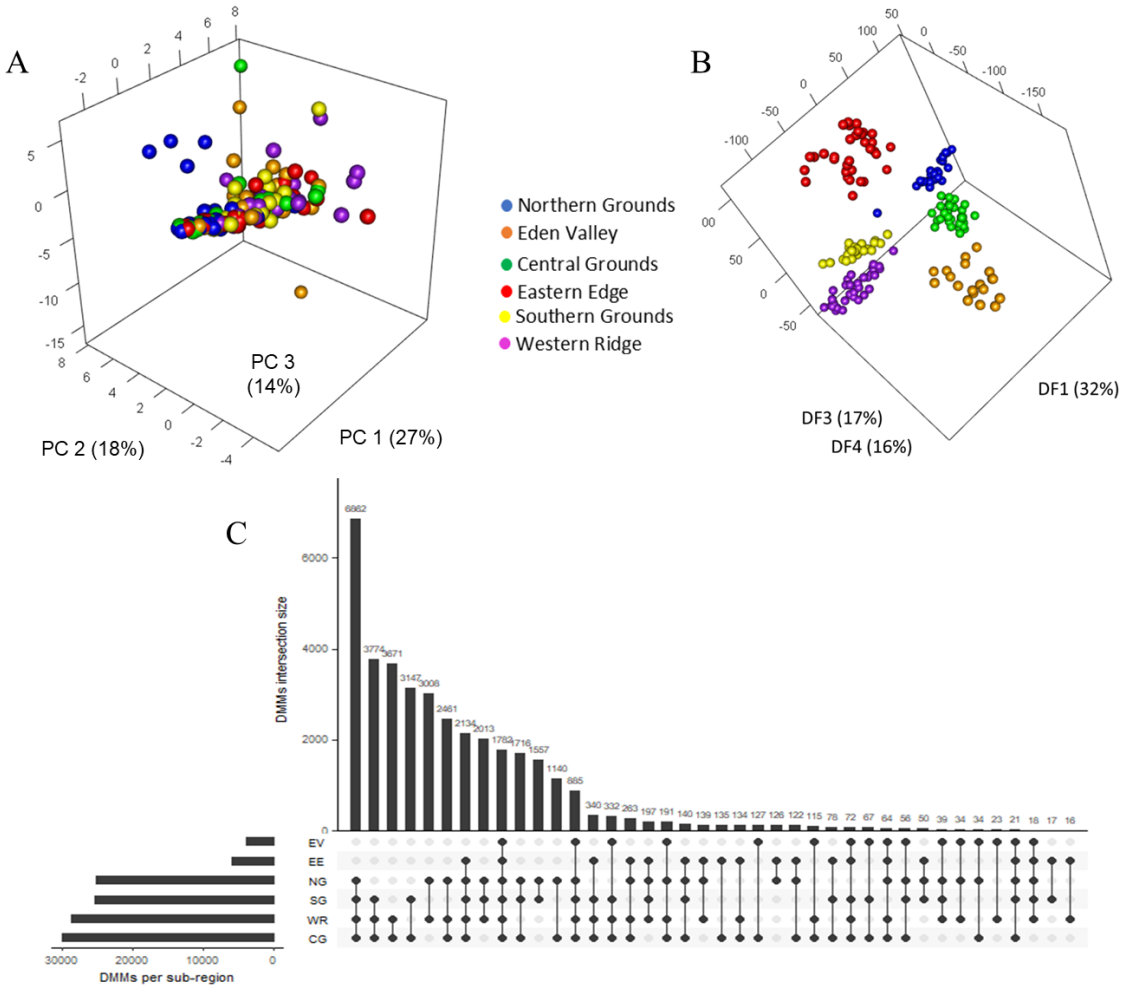


Figure 2. Visualization of genetic and epigenetic diversity between grapevine samples from six sub-regions of the Barossa Wine Zone. A) Three-dimensional plot of PCA of 4983 SNPs presenting the genetic diversity of leaves samples. All the samples are mapped on the first three major components of variability (PC1, PC2 and PC3). B) Three-dimensional plot of PC-LDA of 450,501 msGBS markers showing the DNA methylation diversity of 179 samples on the three axis between-group variation (discriminant factor, DF). C) UpSet plot summarizing the number of DMMs between sub-regions. The horizontal bar graphs show the total number DMMs for each sub-region. Connected black circles in each figure represent the overlapping combination between sub-regions. The top bar graphs summarize the number of DMMs between connected sub- regions.

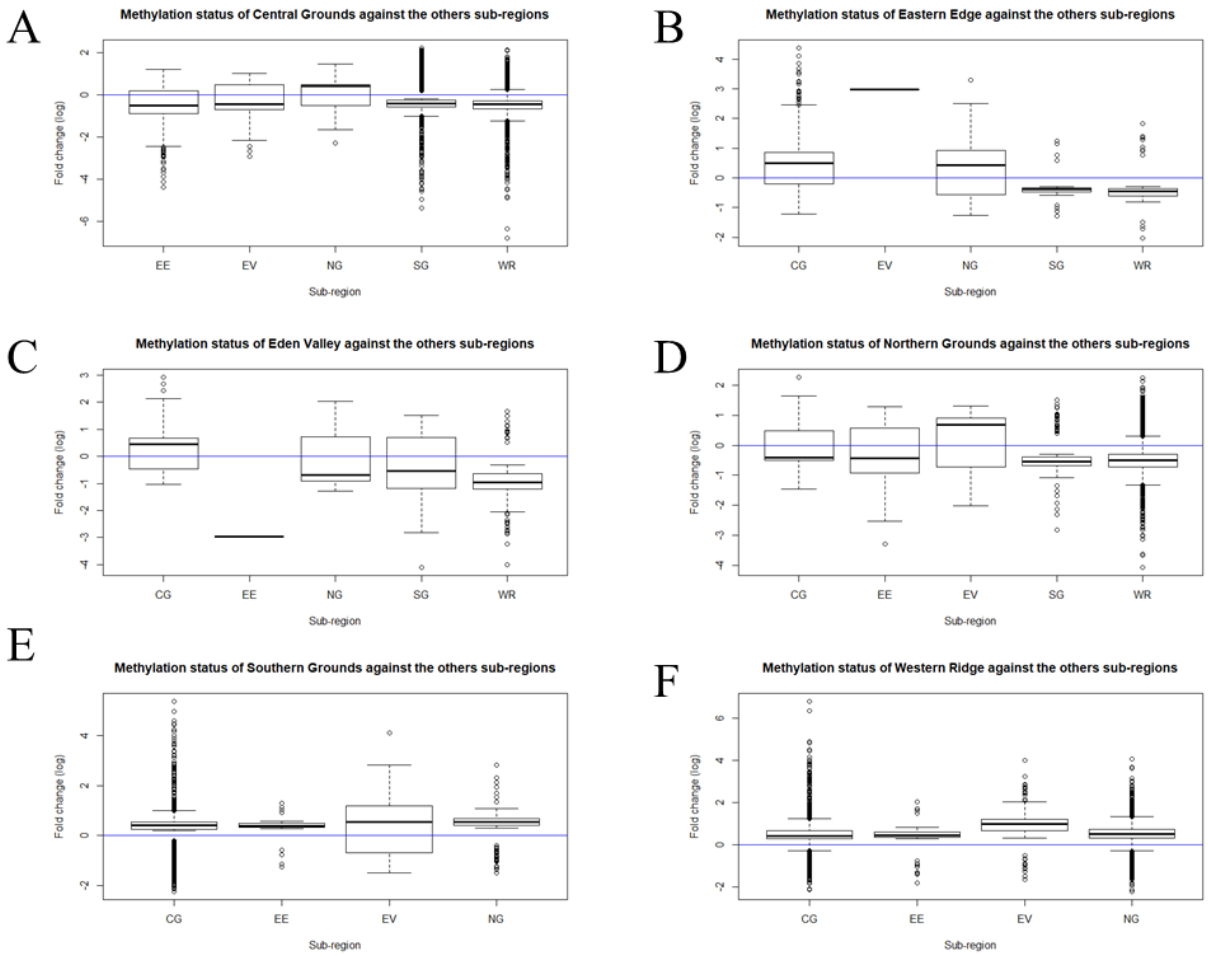


Figure 3. Boxplots of the DNA methylation status in sub-regions. Differentially methylated markers (DMMs), generated from the msGBS, were identified in pairwise comparison between sub-regions from the Barossa Wine Zone. Boxplots show the distribution of the DNA methylation changes (-log10 fold change) of the DMMs between one sub-region against the other five, A) Central Grounds, B) Eastern Edge, C) Eden Valley, D) Northern Grounds, E) Southern Grounds and F) Western Ridge. Median represents the direction of the DNA methylation at whole genome level (positive median = increase global in DNA methylation, negative media = decrease in global DNA methylation), horizontal blue line no change in methylation.

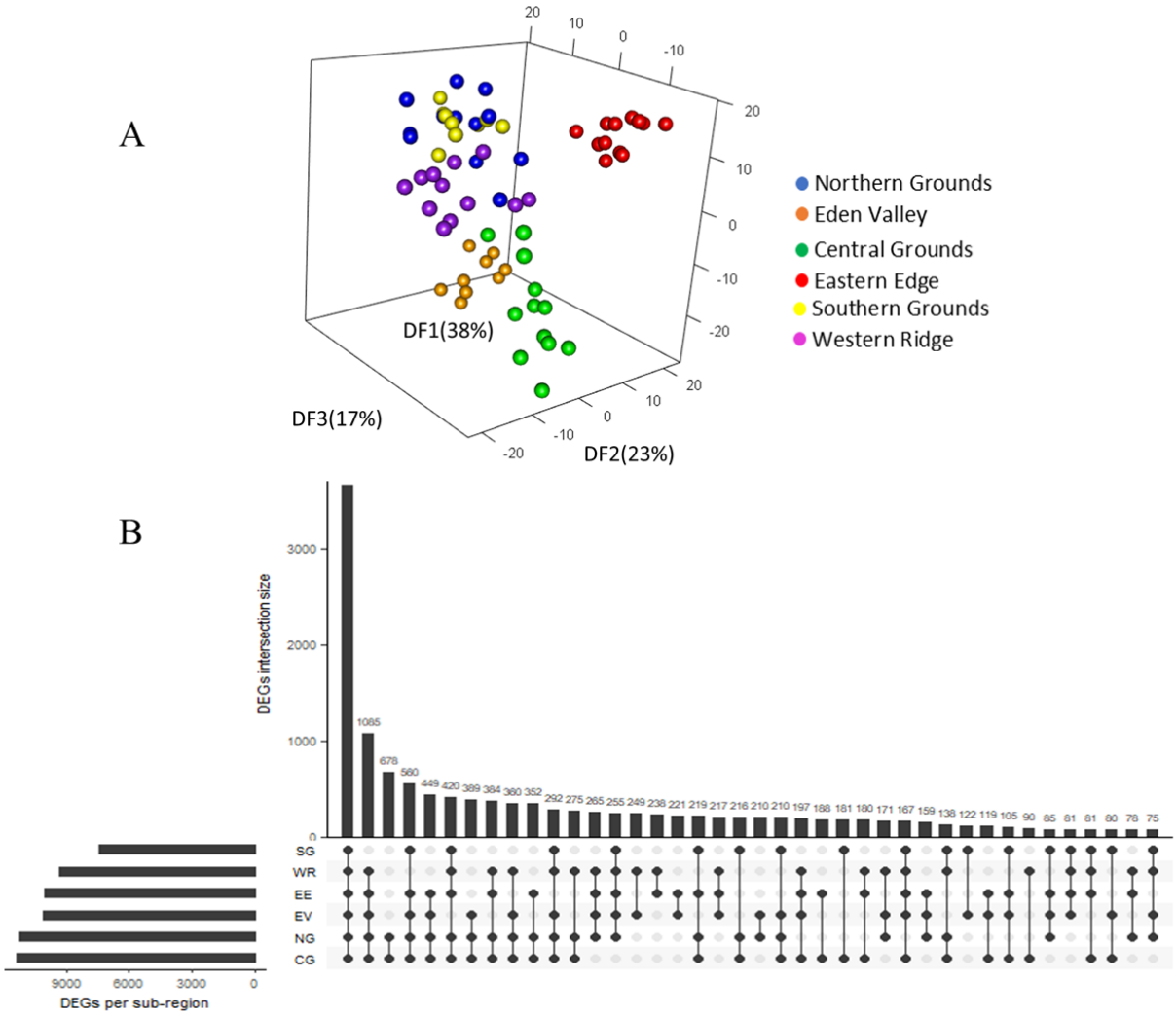


Figure 4. Visualization of gene expression diversity between grapevine samples from six sub-regions of the Barossa Wine Zone. A) Three-dimensional plot of PC-LDA of the gene expression profile of 18,952 genes of 66 samples on the three axis between-group variation (discriminant function, DF). B) UpSet plot summarizing the intersection of DEGs between sub-regions. The horizontal bar graphs show the total number of DEGs per sub-region. Connected black circles in each figure represent the overlapping combination between sub-regions. The top bar graphs summarize the number of DEGs between connected sub- regions.

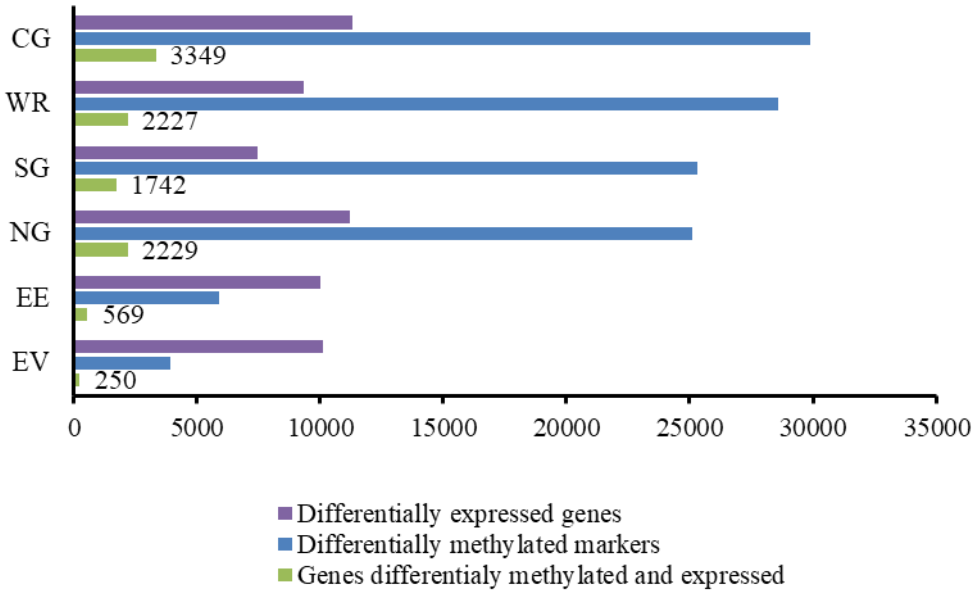
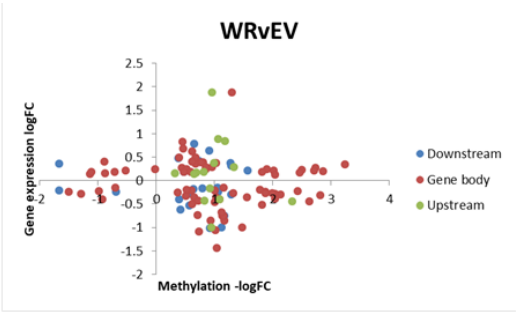
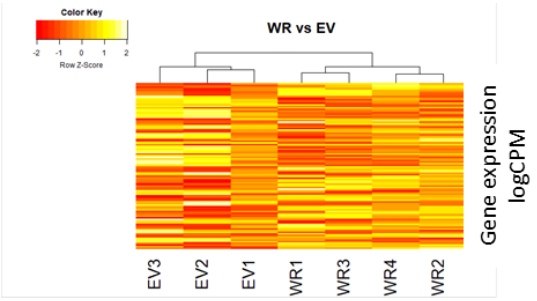


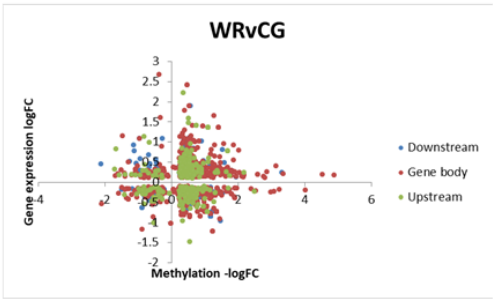
Figure 5. Total numbers of the differentially methylated markers (blue bar), differentially expressed genes (purple bar) and the number of genes differentially methylated and expressed (green bar) for each sub-region of the Barossa Wine Zone.

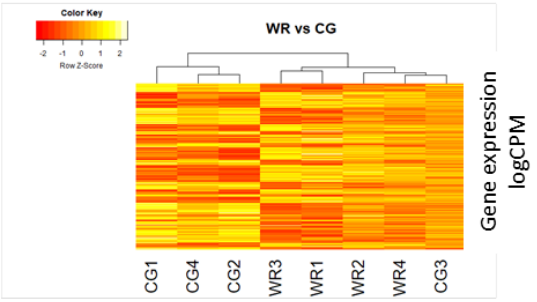
A


WRvEV
Gene expression logFC
Methylation -logFC
Downstream
Gene body
Upstream

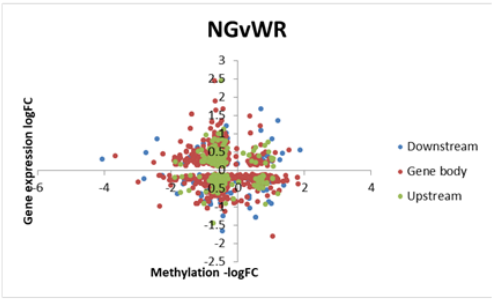

Color Key
Row Z-Score
WR vs EV
EV3
EV2
EV1
WR1
WR3
WR4
WR2
Gene expression logCPM

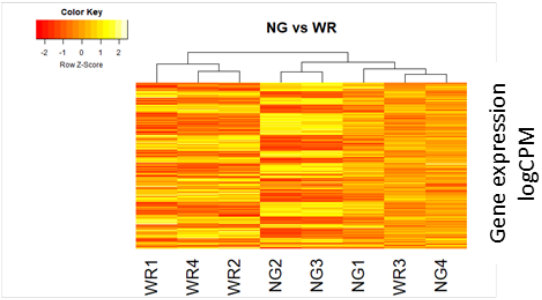
B


WRvCG
Gene expression logFC
Methylation -logFC
Downstream
Gene body
Upstream

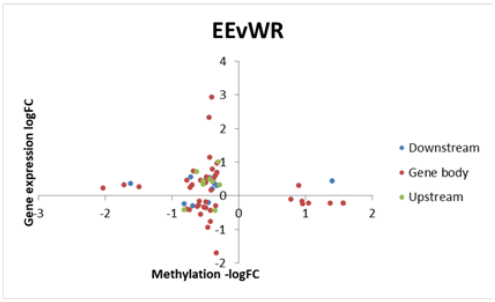

Color Key
Row Z-Score
WR vs CG
CG1
CG4
CG2
WR3
WR1
WR2
WR4
CG3
Gene expression logCPM

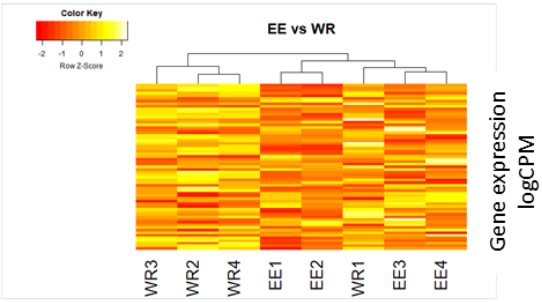
C


NGvWR
Gene expression logFC
Methylation -logFC
Downstream
Gene body
Upstream

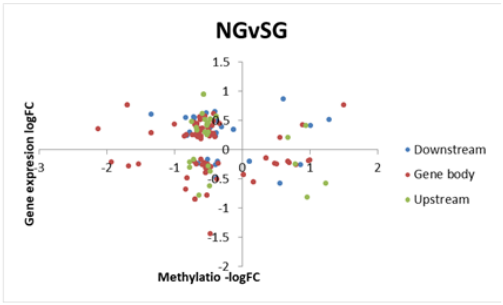

Color Key
Row Z-Score
NG vs WR
WR1
WR4
WR2
NG2
NG3
NG1
WR3
NG4
Gene expression logCPM

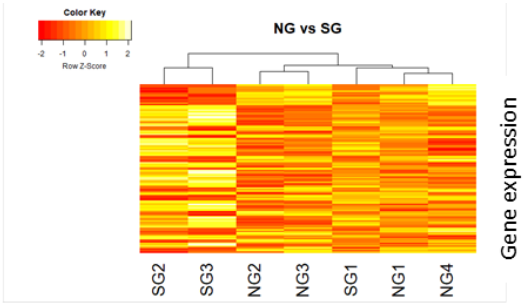
D


EEvWR
Gene expression logFC
Methylation -logFC
Downstream
Gene body
Upstream

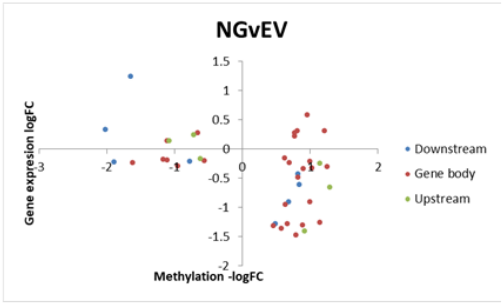

Color Key
Row Z-Score
EE vs WR
WR3
WR2
WR4
EE1
EE2
WR1
EE3
EE4
Gene expression logCPM

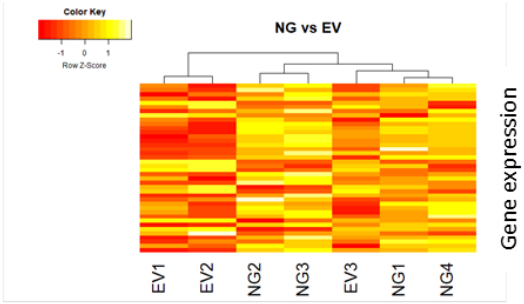
E


NGvSG
Gene expresion logFC
Methylatio -logFC
Downstream
Gene body
Upstream

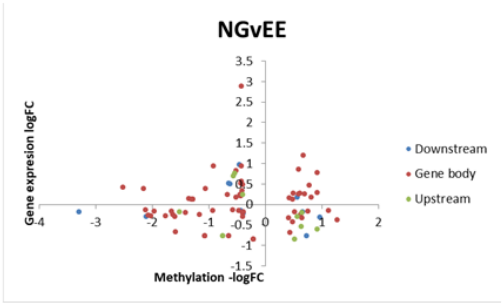

Color Key
Row Z-Score
NG vs SG
SG2
SG3
NG2
NG3
SG1
NG1
NG4
Gene expression logCPM

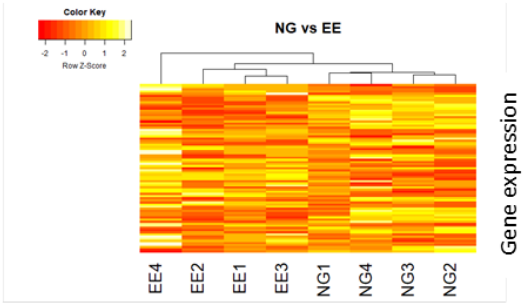
F


NGvEV
Gene expresion logFC
Methylation -logFC
Downstream
Gene body
Upstream

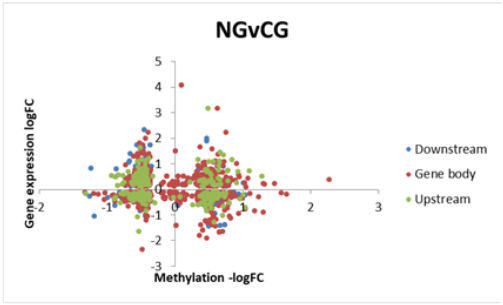

Color Key
Row Z-Score
NG vs EV
EV1
EV2
NG2
NG3
EV3
NG1
NG4
Gene expression logCPM

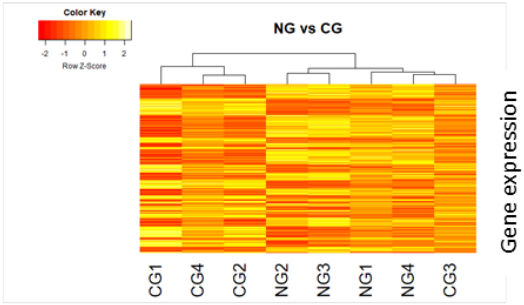
G


NGvEE
Gene expresion logFC
Methylation -logFC
Downstream
Gene body
Upstream

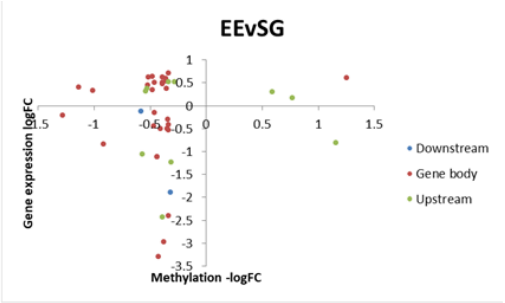

Color Key
Row Z-Score
NG vs EE
EE4
EE2
EE1
EE3
NG1
NG4
NG3
NG2
Gene expression logCPM

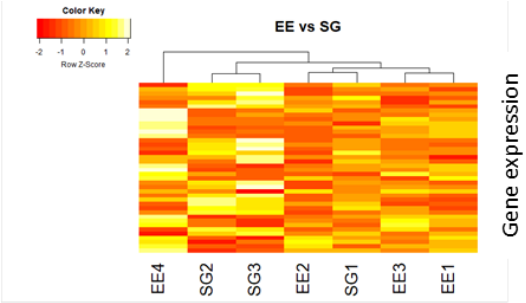
H


NGvCG
Gene expression logFC
Methylation -logFC
Downstream
Gene body
Upstream

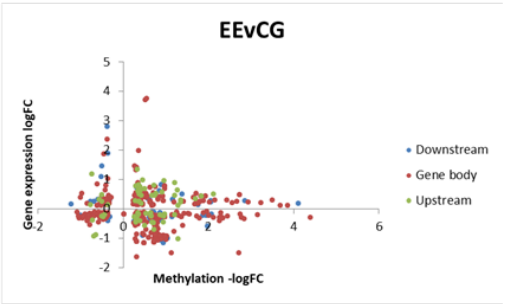

Color Key
Row Z-Score
NG vs CG
CG1
CG4
CG2
NG2
NG3
NG1
NG4
CG3
Gene expression logCPM

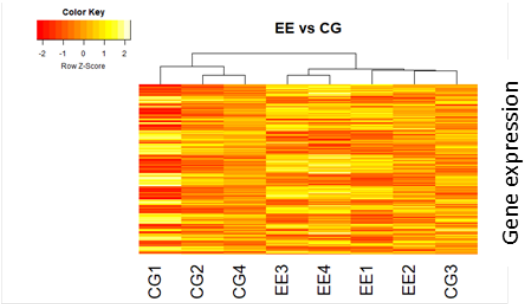
I


EEvSG
Gene expression logFC
Methylation -logFC
Downstream
Gene body
Upstream

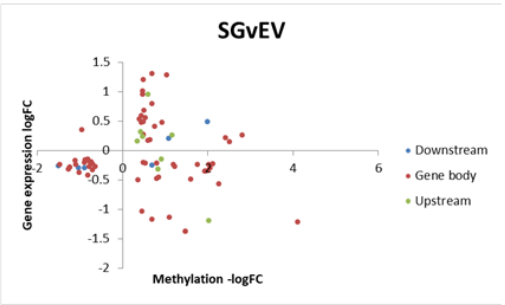

Color Key
Row Z-Score
EE vs SG
EE4
SG2
SG3
EE2
SG1
EE3
EE1
Gene expression logCPM

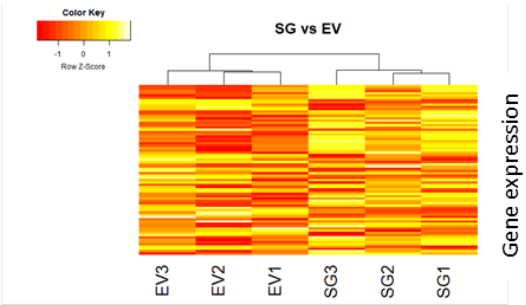
J


EEvCG
Gene expression logFC
Methylation -logFC
Downstream
Gene body
Upstream

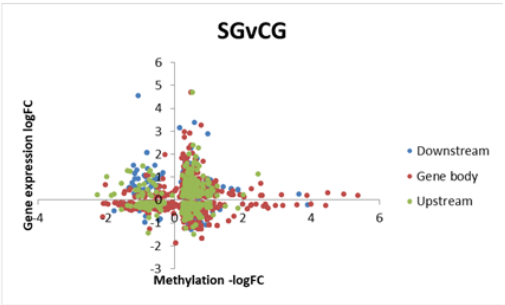

Color Key
Row Z-Score
EE vs CG
CG1
CG2
CG4
EE3
EE4
EE1
EE2
CG3
Gene expression logCPM

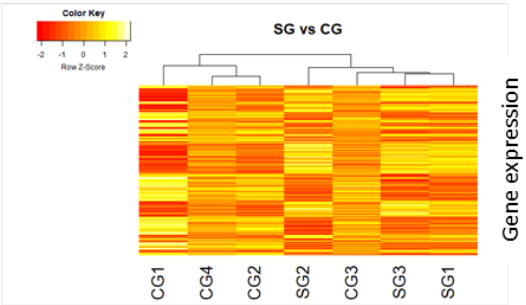
K


SGvEV
Gene expression logFC
Methylation -logFC
Downstream
Gene body
Upstream

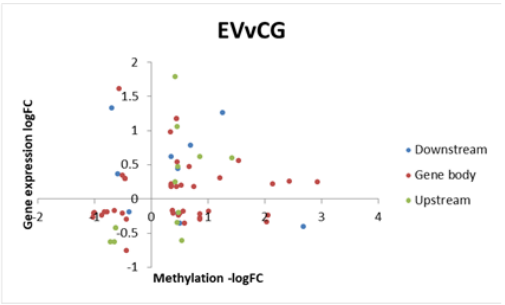

Color Key
Row Z-Score
SG vs EV
EV3
EV2
EV1
SG3
SG2
SG1
Gene expression logCPM

L


SGvCG
Gene expression logFC
Methylation -logFC
Downstream
Gene body
Upstream


Color Key
Row Z-Score
SG vs CG
CG1
CG4
CG2
SG2
CG3
SG3
SG1
Gene expression logCPM

M


EVvCG
Gene expression logFC
Methylation -logFC
Downstream
Gene body
Upstream

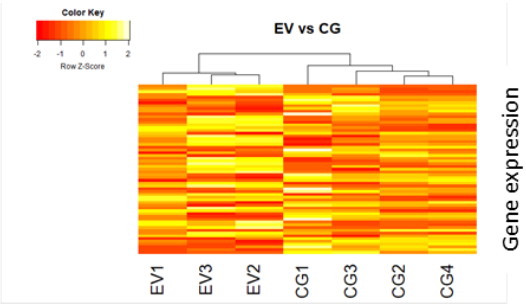

Color Key
Row Z-Score
EV vs CG
EV1
EV3
EV2
CG1
CG3
CG2
CG4
Gene expression logCPM

Figure 6. Relationship between DNA methylation versus gene expression of the differentially methylated and expressed genes by each sub-region combination. Scatter plot of the gene expression (logFC) against changes in the DNA methylation (-logFC) depending on the position of the DMMs with respect to the gene (blue: up to 5 kb downstream the gene, green: up to 5 kb upstream the gene and red: gene body. Heatmap of the gene expression (logCPM) of differentially methylated and expressed genes for each combination of sub-regions. Hierarchical clustering was assembled for the vineyards (columns) depending on their expression similarities. Z-score represents the gene expression standard deviation from the mean, with red for low expression and white for high expression levels.

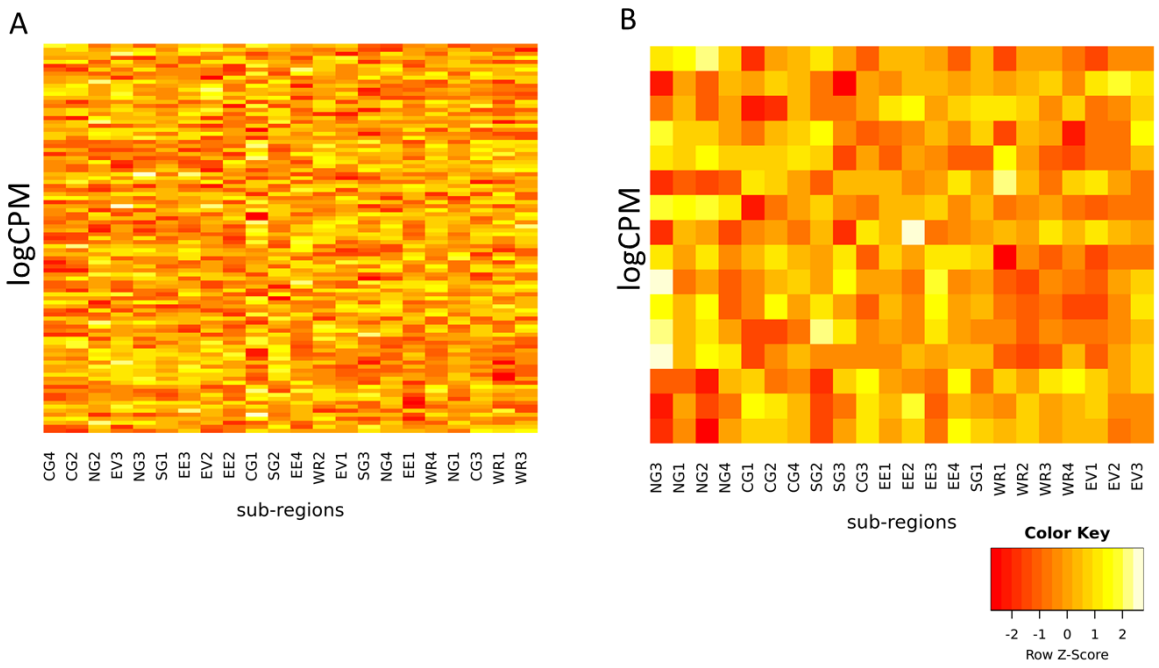


Figure 7. Effects of environmental factor on gene expression. Heatmaps of genes (logCPM) where expression and methylation significantly correlate (FDR-adjusted p-value < 0.05) with A) Planting year and B) Mean annual rainfall (mm). Hierarchical clustering groups vineyards by their gene expression. Z-score represents the gene expression standard deviation from the mean, with red for low expression and white for high expression levels.

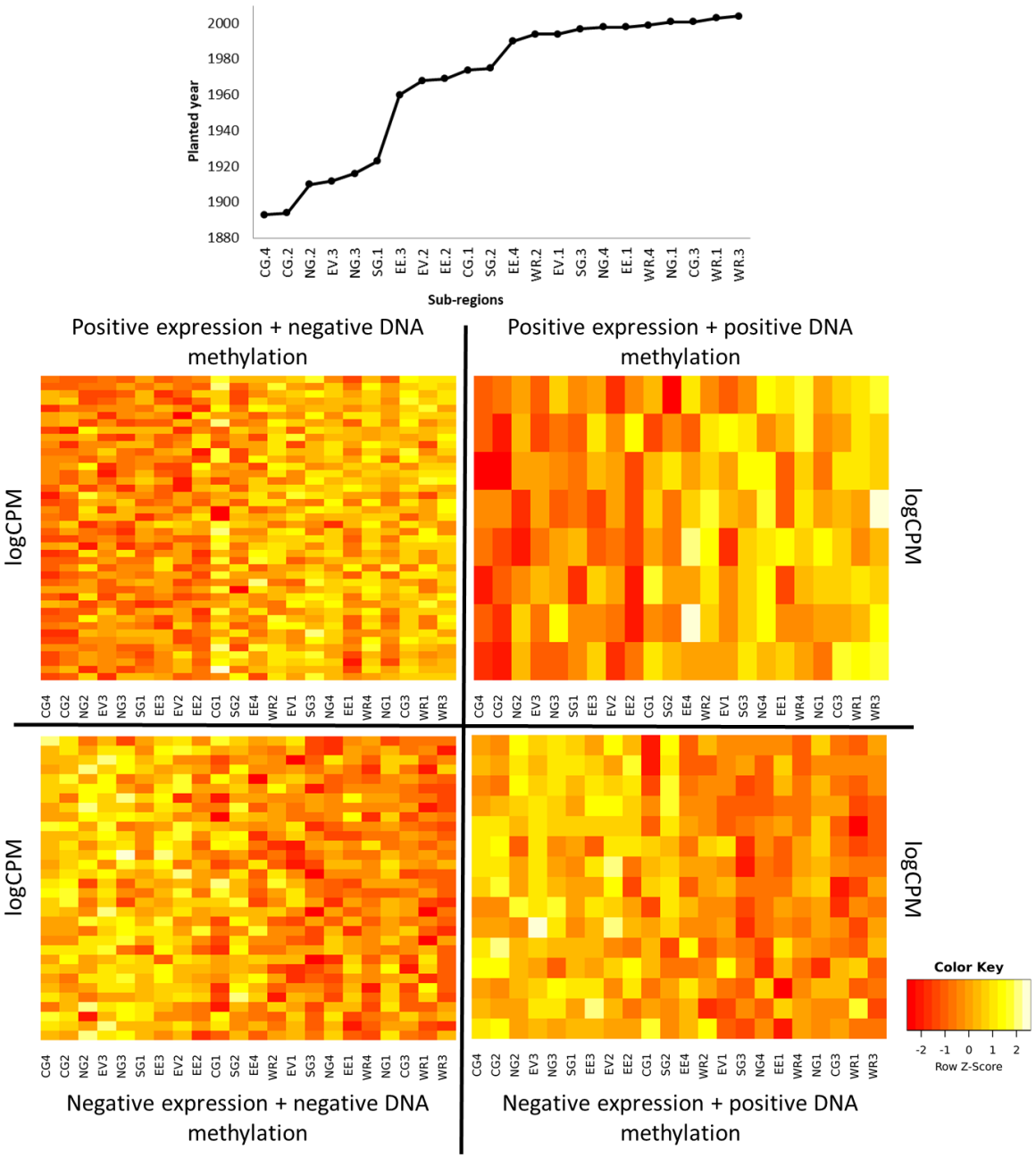


Figure 8. Effects of plant age on gene expression. Heatmap of genes (logCPM) where expression and methylation correlate with planting year of the grapevine. Genes are separated by the direction of the change in gene expression and methylation in relation to planting year. Positive changes in DNA methylation indicate decrease in DNA methylation while negative changes indicate increase in DNA methylation. Vineyards are sorted from left to right from older to younger grapevines. Z-score represents the gene expression standard deviation from the mean,

with red for low expression and white for high expression levels. (Pannels on upper quadrants represent genes which expression has a significantly positive correlation with planting year; panels on lower quadrants represent genes which expression has a significantly negative correlation with planting year; panels on right quadrants represent genes which DNA methylation has a significantly positive correlation with planting year; panels on left quadrants represent genes which DNA methylation has a significantly negative correlation with planting year).

Additional file 1: Table S1. Vineyard characteristics for vineyards sampled in this study: Vineyard region and code: NG: Northern Grounds; CG: Central Grounds; EE: Eastern Edge; SG: Southern Grounds; WR: Western Ridge; EV: Eden Valley. 3-4 sites were sampled by sub-region, numbered as NG1, NG2, NG3, NG4 and so on. Vineyard management information was collected from participating growers.

Vineyards details

Site Code	Site Number	Sub-region	Planted	Clone	Rootstock
NG1	1	Northern Grounds	2001	1654	Own Roots
NG2	2	Northern Grounds	1910	Unknown	Own Roots
NG3	3	Northern Grounds	1916	Unknown	Own Roots
NG4	4	Northern Grounds	1998	1654	Own Roots
CG1	5	Central Grounds	1974	PT15 Griffith Clone	Own Roots
CG2	6	Central Grounds	1894	Unknown	Own Roots
CG3	7	Central Grounds	2001	1654	Own Roots
CG4	8	Central Grounds	1893	Unknown	Own roots
EE1	9	Eastern Ridge	1998	1654	Own Roots
EE2	10	Eastern Ridge	1969	Unknown	Own Roots
EE3	11	Eastern Ridge	1960	Unknown	Own Roots
EE4	12	Eastern Ridge	1990	1654	Own Roots
SG1	13	Southern Grounds	1923	Unknown	Own Roots
SG2	14	Southern Grounds	1975	Unknown	Own Roots
SG3	15	Southern Grounds	1997	1654	Own Roots
WR1	16	Western Ridge	2003	1654	Own Roots
WR2	17	Western Ridge	1994	1654	Own roots
WR3	18	Western Ridge	2004	1654	Own Roots
WR4	19	Western Ridge	1999	1654	Own Roots
EV1	20	Eden Valley	1994	BVRC30	Own Roots
EV2	21	Eden Valley	1968	Unknown	Own Roots
EV3	22	Eden Valley	1912	Unknown	Own Roots

Additional file 1: Table S2. Vineyard climatic characteristics for vineyards sampled in this study: Vineyard region and code: NG: Northern Grounds; CG: Central Grounds; EE: Eastern Edge; SG: Southern Grounds; WR: Western Ridge; EV: Eden Valley. 3-4 sites were sampled by sub-region, numbered as NG1, NG2, NG3, NG4 and so on. Topographic and climatic data was obtained from the Barossa Grounds project (Robinson and Sandercock 2014). Vineyard management information was collected from participating growers.

Climate information

Site Code	Elevation (m)	Mean Annual Rainfall (mm)	Growing Season Rainfall (mm)	Mean January Temperature (°C)	Growing Season Mean Temperature (°C)	Growing Degree Days	Amount of irrigation (ML/ha)
NG1	300	547.18	213.37	20.4	17.43	1548.75	0.8
NG2	300	547.18	213.37	21.3	18.32	1548.75	0.3
NG3	310	431.18	213.37	21.3	19.21	1724.5	0
NG4	295	547.18	213.37	20.4	17.43	1548.75	0.2
CG1	270	547.18	213.37	22.2	19.21	1900.25	0.5
CG2	270	547.18	213.37	22.2	19.21	1900.25	0.1
CG3	300	663.18	245.255	22.2	19.21	1900.25	1.2
CG4	270	547.18	213.37	22.2	19.21	1900.25	0.3
EE1	448	663.18	245.255	21.3	18.32	1548.75	1.1
EE2	300	663.18	245.255	21.3	18.32	1548.75	0.3
EE3	300	663.18	245.255	21.3	18.32	1548.75	0.6
EE4	340	663.18	245.255	21.3	18.32	1724.5	0.7
SG1	229	663.18	213.37	22.2	19.21	1900.25	1.2
SG2	230	547.18	213.37	22.2	19.21	1900.25	0
SG3	250	547.18	213.37	22.2	19.21	1900.25	Not informed
WR1	256	663.18	245.255	22.2	19.21	1900.25	0.275
WR2	300	663.18	245.255	22.2	18.32	1900.25	1
WR3	300	663.18	245.255	21.3	18.32	1900.25	Not informed
WR4	260	663.18	245.255	22.2	19.21	1900.25	0.5
EV1	430	663.18	245.255	20.4	17.43	1548.75	0.1
EV2	490	663.18	245.255	20.4	17.43	1548.75	0.4
EV3	420	663.18	245.255	20.4	17.43	1548.75	0

Additional file 1: Table S3. Vineyard edaphic characteristics for vineyards sampled in this study: Vineyard region and code: NG: Northern Grounds; CG: Central Grounds; EE: Eastern Edge; SG: Southern Grounds; WR: Western Ridge; EV: Eden Valley. 3-4 sites were sampled by sub-region, numbered as NG1, NG2, NG3, NG4 and so on. Topographic and climatic data was obtained from the Barossa Grounds project (Robinson and Sandercock 2014). Soil texture was determined following (Dpi.nsw.gov.au 2016). Soil pH and electrical conductivity were determined on a 1:5 soil/water mixture and then measured using pH/salinity meter (WP-81 Conductivity-Salinity-pH-mV Meter, v6.0, TPS Pty Ltd). Plant-available phosphorus was extracted and measured using Colwell P method.

Soil information

Site Code	Soil Type	Available water holding capacity (mm)	pH	Plant available (Colwell) P (µg P / g dry soil)	Electrical Conductivity (Salinity g/l)	Moisture
NG1	Clay Loam over red clay	70-100	7.0	89.1	210.9	9.9
NG2	Sand over yellow/orange clay	40-70	6.9	59.6	79.8	0.9
NG3	clay loam over red clay with various calcium carbonate layers	70-100	6.3	59.7	58.0	1.5
NG4	Loam over clay over limestone	70-100	6.3	22.1	95.5	9.5
CG1	Loam over clay	>100	6.3	68.7	128.8	2.7
CG2	Sandy alluvial	>100	6.2	128.8	236.5	2.9
CG3	Thick sand over clay	>100	6.2	259.9	667.0	6.3
CG4	deep sandy loam (top soil) 30cm, med-heavy clay (80-100cm)	>100	6.9	151.9	177.6	6.5
EE1	Bubbly loam	>100	5.8	406.7	169.7	4.0
EE2	Sand over red clay	>100	6.4	138.4	98.3	1.8

Site Code	Soil Type	Available water holding capacity (mm)	pH	Plant available (Colwell) P (µg P / g dry soil)	Electrical Conductivity (Salinity g/l)	Moisture
EE3	High sand sandy loam over yellow/red clay	>100	6.9	133.8	182.9	4.9
EE4	Red brown earth over red clay	>100	7.7	100.2	176.3	10.4
SG1	Sandy Loam	20-40	6.5	138.9	84.5	1.2
SG2	Meadow dirt- high organic carbon content	>100	6.1	22.1	452.0	3.8
SG3	Deep Friable Gradational clay loam	>100	7.7	64.9	263.0	11.2
WR1	Clay Loam- Calcium Carbonate	40-70	6.9	116.6	154.6	5.3
WR2	Loam over red clay	70-100	6.4	84.5	123.0	4.4
WR3	Red/Brown earth	70-100	6.5	44.9	97.2	4.3
WR4	Shallow top soil- red clay/Limestone	40-70	6.3	93.3	241.9	6.9
EV1	Sandy loam over yellow Pudsolic clay	70-100	5.5	-1.4	45.9	1.0
EV2	Sandy loam over blocky clay	70-100	6.5	172.0	208.9	6.8
EV3	Red/brown earth - stone fragments - strong clay over shift	70-100	6.7	348.9	468.7	15.0

Additional file 2, Table S1. Reads per sample from msGBS sequencing (pair end 75 bp). In red are the samples that were discarded from downstream analyses due to low coverage or poor alignment to the reference genome (IGGP_12X.V1).

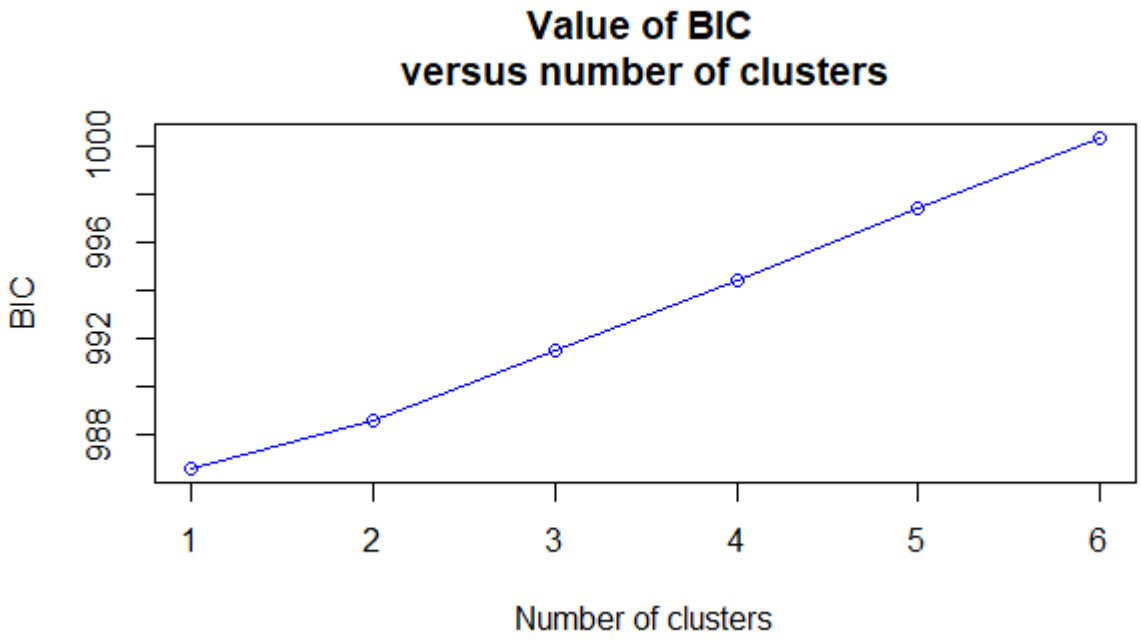
Sample ID	Total reads counts	Sample ID	Total reads counts
CG1.1	3,499,132	CG4.8	6,994,513
CG1.2	1,699,029	CG4.9	3,874,416
CG1.3	2,307,476	EE1.1	2,326,730
CG1.4	1,153,763	EE1.2	2,029,053
CG1.5	907,707	EE1.3	2,296,299
CG1.6	3,639,259	EE1.4	5,675,933
CG1.7	9,622,266	EE1.5	2,509,113
CG1.8	778,766	EE1.6	1,217,586
CG1.9	945,566	EE1.7	1,099,840
CG2.1	3,699,024	EE1.8	506,542
CG2.2	1,765,435	EE1.9	3,616,639
CG2.3	650,304	EE2.1	794,128
CG2.4	2,774,133	EE2.2	787,458
CG2.5	690,000	EE2.3	2,697,833
CG2.6	4,377,751	EE2.4	1,356,520
CG2.7	125,167	EE2.5	1,352,477
CG2.8	3,548,992	EE2.6	12,285
CG2.9	2,309,843	EE2.7	1,988,876
CG3.1	1,451,066	EE2.8	2,043,835
CG3.2	1,966,054	EE2.9	435,065
CG3.3	732,162	EE3.1	2,089,208
CG3.4	48	EE3.2	3,288,054
CG3.5	1,218,867	EE3.3	2,785,581
CG3.6	1,436,000	EE3.4	1,176,652
CG3.7	4,285,736	EE3.5	3,010,388
CG3.8	3,661,601	EE3.6	2,000,205
CG3.9	1,895,751	EE3.7	4,449,103
CG4.1	3,623,913	EE3.8	1,651,607
CG4.2	962,440	EE3.9	1,501,093
CG4.3	6,719,209	EE4.1	3,024,782
CG4.4	6,348,372	EE4.2	2,360,919
CG4.5	3,472,497	EE4.3	1,334,992
CG4.6	4,084,164	EE4.4	1,551,712
CG4.7	2,743,971	EE4.5	2,435,730

Sample ID	Total reads counts	Sample ID	Total reads counts
EE4.6	1,771,483	NG1.8	2,059,896
EE4.7	3,545,022	NG1.9	8,399,578
EE4.8	2,878,869	NG2.1	1,235,927
EE4.9	3,212,936	NG2.2	4,719,281
EV1.1	139	NG2.3	3,824,227
EV1.2	937	NG2.4	3,133,497
EV1.3	1,490,506	NG2.5	45
EV1.4	1,203,266	NG2.6	2,182,872
EV1.5	1,385,453	NG2.7	5,069,634
EV1.6	2,426,068	NG2.8	1,433,163
EV1.7	2,806,846	NG2.9	263,307
EV1.8	2,570,843	NG3.1	628,825
EV1.9	1,376,557	NG3.2	565,459
EV2.1	4,869,349	NG3.3	489,492
EV2.2	3,807,716	NG3.4	569,750
EV2.3	3,495,849	NG3.5	545,315
EV2.4	4,202,091	NG3.6	1,707,093
EV2.5	596,641	NG3.7	0
EV2.6	1,818,155	NG3.8	2,179,086
EV2.7	1,164,788	NG3.9	3,568,831
EV2.8	1,288,232	NG4.1	6,060,251
EV2.9	1,435,446	NG4.2	3,127,156
EV3.1	1,888,422	NG4.3	8,247,482
EV3.2	1,381,833	NG4.4	4,775,380
EV3.3	3,414,920	NG4.5	2,911,406
EV3.4	1,372,927	NG4.6	1,566,639
EV3.5	1,245,525	NG4.7	2,621,120
EV3.6	3,189,423	NG4.8	51
EV3.7	1,109,570	NG4.9	2,401,445
EV3.8	668,157	SG1.1	5,180,336
EV3.9	1,092,528	SG1.2	1,712,296
NG1.1	770,713	SG1.3	1,911,670
NG1.2	2,597,091	SG1.4	2,439,056
NG1.3	1,924,354	SG1.5	1,799,986
NG1.4	2,166,534	SG1.6	2,113,195
NG1.5	2,215,520	SG1.7	1,925,724
NG1.6	3,093,949	SG1.8	5,161,272
NG1.7	2,013,517	SG1.9	6,085,443

Sample ID	Total reads counts	Sample ID	Total reads counts
SG2.1	2,268,824	WR3.2	2,472,970
SG2.2	5,633,717	WR3.3	779,652
SG2.3	2,143,533	WR3.4	2,470,207
SG2.4	2,310,738	WR3.5	1,592,113
SG2.5	2,472,581	WR3.6	1,940,461
SG2.6	6,188,941	WR3.7	2,656,725
SG2.7	3,959,624	WR3.8	1,042,212
SG2.8	2,464,846	WR3.9	1,824,551
SG2.9	841,074	WR4.1	1,889,052
SG3.1	2,891,508	WR4.2	1,602,806
SG3.2	15,276	WR4.3	753,961
SG3.3	1,751,734	WR4.4	5,464,315
SG3.4	2,133,374	WR4.5	5,547,819
SG3.5	4,025,856	WR4.6	0
SG3.6	1,984,150	WR4.7	2,957,734
SG3.7	4,218,769	WR4.8	0
SG3.8	3,815,071	WR4.9	2,923,008
SG3.9	1,439,674	undetermined_1	10,964,300
WR1.1	2,141,752	undetermined_2	11,926,174
WR1.2	3,428,655	Water1	554
WR1.3	964,166	Water2	2,350
WR1.4	2,459,003		
WR1.5	2,997,652		
WR1.6	3,700,487		
WR1.7	2,706,071		
WR1.8	1,918,646		
WR1.9	1,772,633		
WR2.1	3,021,839		
WR2.2	4,298,296		
WR2.3	1,312,352		
WR2.4	3,913,119		
WR2.5	1,418,377		
WR2.6	2,594,070		
WR2.7	3,835,492		
WR2.8	2,969,692		
WR2.9	3,292,219	Total counts	484,776,478
WR3.1	4,077,089	Average counts per sample	2,448,366

Additional file 2, Table S2. Mapped RNA-seq data for leaf samples to the *V. vinifera* to the reference genome (IGGP_12X.V1).

Samples	Reads leaf samples	Samples	Reads leaf samples
CG1.1	25,743,097	EE1.3	26,985,753
CG1.2	27,225,555	EE2.1	42,939,482
CG1.3	26,383,747	EE2.2	29,521,962
CG2.1	20,726,878	EE2.3	25,145,787
CG2.2	37,326,602	EE3.1	29,216,096
CG2.3	20,330,612	EE3.2	18,263,688
CG3.1	25,863,809	EE3.3	21,728,373
CG3.2	20,856,760	EE4.1	11,557,040
CG3.3	19,566,185	EE4.2	13,484,751
CG4.1	25,342,050	EE4.3	8,873,232
CG4.2	22,221,312	SG1.1	15,484,201
CG4.3	22,097,859	SG1.2	17,223,812
EV1.1	23,056,320	SG1.3	17,162,388
EV1.2	25,524,736	SG2.1	16,276,961
EV1.3	17,605,976	SG2.2	18,392,528
EV2.1	25,353,854	SG2.3	16,084,122
EV2.2	13,559,924	SG3.1	14,614,325
EV2.3	21,162,499	SG3.2	17,965,039
EV3.1	10,817,673	SG3.3	23,806,350
EV3.2	21,138,475	WR1.1	16,338,124
EV3.3	25,676,349	WR1.2	20,531,631
NG1.1	29,778,866	WR1.3	42,524,626
NG1.2	24,798,843	WR2.1	21,867,390
NG1.3	54,236,656	WR2.2	21,709,202
NG2.1	24,638,270	WR2.3	17,494,025
NG2.2	29,024,880	WR3.1	21,713,180
NG2.3	18,360,950	WR3.2	18,808,425
NG3.1	32,656,908	WR3.3	23,660,660
NG3.2	15,254,356	WR4.1	20,540,628
NG3.3	21,842,503	WR4.2	44,712,982
NG4.1	18,310,992	WR4.3	17,825,928
NG4.2	27,310,080		
NG4.3	19,468,484	total	1,526,917,359
EE1.1	31,012,663	average	23,135,111.50
EE1.2	30,189,945		



Additional file 3. Discriminant analysis of principal components (DAPC) for the identification of genetics clusters from grapevine leaf samples from twenty-two vineyards from the Barossa Wine Zone. The optimal number of clusters is calculated by using k-means with increasing values of k (number of groups) that maximize the variation between groups. The different numbers of clustering options are compared using Bayesian information criterion (BIC), where the optimal number of clusters correspond to the lowest BIC.

Additional file 4, Table S1: Summary of the pairwise comparisons of the gene expression, differentially methylated markers, differentially methylated genes and genes that are both differentially methylated and expressed between sub-regions.

*Significance level FDR < 0.05.

^Genes that where +/- 5kb from a differentially methylated marker.

° Intersection of genes that where differentially expressed and methylated.

Region 1	Region 2	Number of differentially expressed genes*	Number of differentially methylated markers*	Number of differentially methylated genes*^	Number of differentially methylated and expressed genes°
CG	EE	5,822	3,468	417	417
CG	EV	957	2,393	61	61
CG	NG	8,405	6,751	1,353	1,353
CG	SG	4,949	18,780	1,604	1,604
CG	WR	3,772	20,530	1,267	1,267
EE	EV	5,242	2	1	1
EE	NG	1,496	1,852	79	79
EE	SG	1,084	770	40	40
EE	WR	3,248	676	63	63
EV	NG	6,638	237	40	40
EV	SG	3,610	497	74	74
EV	WR	4,130	1,231	104	104
NG	SG	380	13,695	113	113
NG	WR	2,652	16,902	1,007	1,007
SG	WR	759	6	-	-

Additional file 4, Table S2: Summary of pairwise comparison of the gene expression and DNA methylation data. DEGs: Differentially expressed genes, obtained using the R package EdgeR (p-values < 0.05). DMMs: Differentially methylated markers, obtained using the R package msgbsR (p-values <0.05). DMEGs: Differentially methylated and expressed genes obtained from the intersection between DEGs and DMMs. CG: Central Grounds, EE: Eastern Edge, EV: Eden Valley, NG: Northern Grounds, SG: Southern Grounds and WR: Western Ridge

	Number of:			Location of the methylation:			Total unique genes
Combination	**DEGs**	**DMMs**	**DMEGs**	**upstream**	**down**	**in the gene**	
CG vs EE	5822	3468	574	93	96	385	417
CG vs EV	957	2393	67	11	12	44	61
CG vs NG	8405	6751	1596	277	270	1049	1353
CG vs SG	4949	18780	2483	494	422	1567	1604
CG vs WR	3772	20530	1994	358	351	1285	1267
EE vs EV	5242	2	1	0	0	1	1
EE vs NG	1496	1852	90	15	7	68	79
EE vs SG	1084	770	41	7	5	29	40
EE vs WR	3248	676	65	11	10	44	63
EV vs NG	6638	237	41	5	9	27	40
EV vs SG	3610	497	76	10	4	62	74
EV vs WR	4130	1231	167	33	32	102	104
NG vs SG	380	13695	157	36	28	93	113
NG vs WR	2652	16902	1572	246	228	1098	1007
SG vs WR	759	6	0	0	0	0	0

Additional file 5, Table S1. Genes that their methylation and expression significantly (FDR-adjusted p-values < 0.05) correlate with planting year obtained by fitting a generalized linear model. Grapevine gene function, gene ontology motifs and best *Arabidopsis* match was obtained from Cramer et al. (2020).

Gene ID	Description	GOs	GOs description	Best *Arabidopsis* match	*Arabidopsis* gene description
VIT_05s0049g00160	enoyl- hydratase isomerase family	F:GO:0003824, P:GO:0008152	F:catalytic activity, P:metabolic process	AT4G14430.1	indole-3-butyric acid response 10
VIT_18s0122g00780	phosphate transporter family	C:GO:0005634, P:GO:0007154, P:GO:0009991, P:GO:0009719, C:GO:0005794, F:GO:0005488, C:GO:0005773, C:GO:0005886, F:GO:0005215, P:GO:0006950, F:GO:0016787	C:nucleus, P:cell communication, P:response to extracellular stimulus, P:response to endogenous stimulus, C:Golgi apparatus, F:binding, C:vacuole, C:plasma membrane, F:transporter activity, P:response to stress, F:hydrolase activity	AT1G76430.1	phosphate transporter 1%3B9
VIT_03s0038g03530	transmembrane amino acid transporter family	P:GO:0007267, F:GO:0004872, F:GO:0016740, F:GO:0004871, P:GO:0006139, C:GO:0005886, F:GO:0005215	P:cell-cell signaling, F:receptor activity, F:transferase activity, F:signal transducer activity, P:nucleobase-containing compound metabolic process, C:plasma membrane, F:transporter activity	AT1G47670.1	Transmembrane amino acid transporter family protein
VIT_04s0023g01250	BES1 BZR1 homolog 2-like	P:GO:0007165, P:GO:0009058, C:GO:0005773, P:GO:0006139, P:GO:0040007, P:GO:0006950, F:GO:0003677, C:GO:0005634, P:GO:0009607, F:GO:0003700, P:GO:0009605, P:GO:0009908, P:GO:0009719, C:GO:0005829	P:signal transduction, P:biosynthetic process, C:vacuole, P:nucleobase-containing compound metabolic process, P:growth, P:response to stress, F:DNA binding, C:nucleus, P:response to biotic stimulus, F:transcription factor activity, sequence-specific DNA binding, P:response to external stimulus, P:flower development, P:response to endogenous stimulus, C:cytosol	AT1G75080.1	Brassinosteroid signaling positive regulator (BZR1) family protein
VIT_01s0011g05030	fold protein	P:GO:0009607, C:GO:0016020, P:GO:0006950	P:response to biotic stimulus, C:membrane, P:response to stress	AT1G23110.1	fold protein
VIT_04s0023g02650	3-epi-6-deoxocathasterone 23-monooxygenase-like	F:GO:0003824, P:GO:0009908, C:GO:0016020, P:GO:0006629, P:GO:0016043, P:GO:0009653, P:GO:0009058, F:GO:0005488, P:GO:0016049	F:catalytic activity, P:flower development, C:membrane, P:lipid metabolic process, P:cellular component organization, P:anatomical structure morphogenesis, P:biosynthetic process, F:binding, P:cell growth	AT4G36380.1	Cytochrome P450 superfamily protein
VIT_06s0061g01170	adaptor complex AP-alpha subunit	C:GO:0005737, C:GO:0016020, F:GO:0005488, F:GO:0005215	C:cytoplasm, C:membrane, F:binding, F:transporter activity	AT5G22770.1	alpha-adaptin
VIT_18s0001g11040	heme oxygenase 2	P:GO:0009056, P:GO:0007165, P:GO:0009791, P:GO:0009058, C:GO:0009536, P:GO:0006950, P:GO:0040007, P:GO:0000003, F:GO:0003677, P:GO:0009628, F:GO:0003824, C:GO:0016020, P:GO:0006629	P:catabolic process, P:signal transduction, P:post-embryonic development, P:biosynthetic process, C:plastid, P:response to stress, P:growth, P:reproduction, F:DNA binding, P:response to abiotic stimulus, F:catalytic activity, C:membrane, P:lipid metabolic process	AT2G26550.2	heme oxygenase 2
VIT_17s0000g02860	uncharacterized protein LOC109814835	P:GO:0019725, C:GO:0005737, P:GO:0007275, P:GO:0030154, F:GO:0005488, C:GO:0005886	P:cellular homeostasis, C:cytoplasm, P:multicellular organism development, P:cell differentiation, F:binding, C:plasma membrane	ATMG01250.1	RNA-directed DNA polymerase (reverse transcriptase)

VIT_14s0068g00010	Receptor kinase FERONIA	P:GO:0009791, P:GO:0006464, F:GO:0004871, P:GO:0009856, C:GO:0005622, C:GO:0005886, P:GO:0006950, F:GO:0000166, P:GO:0009607, F:GO:0005515, P:GO:0009605, C:GO:0005618, P:GO:0009719, P:GO:0016043, F:GO:0016301, P:GO:0016049	P:post-embryonic development, P:cellular protein modification process, F:signal transducer activity, P:pollination, C:intracellular, C:plasma membrane, P:response to stress, F:nucleotide binding, P:response to biotic stimulus, F:protein binding, P:response to external stimulus, C:cell wall, P:response to endogenous stimulus, P:cellular component organization, F:kinase activity, P:cell growth	AT3G51550.1	Malectin/receptor-like protein kinase family protein
VIT_06s0004g06690	ribosomal S18	P:GO:0006629, F:GO:0005198, C:GO:0005829, C:GO:0005840, F:GO:0016740, P:GO:0006412	P:lipid metabolic process, F:structural molecule activity, C:cytosol, C:ribosome, F:transferase activity, P:translation	AT1G07210.1	Ribosomal protein S18
VIT_01s0011g03830	metallo-beta-lactamase superfamily	No GO term	No GO term	AT1G25375.1	Metallo-hydrolase/oxidoreductase superfamily protein
VIT_00s0125g00060	Uncharacterized protein	No GO term	No GO term		
VIT_06s0061g00610	actin 2 3 complex subunit	No GO term	No GO term	AT2G28130.1	actin protein 2/3 complex subunit-like protein
VIT_04s0023g01890	embryonic DC-8	No GO term	No GO term	-	-
VIT_18s0001g01350	Uncharacterized protein	No GO term	No GO term	-	-
VIT_05s0020g01560	pentatricopeptide repeat-containing family	F:GO:0000166, F:GO:0003723, F:GO:0005515, P:GO:0007165, F:GO:0004518, P:GO:0006464, F:GO:0004872, F:GO:0016301, C:GO:0005886, C:GO:0005622	F:nucleotide binding, F:RNA binding, F:protein binding, P:signal transduction, F:nuclease activity, P:cellular protein modification process, F:receptor activity, F:kinase activity, C:plasma membrane, C:intracellular	AT4G15720.1	Tetratricopeptide repeat (TPR)-like superfamily protein
VIT_17s0000g05010	F-box CPR30	F:GO:0005515, P:GO:0009056, P:GO:0006950, F:GO:0016787	F:protein binding, P:catabolic process, P:response to stress, F:hydrolase activity	AT5G49610.1	F-box family protein
VIT_00s0505g00040	Uncharacterized protein	No GO term	No GO term	-	-
VIT_05s0077g00630	acidic leucine-rich nuclear phospho 32-related 1-like	C:GO:0005829	C:cytosol	-	-
VIT_17s0000g06650	probable serine threonine- kinase At1g54610	F:GO:0000166, C:GO:0005634, F:GO:0005515, P:GO:0006464, F:GO:0016301, C:GO:0005886, C:GO:0009536, P:GO:0007049	F:nucleotide binding, C:nucleus, F:protein binding, P:cellular protein modification process, F:kinase activity, C:plasma membrane, C:plastid, P:cell cycle	AT1G18670.1	Protein kinase superfamily protein
VIT_01s0011g02740	phosphoenolpyruvate carboxylase 4-like	P:GO:0019538, P:GO:0006091, P:GO:0009056, P:GO:0007275, P:GO:0007165, C:GO:0005794, P:GO:0009058, F:GO:0005102, C:GO:0005886, F:GO:0005215, P:GO:0006950, P:GO:0005975, P:GO:0008219, P:GO:0030154, C:GO:0005829, F:GO:0016740	P:protein metabolic process, P:generation of precursor metabolites and energy, P:catabolic process, P:multicellular organism development, P:signal transduction, C:Golgi apparatus, P:biosynthetic process, F:receptor binding, C:plasma membrane, F:transporter activity, P:response to stress, P:carbohydrate metabolic process, P:cell death, P:cell differentiation, C:cytosol, F:transferase activity	AT1G68750.1	phosphoenolpyruvate carboxylase 4
VIT_03s0038g04040	N-acyl-L-amino-acid amidohydrolase	P:GO:0019538, F:GO:0005515, C:GO:0016020, C:GO:0005783, C:GO:0005794, C:GO:0005773,	P:protein metabolic process, F:protein binding, C:membrane, C:endoplasmic reticulum, C:Golgi apparatus, C:vacuole, P:cellular process, F:hydrolase activity	AT4G38220.1	Peptidase M20/M25/M40 family protein

		P:GO:0009987, F:GO:0016787			
VIT_17s0000g05550	peptide nitrate transporter	P:GO:0009056, P:GO:0007165, P:GO:0009653, F:GO:0005215, P:GO:0040029, F:GO:0003676, C:GO:0005634, C:GO:0005737, C:GO:0016020, F:GO:0004518, P:GO:0009719, P:GO:0016043, P:GO:0016049	P:catabolic process, P:signal transduction, P:anatomical structure morphogenesis, F:transporter activity, P:regulation of gene expression, epigenetic, F:nucleic acid binding, C:nucleus, C:cytoplasm, C:membrane, F:nuclease activity, P:response to endogenous stimulus, P:cellular component organization, P:cell growth	AT5G28470.1	Major facilitator superfamily protein
VIT_00s0391g00030	EPIDERMAL PATTERNING FACTOR 1	F:GO:0003676, C:GO:0005635, P:GO:0005975, F:GO:0005515, P:GO:0009791, C:GO:0016020, P:GO:0030154, C:GO:0005829, P:GO:0009653, C:GO:0005576, F:GO:0005215, F:GO:0016787	F:nucleic acid binding, C:nuclear envelope, P:carbohydrate metabolic process, F:protein binding, P:post-embryonic development, C:membrane, P:cell differentiation, C:cytosol, P:anatomical structure morphogenesis, C:extracellular region, F:transporter activity, F:hydrolase activity	AT5G10310.1	EPIDERMAL PATTERNING FACTOR-like protein
VIT_00s0361g00040	anthocyanidin reductase	C:GO:0005737, P:GO:0007165, P:GO:0007267, C:GO:0005576, F:GO:0005102, P:GO:0006810, C:GO:0005886	C:cytoplasm, P:signal transduction, P:cell-cell signaling, C:extracellular region, F:receptor binding, P:transport, C:plasma membrane	AT1G61720.1	NAD(P)-binding Rossmann-fold superfamily protein
VIT_05s0020g03460	RNA recognition motif XS domain	F:GO:0003676, C:GO:0016020, P:GO:0008152, C:GO:0005730, P:GO:0009987	F:nucleic acid binding, C:membrane, P:metabolic process, C:nucleolus, P:cellular process	AT3G22430.1	RNA recognition motif XS domain protein
VIT_07s0031g00030	Rer1 family	P:GO:0019538, F:GO:0005515, C:GO:0016020, C:GO:0005783, C:GO:0005794, P:GO:0006810, F:GO:0016787	P:protein metabolic process, F:protein binding, C:membrane, C:endoplasmic reticulum, C:Golgi apparatus, P:transport, F:hydrolase activity	AT4G39220.1	Rer1 family protein
VIT_14s0030g00680	ribosomal S4 (chloroplast)	F:GO:0003723, F:GO:0005515, F:GO:0005198, C:GO:0005840, C:GO:0009536, P:GO:0006412	F:RNA binding, F:protein binding, F:structural molecule activity, C:ribosome, C:plastid, P:translation	ATCG00380.1	chloroplast ribosomal protein S4
VIT_08s0007g02870	mRNA-decapping enzyme	F:GO:0003723, C:GO:0005737, P:GO:0009056, F:GO:0030234, P:GO:0006139	F:RNA binding, C:cytoplasm, P:catabolic process, F:enzyme regulator activity, P:nucleobase-containing compound metabolic process	AT1G08370.1	decapping 1
VIT_04s0044g01920	FAR1-RELATED SEQUENCE 5	P:GO:0009058, F:GO:0005488, P:GO:0006139	P:biosynthetic process, F:binding, P:nucleobase-containing compound metabolic process	AT4G38180.1	FAR1-related sequence 5
VIT_05s0051g00910	Non-lysosomal glucosylceramidase	P:GO:0005975, P:GO:0009056, P:GO:0006629, C:GO:0005794, C:GO:0005886, P:GO:0009987, F:GO:0016787	P:carbohydrate metabolic process, P:catabolic process, P:lipid metabolic process, C:Golgi apparatus, C:plasma membrane, P:cellular process, F:hydrolase activity	AT3G24180.1	Beta-glucosidase%2C GBA2 type family protein
VIT_07s0104g01300	mobidb-lite (MOBIDB_LITE)	No GO term	No GO term	-	-
VIT_07s0005g00070	bifunctional 2	C:GO:0005829, P:GO:0009058, F:GO:0016740, F:GO:0005488, P:GO:0006259, C:GO:0009536, F:GO:0016787	C:cytosol, P:biosynthetic process, F:transferase activity, F:binding, P:DNA metabolic process, C:plastid, F:hydrolase activity	AT3G12290.1	Amino acid dehydrogenase family protein
VIT_16s0050g00450	DNA-directed RNA polymerase III subunit RPC3	F:GO:0003677, P:GO:0009058, F:GO:0016740, C:GO:0005730, P:GO:0006139, C:GO:0009536, C:GO:0005654	F:DNA binding, P:biosynthetic process, F:transferase activity, C:nucleolus, P:nucleobase-containing compound metabolic process, C:plastid, C:nucleoplasm	AT3G49000.1	RNA polymerase III subunit RPC82 family protein
VIT_02s0012g01330	ADP-ribosylation	P:GO:0009790, P:GO:0009791,	P:embryo development, P:post-embryonic development,	AT4G17890.1	ARF-GAP domain 8

	factor GTPase-activating AGD10	P:GO:0009719, P:GO:0009058, F:GO:0005488, P:GO:0006139, F:GO:0030234, P:GO:0000003, F:GO:0016787	P:response to endogenous stimulus, P:biosynthetic process, F:binding, P:nucleobase-containing compound metabolic process, F:enzyme regulator activity, P:reproduction, F:hydrolase activity		
VIT_00s0753g00010	TMhelix (TMHMM), TMhelix (TMHMM), TMhelix (TMHMM)	F:GO:0000166, C:GO:0005737, C:GO:0016020, C:GO:0005783, P:GO:0009058, P:GO:0006139, F:GO:0005215, F:GO:0016787	F:nucleotide binding, C:cytoplasm, C:membrane, C:endoplasmic reticulum, P:biosynthetic process, P:nucleobase-containing compound metabolic process, F:transporter activity, F:hydrolase activity	AT1G49420.1	Heavy metal transport/detoxification superfamily protein
VIT_13s0019g05290	acyl- -binding domain-containing 4	C:GO:0005634, F:GO:0005515, C:GO:0016020, C:GO:0005829	C:nucleus, F:protein binding, C:membrane, C:cytosol	AT5G04420.1	Galactose oxidase/kelch repeat superfamily protein
VIT_19s0014g01300	histidyl-tRNA synthetase	F:GO:0000166, P:GO:0009790, P:GO:0009791, P:GO:0006464, F:GO:0016301, C:GO:0009536, P:GO:0000003	F:nucleotide binding, P:embryo development, P:post-embryonic development, P:cellular protein modification process, F:kinase activity, C:plastid, P:reproduction	AT3G14900.1	hypothetical protein
VIT_17s0000g07640	ADP-glucose pyrophosphorylase family	C:GO:0005777, F:GO:0003677, P:GO:0005975, C:GO:0016020, F:GO:0005198, C:GO:0005829, C:GO:0005840, F:GO:0016740, C:GO:0009536, F:GO:0005215, P:GO:0006412, C:GO:0009579	C:peroxisome, F:DNA binding, P:carbohydrate metabolic process, C:membrane, F:structural molecule activity, C:cytosol, C:ribosome, F:transferase activity, C:plastid, F:transporter activity, P:translation, C:thylakoid	AT1G74910.1	ADP-glucose pyrophosphorylase family protein
VIT_12s0059g00760	5 -adenylylsulfate reductase chloroplastic-like	F:GO:0000166, C:GO:0005777, P:GO:0019725, F:GO:0003824, P:GO:0006629, P:GO:0009058	F:nucleotide binding, C:peroxisome, P:cellular homeostasis, F:catalytic activity, P:lipid metabolic process, P:biosynthetic process	AT4G21990.1	APS reductase 3
VIT_00s1287g00010	Uncharacterized protein	No GO term	No GO term		
VIT_04s0008g01310	serine threonine-kinase EDR1	F:GO:0000166, P:GO:0006464, F:GO:0016301	F:nucleotide binding, P:cellular protein modification process, F:kinase activity	AT5G11850.1	Protein kinase superfamily protein
VIT_06s0004g01740	transcription factor bHLH122	F:GO:0003677, C:GO:0005634, P:GO:0005975, F:GO:0003700, P:GO:0009628, F:GO:0005515, F:GO:0030528, P:GO:0009791, P:GO:0009058, P:GO:0006139, F:GO:0016301, P:GO:0000003	F:DNA binding, C:nucleus, P:carbohydrate metabolic process, F:transcription factor activity, sequence-specific DNA binding, P:response to abiotic stimulus, F:protein binding, F:obsolete transcription regulator activity, P:post-embryonic development, P:biosynthetic process, P:nucleobase-containing compound metabolic process, F:kinase activity, P:reproduction	AT2G42280.1	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
VIT_01s0010g00690	DNA-directed RNA polymerase mitochondrial-like	F:GO:0003677, F:GO:0003700, F:GO:0003723, C:GO:0005829, P:GO:0009058, C:GO:0005576, F:GO:0016740, C:GO:0005730, P:GO:0006139	F:DNA binding, F:transcription factor activity, sequence-specific DNA binding, F:RNA binding, C:cytosol, P:biosynthetic process, C:extracellular region, F:transferase activity, C:nucleolus, P:nucleobase-containing compound metabolic process	AT1G68990.1	male gametophyte defective 3
VIT_15s0046g02440	plant-specific TFIIB-related PTF2	P:GO:0009790, P:GO:0007154, P:GO:0009991, P:GO:0009791, F:GO:0008135, C:GO:0005840, P:GO:0006139, P:GO:0009856, C:GO:0009536, P:GO:0006950, F:GO:0003677, C:GO:0005634, P:GO:0009628, F:GO:0005515, P:GO:0009719, P:GO:0016043	P:embryo development, P:cell communication, P:response to extracellular stimulus, P:post-embryonic development, F:translation factor activity, RNA binding, C:ribosome, P:nucleobase-containing compound metabolic process, P:pollination, C:plastid, P:response to stress, F:DNA binding, C:nucleus, P:response to abiotic stimulus, F:protein binding, P:response to endogenous stimulus, P:cellular component organization	AT4G35540.1	POLLEN-EXPRESSED TRANSCRIPTION FACTOR 2
VIT_00s0441g00010	DEAD-box ATP-dependent RNA	F:GO:0000166, P:GO:0019748, F:GO:0003723, P:GO:0009056,	F:nucleotide binding, P:secondary metabolic process, F:RNA binding, P:catabolic process, C:extracellular	AT4G34910.1	P-loop containing nucleoside triphosphate hydrolases

	helicase 16	C:GO:0005576, P:GO:0006139, F:GO:0016787	region, P:nucleobase-containing compound metabolic process, F:hydrolase activity		superfamily protein
VIT_06s0004g01210	PRLI-interacting factor	F:GO:0000166, C:GO:0005622	F:nucleotide binding, C:intracellular	AT5G19900.1	PRLI-interacting factor
VIT_08s0040g01860	senescence regulator	F:GO:0003677, P:GO:0009058, F:GO:0016740, C:GO:0005730, P:GO:0006139	F:DNA binding, P:biosynthetic process, F:transferase activity, C:nucleolus, P:nucleobase-containing compound metabolic process	AT5G60680.1	transcription initiation factor TFIID subunit (Protein of unknown function%2C DUF584)
VIT_00s0246g00040	Uncharacterized protein	No GO term	No GO term	-	-
VIT_15s0048g01310	Uncharacterized protein	No GO term	No GO term	-	-
VIT_00s0629g00010	Uncharacterized protein	No GO term	No GO term	-	-
VIT_13s0067g01550	inactive receptor-like serine threonine-kinase At2g40270	F:GO:0000166, P:GO:0006464, F:GO:0004872, F:GO:0016301, F:GO:0004871, C:GO:0005886	F:nucleotide binding, P:cellular protein modification process, F:receptor activity, F:kinase activity, F:signal transducer activity, C:plasma membrane	AT3G56050.1	Protein kinase family protein
VIT_18s0001g00470	multi-copper oxidase	P:GO:0005975, F:GO:0003824, C:GO:0005618, C:GO:0016020, C:GO:0005576, F:GO:0005488, P:GO:0009987	P:carbohydrate metabolic process, F:catalytic activity, C:cell wall, C:membrane, C:extracellular region, F:binding, P:cellular process	AT1G76160.1	SKU5 similar 5
VIT_00s1348g00010	Uncharacterized protein	No GO term	No GO term	-	-
VIT_08s0007g02020	NRT1 PTR FAMILY -like	F:GO:0000166, F:GO:0003676, C:GO:0016020, F:GO:0005215	F:nucleotide binding, F:nucleic acid binding, C:membrane, F:transporter activity	AT2G02040.1	peptide transporter 2
VIT_00s0372g00040	(3S,6E)-nerolidol synthase chloroplastic-like	P:GO:0019725, P:GO:0019538, F:GO:0005515, C:GO:0005737, F:GO:0003824, P:GO:0009056, P:GO:0016043	P:cellular homeostasis, P:protein metabolic process, F:protein binding, C:cytoplasm, F:catalytic activity, P:catabolic process, P:cellular component organization	AT1G61680.1	terpene synthase 14
VIT_00s0246g00110	Uncharacterized protein	No GO term	No GO term	-	-
VIT_06s0004g04600	phosphatase 2C family	F:GO:0000166, P:GO:0006464, C:GO:0005622, F:GO:0016787	F:nucleotide binding, P:cellular protein modification process, C:intracellular, F:hydrolase activity	AT2G28890.1	poltergeist like 4
VIT_04s0008g01910	hypothetical protein L484_018601	F:GO:0000166, F:GO:0003824, C:GO:0005739, C:GO:0016020, P:GO:0006629, P:GO:0016043, P:GO:0009058, C:GO:0005576, P:GO:0006810	F:nucleotide binding, F:catalytic activity, C:mitochondrion, C:membrane, P:lipid metabolic process, P:cellular component organization, P:biosynthetic process, C:extracellular region, P:transport	AT1G02180.1	ferredoxin-like protein
VIT_08s0007g08380	cellulose synthase family	C:GO:0005794, P:GO:0009058, P:GO:0006139, C:GO:0005886, C:GO:0009536, P:GO:0007049, P:GO:0006950, F:GO:0000166, P:GO:0005975, P:GO:0009628, C:GO:0005768, P:GO:0016043, F:GO:0016301	C:Golgi apparatus, P:biosynthetic process, P:nucleobase-containing compound metabolic process, C:plasma membrane, C:plastid, P:cell cycle, P:response to stress, F:nucleotide binding, P:carbohydrate metabolic process, P:response to abiotic stimulus, C:endosome, P:cellular component organization, F:kinase activity	AT5G05170.1	Cellulose synthase family protein
VIT_18s0001g07140	zinc-finger homeodomain 1-like	P:GO:0019748, P:GO:0009791, P:GO:0006464, P:GO:0009875, P:GO:0009058, P:GO:0006139, F:GO:0000166, F:GO:0003677,	P:secondary metabolic process, P:post-embryonic development, P:cellular protein modification process, P:pollen-pistil interaction, P:biosynthetic process, P:nucleobase-containing compound metabolic process,	AT4G24660.1	homeobox protein 22

		C:GO:0005634, F:GO:0005515, C:GO:0005856, P:GO:0006629, C:GO:0016020, F:GO:0030246, F:GO:0016301	F:nucleotide binding, F:DNA binding, C:nucleus, F:protein binding, C:cytoskeleton, P:lipid metabolic process, C:membrane, F:carbohydrate binding, F:kinase activity		
VIT_15s0048g00680	LRR receptor-like kinase	F:GO:0000166, F:GO:0005515, P:GO:0006464, F:GO:0016301, P:GO:0006139, C:GO:0005886	F:nucleotide binding, F:protein binding, P:cellular protein modification process, F:kinase activity, P:nucleobase-containing compound metabolic process, C:plasma membrane	AT2G45340.1	Leucine-rich repeat protein kinase family protein
VIT_17s0000g03870	serine threonine-phosphatase 2A 65 kDa regulatory subunit A beta isoform	C:GO:0005634, F:GO:0005515, P:GO:0007165, C:GO:0005618, P:GO:0009719, P:GO:0006464, C:GO:0005829, P:GO:0006810, C:GO:0005886, P:GO:0006950, F:GO:0016787	C:nucleus, F:protein binding, P:signal transduction, C:cell wall, P:response to endogenous stimulus, P:cellular protein modification process, C:cytosol, P:transport, C:plasma membrane, P:response to stress, F:hydrolase activity	AT3G25800.1	protein phosphatase 2A subunit A2
VIT_11s0118g00210	calmodulin-binding receptor-like cytoplasmic kinase 3	F:GO:0000166, C:GO:0016020, P:GO:0006464, F:GO:0016301	F:nucleotide binding, C:membrane, P:cellular protein modification process, F:kinase activity	AT2G11520.1	calmodulin-binding receptor-like cytoplasmic kinase 3
VIT_05s0049g00920	inorganic phosphate transporter 1-4-like	F:GO:0000166, C:GO:0016020, F:GO:0005215, F:GO:0016787	F:nucleotide binding, C:membrane, F:transporter activity, F:hydrolase activity	AT2G38940.1	phosphate transporter 1%3B4
VIT_11s0052g01660	alpha beta-Hydrolases superfamily	No GO term	No GO term	AT4G25770.1	alpha/beta-Hydrolases superfamily protein
VIT_09s0002g00590	phosphoinositide phosphatase SAC4	F:GO:0003677, P:GO:0006629, P:GO:0016043, P:GO:0009058, F:GO:0016787	F:DNA binding, P:lipid metabolic process, P:cellular component organization, P:biosynthetic process, F:hydrolase activity	AT3G14205.1	Phosphoinositide phosphatase family protein
VIT_02s0025g00510	respiratory burst oxidase homolog A-like	P:GO:0006091, P:GO:0007165, P:GO:0009791, C:GO:0005794, P:GO:0009653, P:GO:0009058, F:GO:0003682, P:GO:0006259, C:GO:0005886, P:GO:0006950, P:GO:0009607, C:GO:0005634, F:GO:0003677, P:GO:0009628, P:GO:0008219, P:GO:0009605, F:GO:0005515, F:GO:0003824, P:GO:0009719, P:GO:0016043, P:GO:0030154	P:generation of precursor metabolites and energy, P:signal transduction, P:post-embryonic development, C:Golgi apparatus, P:anatomical structure morphogenesis, P:biosynthetic process, F:chromatin binding, P:DNA metabolic process, C:plasma membrane, P:response to stress, P:response to biotic stimulus, C:nucleus, F:DNA binding, P:response to abiotic stimulus, P:cell death, P:response to external stimulus, F:protein binding, F:catalytic activity, P:response to endogenous stimulus, P:cellular component organization, P:cell differentiation	AT1G64060.1	respiratory burst oxidase protein F
VIT_01s0011g06090	THO complex subunit 2	P:GO:0005975, F:GO:0003723, C:GO:0005618, C:GO:0005576, F:GO:0016740, P:GO:0006139, P:GO:0006810, P:GO:0040029, C:GO:0005654, P:GO:0006412, F:GO:0016787	P:carbohydrate metabolic process, F:RNA binding, C:cell wall, C:extracellular region, F:transferase activity, P:nucleobase-containing compound metabolic process, P:transport, P:regulation of gene expression, epigenetic, C:nucleoplasm, P:translation, F:hydrolase activity	AT1G24706.7	THO2
VIT_06s0004g01420	glycoside hydrolase family 1	P:GO:0019748, P:GO:0009056, F:GO:0004872, C:GO:0005576, F:GO:0004871, C:GO:0005886, F:GO:0005215, P:GO:0006950, P:GO:0005975, P:GO:0009628, C:GO:0005618, P:GO:0009719, P:GO:0006629, F:GO:0016787	P:secondary metabolic process, P:catabolic process, F:receptor activity, C:extracellular region, F:signal transducer activity, C:plasma membrane, F:transporter activity, P:response to stress, P:carbohydrate metabolic process, P:response to abiotic stimulus, C:cell wall, P:response to endogenous stimulus, P:lipid metabolic process, F:hydrolase activity	AT5G42260.1	beta glucosidase 12

VIT_18s0166g00210	E3 ubiquitin- ligase HERC2-like	F:GO:0003677, C:GO:0005634, P:GO:0006464, F:GO:0016740, C:GO:0005886	F:DNA binding, C:nucleus, P:cellular protein modification process, F:transferase activity, C:plasma membrane	AT1G76950.1	Regulator of chromosome condensation (RCC1) family with FYVE zinc finger domain-containing protein
VIT_16s0098g01770	SNARE-like superfamily	No GO term	No GO term	AT5G52990.1	SNARE-like superfamily protein
VIT_12s0034g01110	DUF3128 family	F:GO:0000166, P:GO:0009628, C:GO:0016020, P:GO:0006464, F:GO:0016301, P:GO:0006950	F:nucleotide binding, P:response to abiotic stimulus, C:membrane, P:cellular protein modification process, F:kinase activity, P:response to stress	AT4G25315.2	Expressed protein
VIT_08s0040g01780	cytokinin riboside 5 - monophosphate phosphoribohydrolase	C:GO:0005634, C:GO:0005829, P:GO:0009058, P:GO:0009987, F:GO:0016787	C:nucleus, C:cytosol, P:biosynthetic process, P:cellular process, F:hydrolase activity	AT2G37210.1	lysine decarboxylase family protein
VIT_01s0011g04680	glycosyltransferase family 14 family	P:GO:0019538, C:GO:0016020, C:GO:0005794, P:GO:0009058, F:GO:0005488, F:GO:0016740, P:GO:0016049	P:protein metabolic process, C:membrane, C:Golgi apparatus, P:biosynthetic process, F:binding, F:transferase activity, P:cell growth	AT1G71070.1	Core-2/I-branching beta-1%2C6-N-acetylglucosaminyltransferase family protein
VIT_04s0023g01880	zinc finger family	P:GO:0019538, F:GO:0003677, F:GO:0003700, C:GO:0005615, F:GO:0005515, P:GO:0009719, P:GO:0016043, P:GO:0009058, F:GO:0016740, P:GO:0006139, C:GO:0005654	P:protein metabolic process, F:DNA binding, F:transcription factor activity, sequence-specific DNA binding, C:extracellular space, F:protein binding, P:response to endogenous stimulus, P:cellular component organization, P:biosynthetic process, F:transferase activity, P:nucleobase-containing compound metabolic process, C:nucleoplasm	AT2G18350.1	homeobox protein 24
VIT_06s0004g02880	probable serine threonine- kinase PBL28	F:GO:0000166, P:GO:0006464, F:GO:0016301	F:nucleotide binding, P:cellular protein modification process, F:kinase activity	AT3G26700.1	Protein kinase superfamily protein
VIT_04s0023g03750	L-ascorbate peroxidase peroxisomal	P:GO:0009056, P:GO:0009791, P:GO:0009653, C:GO:0005773, C:GO:0009536, P:GO:0006950, P:GO:0000003, C:GO:0005777, P:GO:0005975, F:GO:0005515, F:GO:0003824, C:GO:0005739, C:GO:0016020, P:GO:0009719, P:GO:0016043, P:GO:0030154, P:GO:0016049	P:catabolic process, P:post-embryonic development, P:anatomical structure morphogenesis, C:vacuole, C:plastid, P:response to stress, P:reproduction, C:peroxisome, P:carbohydrate metabolic process, F:protein binding, F:catalytic activity, C:mitochondrion, C:membrane, P:response to endogenous stimulus, P:cellular component organization, P:cell differentiation, P:cell growth	AT4G35000.1	ascorbate peroxidase 3
VIT_05s0020g03710	Acyl- N-acyltransferases (NAT) superfamily	P:GO:0019538, P:GO:0009058, F:GO:0016740, P:GO:0009987	P:protein metabolic process, P:biosynthetic process, F:transferase activity, P:cellular process	AT2G32030.1	Acyl-CoA N-acyltransferases (NAT) superfamily protein
VIT_12s0028g03210	aminotransferase-related family	P:GO:0019748, P:GO:0007154, P:GO:0009991, P:GO:0009056, P:GO:0009058, F:GO:0005488, P:GO:0006950, C:GO:0016020, P:GO:0009719, P:GO:0016043, C:GO:0005829, F:GO:0016740, P:GO:0016049	P:secondary metabolic process, P:cell communication, P:response to extracellular stimulus, P:catabolic process, P:biosynthetic process, F:binding, P:response to stress, C:membrane, P:response to endogenous stimulus, P:cellular component organization, C:cytosol, F:transferase activity, P:cell growth	AT5G36160.1	Tyrosine transaminase family protein
VIT_16s0022g01860	Cleavage and polyadenylation specificity factor subunit 1	No GO term	No GO term	AT5G51660.1	cleavage and polyadenylation specificity factor 160

VIT_10s0071g00210	myosin 1	F:GO:0000166, C:GO:0005768, C:GO:0005856, F:GO:0005515, C:GO:0005783, P:GO:0006810, C:GO:0005886, P:GO:0009987, F:GO:0003774	F:nucleotide binding, C:endosome, C:cytoskeleton, F:protein binding, C:endoplasmic reticulum, P:transport, C:plasma membrane, P:cellular process, F:motor activity	AT1G50360.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein
VIT_04s0008g05220	cellulose synthase A catalytic subunit 1	C:GO:0005794, P:GO:0009058, P:GO:0006139, C:GO:0005886, C:GO:0009536, P:GO:0007049, P:GO:0006950, F:GO:0000166, P:GO:0005975, P:GO:0009628, C:GO:0005768, P:GO:0016043, F:GO:0016301	C:Golgi apparatus, P:biosynthetic process, P:nucleobase-containing compound metabolic process, C:plasma membrane, C:plastid, P:cell cycle, P:response to stress, F:nucleotide binding, P:carbohydrate metabolic process, P:response to abiotic stimulus, C:endosome, P:cellular component organization, F:kinase activity	AT4G32410.1	cellulose synthase 1
VIT_11s0065g00360	biotin carboxylase chloroplastic	F:GO:0000166, P:GO:0009056, F:GO:0003824, C:GO:0005739, P:GO:0006629, C:GO:0005829, P:GO:0009058, C:GO:0005840, C:GO:0009536, P:GO:0009987	F:nucleotide binding, P:catabolic process, F:catalytic activity, C:mitochondrion, P:lipid metabolic process, C:cytosol, P:biosynthetic process, C:ribosome, C:plastid, P:cellular process	AT5G35360.1	acetyl Co-enzyme a carboxylase biotin carboxylase subunit
VIT_13s0156g00640	serine threonine-kinase	C:GO:0005634, F:GO:0030528, P:GO:0009058, P:GO:0006139	C:nucleus, F:obsolete transcription regulator activity, P:biosynthetic process, P:nucleobase-containing compound metabolic process	-	-
VIT_18s0001g03300	aspartate aminotransferase family	P:GO:0009058, F:GO:0016740, F:GO:0005488, C:GO:0009536, P:GO:0009987	P:biosynthetic process, F:transferase activity, F:binding, C:plastid, P:cellular process	AT1G77670.1	Pyridoxal phosphate (PLP)-dependent transferases superfamily protein
VIT_11s0052g00980	hypothetical protein L484_010385	No GO term	No GO term	AT4G30260.2	Integral membrane Yip1 family protein
VIT_16s0022g00160	Uncharacterized protein	No GO term	No GO term	-	-
VIT_07s0130g00470	FYVE zinc finger	F:GO:0005515, F:GO:0016301	F:protein binding, F:kinase activity	AT2G26210.1	Ankyrin repeat family protein
VIT_13s0074g00620	acyl- -binding domain-containing 5 isoform X1	No GO term	No GO term	AT2G36360.1	Galactose oxidase/kelch repeat superfamily protein
VIT_08s0007g06000	Uncharacterized protein	No GO term	No GO term	-	-
VIT_15s0046g01830	pentatricopeptide repeat-containing mitochondrial-like	F:GO:0000166, F:GO:0003723, F:GO:0005515, P:GO:0009056, F:GO:0004518, P:GO:0006464, P:GO:0009058, F:GO:0016301, C:GO:0005622	F:nucleotide binding, F:RNA binding, F:protein binding, P:catabolic process, F:nuclease activity, P:cellular protein modification process, P:biosynthetic process, F:kinase activity, C:intracellular	AT4G00330.1	calmodulin-binding receptor-like cytoplasmic kinase 2
VIT_04s0008g05690	TRAF-like family	P:GO:0009607, F:GO:0005515, P:GO:0008219, P:GO:0009605, C:GO:0005737, F:GO:0008289, P:GO:0009058, P:GO:0006259, F:GO:0005215, P:GO:0006950, F:GO:0016787	P:response to biotic stimulus, F:protein binding, P:cell death, P:response to external stimulus, C:cytoplasm, F:lipid binding, P:biosynthetic process, P:DNA metabolic process, F:transporter activity, P:response to stress, F:hydrolase activity	AT2G25320.2	TRAF-like family protein
VIT_05s0094g00470	RING FYVE PHD zinc finger	F:GO:0003824, P:GO:0008152, C:GO:0016020, F:GO:0005488	F:catalytic activity, P:metabolic process, C:membrane, F:binding	AT1G14260.1	RING/FYVE/PHD zinc finger superfamily protein
VIT_19s0015g01750	uncharacterized WD repeat-containing -	F:GO:0003676, F:GO:0005515, C:GO:0005622	F:nucleic acid binding, F:protein binding, C:intracellular	AT3G13340.1	Transducin/WD40 repeat-like superfamily protein

	like				
VIT_12s0028g02000	phospholipase A1-IIdelta	P:GO:0009628, P:GO:0009056, P:GO:0006629, F:GO:0004872, F:GO:0004871, C:GO:0005886, C:GO:0009536, F:GO:0016787	P:response to abiotic stimulus, P:catabolic process, P:lipid metabolic process, F:receptor activity, F:signal transducer activity, C:plasma membrane, C:plastid, F:hydrolase activity	AT2G42690.1	alpha/beta-Hydrolases superfamily protein

Additional file 5, Table S2. Genes that their methylation and expression significantly (FDR-adjusted p-values < 0.05) correlate with mean annual rainfall (mm) obtained by fitting a generalized linear model. Grapevine gene function, gene ontology motifs and best *Arabidopsis* match was obtained from Cramer et al. (2020)

Gene ID	Description	GOs	GOs description	Best Arabidopsis match	Arabidopsis gene description
VIT_08s0040g01480	cell division cycle 48 homolog	P:GO:0019538, P:GO:0009056, C:GO:0005794, P:GO:0009653, P:GO:0009058, C:GO:0005730, C:GO:0005576, P:GO:0009856, C:GO:0005654, F:GO:0000166, C:GO:0005829, C:GO:0005840, P:GO:0006810, C:GO:0005886, P:GO:0007049, P:GO:0006950, C:GO:0005635, C:GO:0005856, F:GO:0005515, C:GO:0005618, P:GO:0006629, P:GO:0016043, P:GO:0030154, F:GO:0016740, P:GO:0016049, F:GO:0016787	P:protein metabolic process, P:catabolic process, C:Golgi apparatus, P:anatomical structure morphogenesis, P:biosynthetic process, C:nucleolus, C:extracellular region, P:pollination, C:nucleoplasm, F:nucleotide binding, C:cytosol, C:ribosome, P:transport, C:plasma membrane, P:cell cycle, P:response to stress, C:nuclear envelope, C:cytoskeleton, F:protein binding, C:cell wall, P:lipid metabolic process, P:cellular component organization, P:cell differentiation, F:transferase activity, P:cell growth, F:hydrolase activity	AT3G53230.1	ATPase%2C AAA-type%2C CDC48 protein
VIT_00s0252g00060	DNAJ heat shock N-terminal domain-containing family	C:GO:0005634, P:GO:0009628, F:GO:0005515, P:GO:0009908, F:GO:0008135, C:GO:0005840, P:GO:0006139, C:GO:0005886, P:GO:0006950	C:nucleus, P:response to abiotic stimulus, F:protein binding, P:flower development, F:translation factor activity, RNA binding, C:ribosome, P:nucleobase-containing compound metabolic process, C:plasma membrane, P:response to stress	AT5G12430.1	Heat shock protein DnaJ with tetratricopeptide repeat-containing protein
VIT_18s0001g10820	26S proteasome non-ATPase regulatory subunit 14 homolog	P:GO:0019538, P:GO:0019725, C:GO:0005634, P:GO:0009628, F:GO:0005515, P:GO:0009056, C:GO:0016020, C:GO:0005829, C:GO:0005783, C:GO:0005576, P:GO:0006950, F:GO:0016787	P:protein metabolic process, P:cellular homeostasis, C:nucleus, P:response to abiotic stimulus, F:protein binding, P:catabolic process, C:membrane, C:cytosol, C:endoplasmic reticulum, C:extracellular region, P:response to stress, F:hydrolase activity	AT5G23540.1	Mov34/MPN/PAD-1 family protein
VIT_09s0002g00770	Alpha-ketoglutarate-	F:GO:0003824, P:GO:0006259,	F:catalytic activity, P:DNA metabolic	AT3G14160.1	2-oxoglutarate-dependent dioxygenase family

	dependent dioxygenase	P:GO:0006950	process, P:response to stress		protein
VIT_06s0004g08230	root phototropism 2-like	P:GO:0009790, P:GO:0006464, P:GO:0007267, P:GO:0009653, P:GO:0009058, F:GO:0003682, P:GO:0006139, F:GO:0004871, F:GO:0005102, P:GO:0006810, P:GO:0006950, C:GO:0005654, F:GO:0003677, F:GO:0003700, P:GO:0009606, P:GO:0009628, P:GO:0008219, C:GO:0005737, P:GO:0009719, P:GO:0030154, P:GO:0016043	P:embryo development, P:cellular protein modification process, P:cell-cell signaling, P:anatomical structure morphogenesis, P:biosynthetic process, F:chromatin binding, P:nucleobase-containing compound metabolic process, F:signal transducer activity, F:receptor binding, P:transport, P:response to stress, C:nucleoplasm, F:DNA binding, F:transcription factor activity, sequence-specific DNA binding, P:tropism, P:response to abiotic stimulus, P:cell death, C:cytoplasm, P:response to endogenous stimulus, P:cell differentiation, P:cellular component organization	AT2G30520.1	Phototropic-responsive NPH3 family protein
VIT_18s0001g15520	trichome berefringence-like 7	C:GO:0016020	C:membrane	AT1G48880.1	TRICHOME BIREFRINGENCE-LIKE 7
VIT_13s0084g00360	NBS-LRR resistance gene ARGH30	F:GO:0000166, F:GO:0008135, C:GO:0005840, C:GO:0005886	F:nucleotide binding, F:translation factor activity, RNA binding, C:ribosome, C:plasma membrane	AT3G14470.1	NB-ARC domain-containing disease resistance protein
VIT_19s0014g03700	cyclic nucleotide-gated cation channel 4	C:GO:0005634, F:GO:0003677, C:GO:0016020, P:GO:0009719, P:GO:0009058, P:GO:0006139, F:GO:0005215	C:nucleus, F:DNA binding, C:membrane, P:response to endogenous stimulus, P:biosynthetic process, P:nucleobase-containing compound metabolic process, F:transporter activity	AT5G54250.1	cyclic nucleotide-gated cation channel 4
VIT_03s0038g04460	myb-like DNA-binding domain	P:GO:0007275, P:GO:0007165, C:GO:0005794, P:GO:0009058, P:GO:0006139, F:GO:0005215, C:GO:0005886, P:GO:0000003, C:GO:0005764, F:GO:0000166, C:GO:0005634, F:GO:0003677, F:GO:0003723, C:GO:0005856, P:GO:0016043, P:GO:0030154, C:GO:0005829, F:GO:0030234, F:GO:0016787	P:multicellular organism development, P:signal transduction, C:Golgi apparatus, P:biosynthetic process, P:nucleobase-containing compound metabolic process, F:transporter activity, C:plasma membrane, P:reproduction, C:lysosome, F:nucleotide binding, C:nucleus, F:DNA binding, F:RNA binding, C:cytoskeleton, P:cellular component organization, P:cell differentiation, C:cytosol, F:enzyme regulator activity, F:hydrolase activity	AT4G39160.1	Homeodomain-like superfamily protein

VIT_16s0098g01720	ABL interactor 2	P:GO:0007275, P:GO:0009653, P:GO:0009058, F:GO:0003682, C:GO:0005576, P:GO:0006810, C:GO:0005886, P:GO:0006950, C:GO:0005654, F:GO:0000166, P:GO:0005975, F:GO:0003677, P:GO:0008219, C:GO:0005856, F:GO:0005515, C:GO:0005739, P:GO:0016043, P:GO:0030154, C:GO:0005829, F:GO:0016301, F:GO:0016787	P:multicellular organism development, P:anatomical structure morphogenesis, P:biosynthetic process, F:chromatin binding, C:extracellular region, P:transport, C:plasma membrane, P:response to stress, C:nucleoplasm, F:nucleotide binding, P:carbohydrate metabolic process, F:DNA binding, P:cell death, C:cytoskeleton, F:protein binding, C:mitochondrion, P:cellular component organization, P:cell differentiation, C:cytosol, F:kinase activity, F:hydrolase activity	AT5G24310.1	ABL interactor-like protein 3
VIT_06s0004g06780	PHD finger-like domain-containing 5B	C:GO:0005634, F:GO:0003824, P:GO:0006139	C:nucleus, F:catalytic activity, P:nucleobase-containing compound metabolic process	AT1G07170.1	PHF5-like protein
VIT_13s0067g01690	probable methyltransferase PMT23	C:GO:0005768, P:GO:0007165, C:GO:0016020, P:GO:0008152, C:GO:0005794, C:GO:0005783, F:GO:0004872, F:GO:0016740	C:endosome, P:signal transduction, C:membrane, P:metabolic process, C:Golgi apparatus, C:endoplasmic reticulum, F:receptor activity, F:transferase activity	AT2G40280.1	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
VIT_18s0001g11760	MATE efflux family	P:GO:0009607, P:GO:0009605, C:GO:0016020, C:GO:0005773, F:GO:0005215	P:response to biotic stimulus, P:response to external stimulus, C:membrane, C:vacuole, F:transporter activity	AT1G47530.1	MATE efflux family protein
VIT_18s0001g08370	endo-1,4-beta-xylanase A-like	P:GO:0005975, P:GO:0009056, C:GO:0005618, P:GO:0009987, F:GO:0016787	P:carbohydrate metabolic process, P:catabolic process, C:cell wall, P:cellular process, F:hydrolase activity	AT1G58370.2	glycosyl hydrolase family 10 protein / carbohydrate-binding domain-containing protein
VIT_00s0179g00100	ETO1 1	F:GO:0005515, P:GO:0007165, P:GO:0009991, P:GO:0009719, P:GO:0009791, P:GO:0006464, P:GO:0009058, C:GO:0005622, P:GO:0006950	F:protein binding, P:signal transduction, P:response to extracellular stimulus, P:response to endogenous stimulus, P:post-embryonic development, P:cellular protein modification process, P:biosynthetic process, C:intracellular, P:response to stress	AT4G02680.1	ETO1-like 1
VIT_09s0002g00690	chaperone domain	F:GO:0003723, F:GO:0004518	F:RNA binding, F:nuclease activity	AT3G14200.1	Chaperone DnaJ-domain superfamily protein

Additional file 5, Table S3. Genes that their methylation and expression significantly (FDR-adjusted p-values < 0.05) correlate with soil pH obtained by fitting a generalized linear model. Grapevine gene function, gene ontology motifs and best *Arabidopsis* match was obtained from Cramer et al. (2020).

Gene ID	Description	GOs	GOs description	Best Arabidopsis match	Arabidopsis gene description
VIT_12s0028g02130	MAP kinase kinase kinase	F:GO:0000166, F:GO:0003723, P:GO:0006464, P:GO:0009058, P:GO:0006259, F:GO:0016301, F:GO:0004871, C:GO:0005622	F:nucleotide binding, F:RNA binding, P:cellular protein modification process, P:biosynthetic process, P:DNA metabolic process, F:kinase activity, F:signal transducer activity, C:intracellular	AT3G58640.1	Mitogen activated protein kinase kinase kinase-like protein
VIT_13s0101g00510	Ribonuclease H2 subunit A	F:GO:0000166, F:GO:0003723, F:GO:0005515, P:GO:0008219, F:GO:0004518, P:GO:0006950	F:nucleotide binding, F:RNA binding, F:protein binding, P:cell death, F:nuclease activity, P:response to stress	AT2G25100.1	Polynucleotidyl transferase%2C ribonuclease H-like superfamily protein
VIT_14s0006g00910	senescence-associated family	F:GO:0005488	F:binding	AT1G66330.1	senescence-associated family protein
VIT_09s0054g00520	orf257 (mitochondrion)	C:GO:0016020	C:membrane	ATCG01000.1	Ycf1 protein
VIT_19s0014g00070	Solanesyl diphosphate synthase 1	P:GO:0006629, P:GO:0015979, P:GO:0009058, F:GO:0016740	P:lipid metabolic process, P:photosynthesis, P:biosynthetic process, F:transferase activity	AT1G17050.1	solanesyl diphosphate synthase 2

Additional file 5, Table S4. Genes that their methylation and expression significantly (FDR-adjusted p-values < 0.05) correlate with mean seasonal rainfall (mm) obtained by fitting a generalized linear model. Grapevine gene function, gene ontology motifs and best *Arabidopsis* match was obtained from Cramer et al. (2020).

Gene ID	Description	GOs	GOs description	Best Arabidopsis match	Arabidopsis gene description
VIT_06s0004g02820	vacuolar fusion MON1 homolog isoform X2	C:GO:0005768, P:GO:0007275, P:GO:0016043, F:GO:0004872, F:GO:0005488, P:GO:0006810, F:GO:0004871, C:GO:0005886	C:endosome, P:multicellular organism development, P:cellular component organization, F:receptor activity, F:binding, P:transport, F:signal transducer activity, C:plasma membrane	AT2G28390.1	SAND family protein
VIT_11s0016g04530	tudor domain-containing 3	F:GO:0000166, P:GO:0009058, F:GO:0016301, C:GO:0009536	F:nucleotide binding, P:biosynthetic process, F:kinase activity, C:plastid	AT5G19950.1	tudor domain protein (DUF1767)

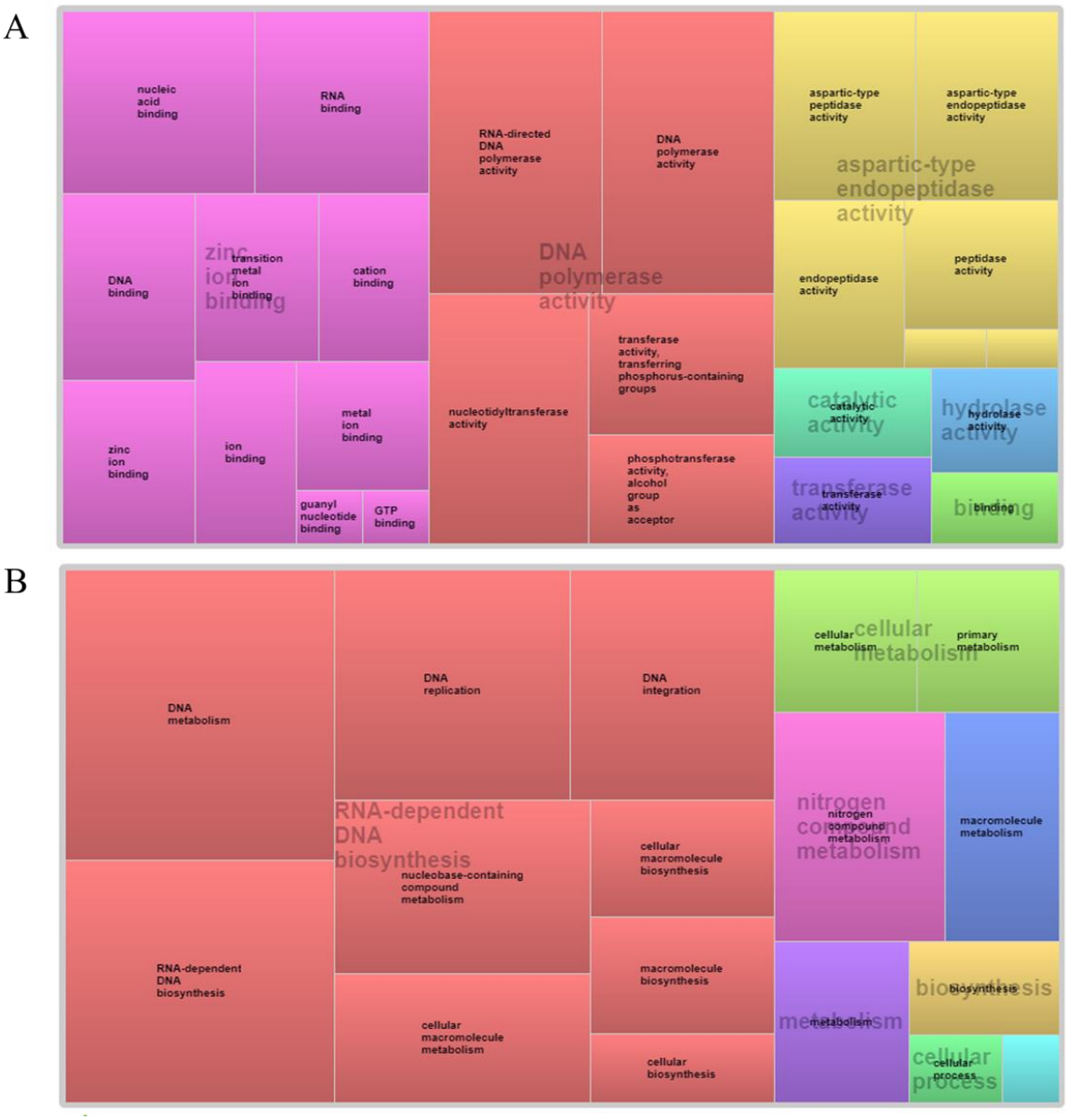
Additional file 5, Table S5. Differentially expressed genes that significantly (FDR-adjusted p-value < 0.05) correlated with vineyard altitude that were also reported to be affected by UV light in grapevine at high altitudes (Carbonell-Bejerano et al. 2014).

Gene ID	Functional annotation	Best Arabidopsis match	Functional category
VIT_15s0021g02060	Arbutin synthase	AT4G01070	Metabolism.Primary metabolism.Carbohydrate metabolism.Polysaccharide metabolism.Starch and sucrose metabolism
VIT_19s0090g00080	Ubiquitin-conjugating enzyme E2 variant	AT4G27030	Metabolism.Primary metabolism.Protein metabolism and modification.Proteolysis.cytoplasmic and nuclear protein degradation.Ubiquitin-mediated proteolysis
VIT_02s0109g00350	Carboxyl-terminal proteinase	AT2G44220	Metabolism.Primary metabolism.Protein metabolism and modification.Proteolysis.Peptidase-mediated proteolysis
VIT_11s0052g01630	VvGT6 is a bifunctional UDP-glucose/UDP-galactose:flavonol-3-O-glucosyltransferase/galactosyltransferase		Metabolism.Secondary metabolism.Phenylpropanoid metabolism.Flavonoid metabolism.Flavonoid biosynthesis.Anthocyanin biosynthesis.Anthocyanin-glycoside biosynthesis
VIT_00s0385g00020	Linalool synthase	AT1G61680	Metabolism.Secondary metabolism.Terpenoid metabolism.Terpenoid biosynthesis
VIT_04s0023g03520	Cold-shock DNA-binding	AT2G17870	Regulation overview.Regulation of gene expression.Regulation of transcription.Transcription factor.CSD family transcription factor
VIT_03s0038g02130	Cold shock protein-1	AT4G38680	Regulation overview. Regulation of gene expression.Regulation of transcription.Transcription factor.CSD family transcription factor
VIT_13s0019g02940	Cis-zeatin O-beta-D-glucosyltransferase	AT1G22400	Signaling. Hormone Signaling. Cytokinin Signaling. Cytokinin metabolism. Cytokinin inactivation by glycosylation
VIT_04s0008g01290	CGA1 (cytokinin-responsive GATA factor 1)	AT4G26150	Signaling. Hormone Signaling. Cytokinin Signaling. Cytokinin-mediated Signaling pathway
VIT_19s0014g01370	Curculin (mannose-binding) lectin	AT1G78850	Unknown

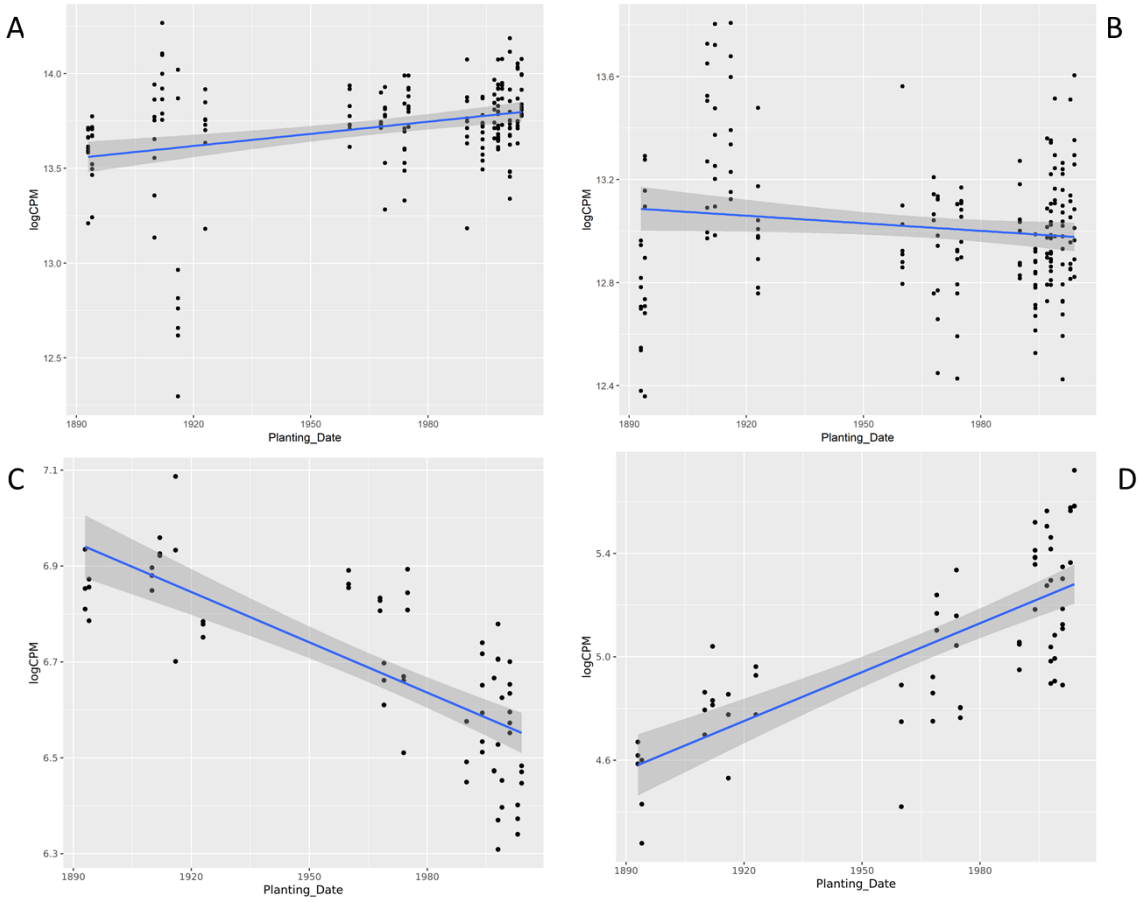
Additional_file_6, Table S1. Significantly enriched gene ontology (GO) terms for genes that their methylation and expression significantly correlate with planting year. Enrichment analysis was performed in the web-based tool AgriGO that specially focus on agricultural species. MF: molecular function, BP: biological process.

GO term	Ontology	Description	Number in input list	p-value	FDR
GO:0006259	BP	DNA metabolic process	43	9.20E-21	2.90E-18
GO:0006278	BP	RNA-dependent DNA replication	27	1.80E-17	2.80E-15
GO:0006260	BP	DNA replication	27	6.80E-15	7.20E-13
GO:0015074	BP	DNA integration	24	3.90E-13	3.10E-11
GO:0006139	BP	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	53	2.00E-12	1.30E-10
GO:0006807	BP	nitrogen compound metabolic process	55	3.50E-11	1.90E-09
GO:0044260	BP	cellular macromolecule metabolic process	62	9.70E-10	4.40E-08
GO:0043170	BP	macromolecule metabolic process	63	3.70E-08	1.50E-06
GO:0008152	BP	metabolic process	76	4.10E-07	1.40E-05
GO:0034645	BP	cellular macromolecule biosynthetic process	38	5.30E-07	1.70E-05
GO:0009059	BP	macromolecule biosynthetic process	38	6.20E-07	1.80E-05
GO:0044237	BP	cellular metabolic process	68	1.00E-06	2.70E-05
GO:0044238	BP	primary metabolic process	69	1.20E-06	3.00E-05
GO:0009058	BP	biosynthetic process	44	3.00E-05	0.00068
GO:0044249	BP	cellular biosynthetic process	42	7.20E-05	0.0015
GO:0009987	BP	cellular process	72	0.0021	0.04
GO:0003964	MF	RNA-directed DNA polymerase activity	27	1.40E-17	3.20E-15
GO:0034061	MF	DNA polymerase activity	27	3.00E-17	3.50E-15
GO:0016779	MF	nucleotidyltransferase activity	27	1.60E-14	1.30E-12
GO:0003676	MF	nucleic acid binding	54	6.10E-13	3.60E-11
GO:0003723	MF	RNA binding	29	8.10E-12	3.80E-10
GO:0070001	MF	aspartic-type peptidase activity	20	3.10E-10	1.00E-08
GO:0004190	MF	aspartic-type endopeptidase activity	20	3.10E-10	1.00E-08
GO:0016772	MF	transferase activity, transferring phosphorus-containing groups	38	5.70E-10	1.70E-08
GO:0003677	MF	DNA binding	37	1.60E-09	3.90E-08
GO:0016301	MF	kinase activity	31	1.60E-09	3.90E-08
GO:0004175	MF	endopeptidase activity	20	1.50E-08	3.20E-07
GO:0008270	MF	zinc ion binding	31	1.80E-08	3.50E-07
GO:0046914	MF	transition metal ion binding	38	4.00E-08	7.20E-07
GO:0016773	MF	phosphotransferase activity, alcohol group as acceptor	27	6.80E-08	1.10E-06
GO:0008233	MF	peptidase activity	23	8.10E-08	1.30E-06
GO:0043169	MF	cation binding	44	2.60E-07	3.40E-06
GO:0043167	MF	ion binding	44	2.70E-07	3.40E-06
GO:0003899	MF	DNA-directed RNA polymerase activity	7	2.80E-07	3.40E-06
GO:0070011	MF	peptidase activity, acting on L-amino acid peptides	20	2.60E-07	3.40E-06
GO:0034062	MF	RNA polymerase activity	7	3.60E-07	4.30E-06

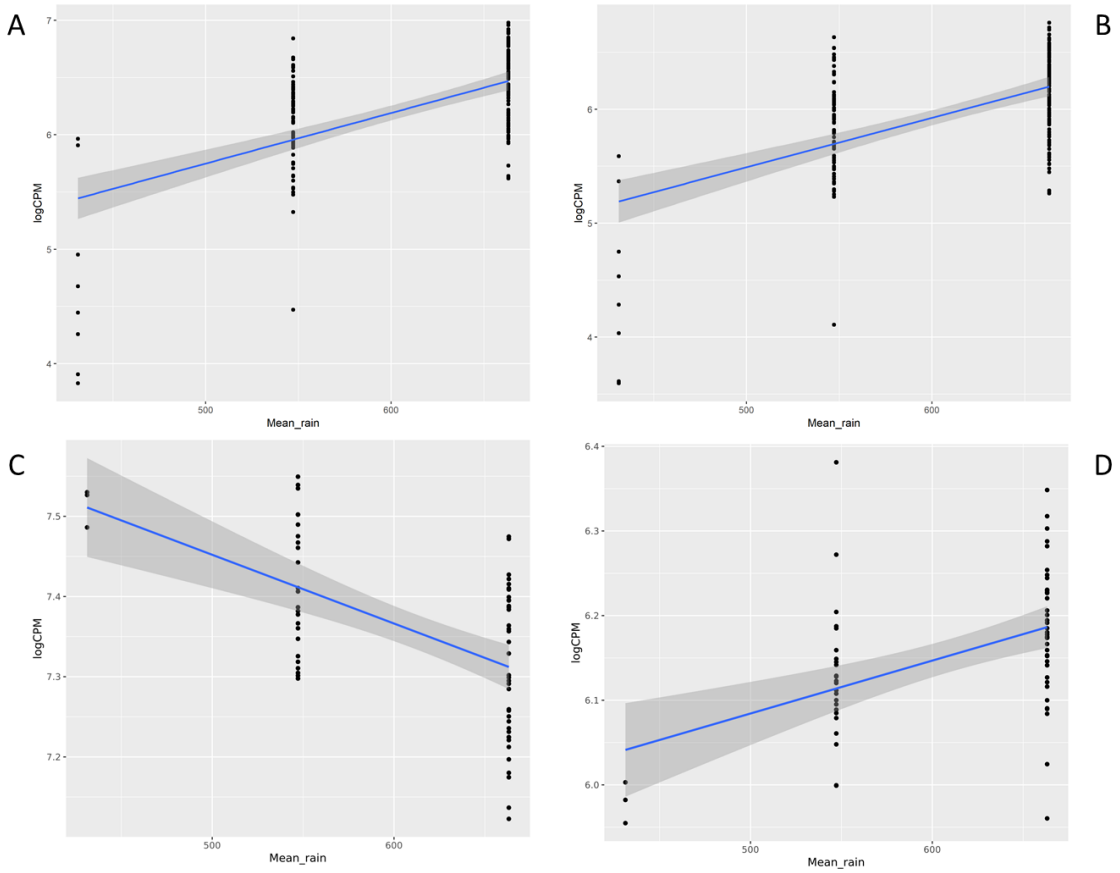
GO term	Ontology	Description	Number in input list	p-value	FDR
GO:0046872	MF	metal ion binding	42	7.60E-07	8.50E-06
GO:0003824	MF	catalytic activity	72	6.50E-06	6.90E-05
GO:0016740	MF	transferase activity	41	9.40E-06	9.60E-05
GO:0016787	MF	hydrolase activity	40	1.30E-05	0.00012
GO:0004672	MF	protein kinase activity	21	3.20E-05	0.0003
GO:0005488	MF	binding	78	0.00024	0.0021



Additional file 7: TreeMap from Revigo of enriched GO terms for planting year. Each rectangle is a single cluster representative. The representatives are joined into 'superclusters' of loosely elated terms, visualized with different colors. The size of the rectangles reflects the p-value of GO term. A: Molecular Function, B: Biological Process.

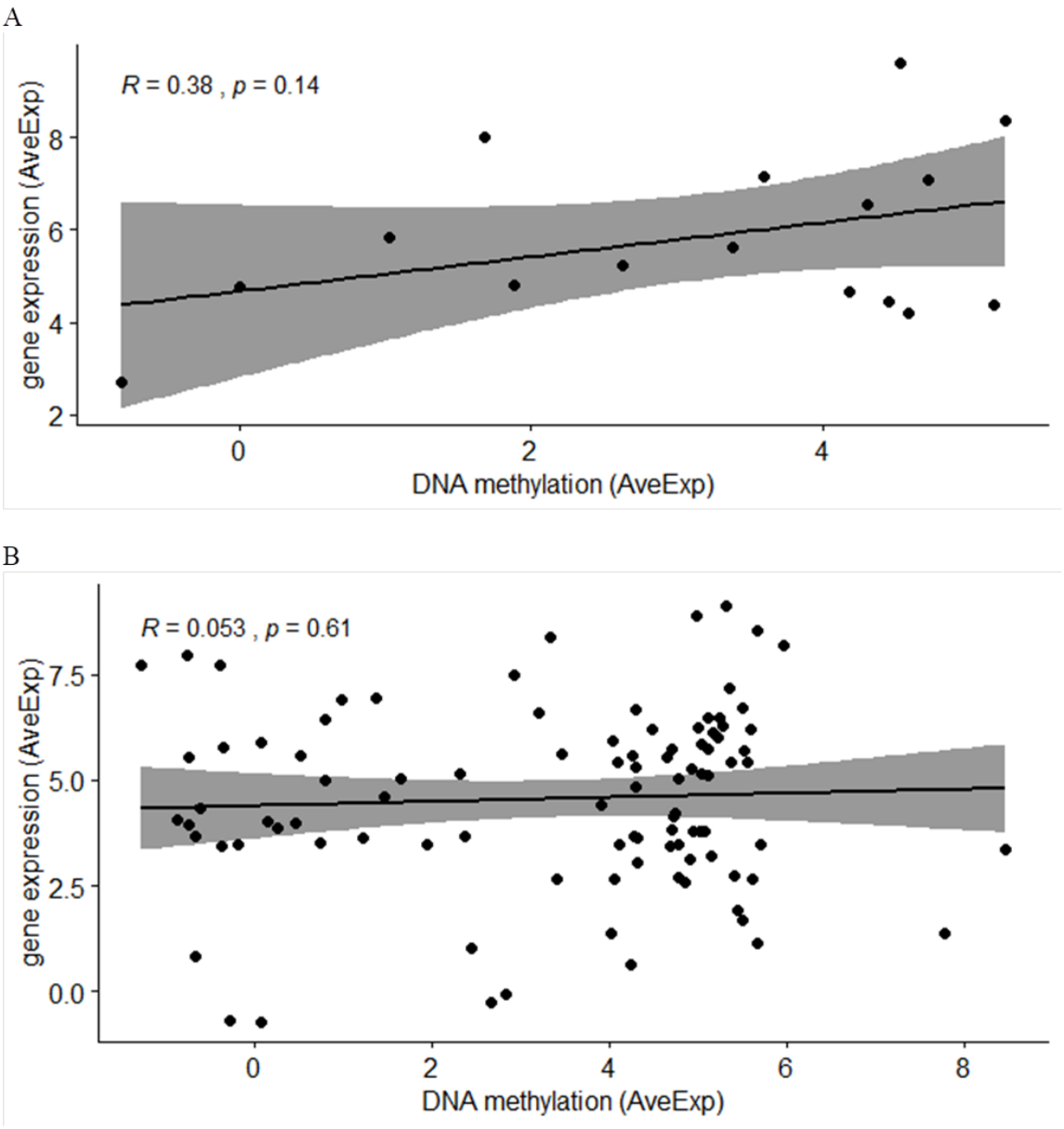


Additional file 8, Figure S1: DNA methylation loci (top panel) and genes significantly associated (FDR-adjusted p-values < 0.05) with planting date from samples of the twenty-two vineyards across the Barossa Wine Zone. The scatter plots show the association between the read counts (logCPM) for the methylation data (top panel) and gene expression data (bottom panel) with planting date of the vineyards. Two loci significantly associated with planting year date were selected for plotting, A) chr15:1:18157053 (FDR-adj. p-value = 2.49e-10) and B) chr19:-1:51468 (FDR-adj. p-value = 2.78e-13). Additionally, two differentially expressed genes significantly associated with planting date were selected for plotting, C) VIT_00s0583g00030 (FDR-adj. p-value = 9.76e-10) 9and D) VIT_02s0025g02310 (FDR-adj. p-value = 3.91e-9).



Additional file 8, Figure S2: DNA methylation (top panel) and genes significantly associated (FDR-adjusted p-values < 0.05) with mean annual rainfall from samples of the twenty-two vineyards across the Barossa Wine Zone. The scatter plots show the association between the read counts (logCPM) for the methylation data (top panel) and gene expression data (bottom panel) with mean annual rainfall of the vineyards. Two loci significantly associated with mean annual rainfall were selected for plotting, A) chr02:-1:3790861 (FDR-adj. p-value = 5.06e-10) and B) chr05:1:444456 (FDR-adj. p-value =1.57e-09). Additionally, two differentially expressed genes significantly associated with mean annual rainfall were also selected for plotting, C) VIT_09s0002g03380 FDR-adj. p-value = 0.0131) and D) VIT_10s0003g00190 FDR-adj. p-value = 0.0156).

.



Additional file 9: Correlation analysis (Pearson) between average read counts (logCPM) of DNA methylation and expression, for genes whose expression and methylation are significantly associated with A) mean annual rainfall (mm) and B) planting year.

References

Adam-Blondon A-F, Martinez-Zapater J-M, Kole C, eds. 2016. *Genetics, Genomics, and Breeding of Grapes*. 1st ed. Science Publishers, Enfield, USA.

Agrawal AA, Tollrian R, Laforsch C. 1999. Transgenerational induction of defences in animals and plants. *Nature* **401**: 60–63.

Anesi A, Stocchero M, Dal Santo S, Commisso M, Zenoni S, Ceoldo S, Tornielli GB, Siebert TE, Herderich M, Pezzotti M, et al. 2015. Towards a scientific interpretation of the *terroir* concept: plasticity of the grape berry metabolome. *BMC Plant Biol* **15**: 191.

Aquea F, Vega A, Timmermann T, Poupin MJ, Arce-Johnson P. 2011. Genome-wide analysis of the SET DOMAIN GROUP family in grapevine. *Plant Cell Rep* **30**: 1087–1097.

Aradhya MK, Dangl GS, Prins BH, Boursiquot J-M, Walker MA, Meredith CP, Simon CJ. 2003. Genetic structure and differentiation in cultivated grape, *Vitis vinifera* L. *Genet Res* **81**: 179–192.

Barossa Grape and Wine Association. Barossa Chapters: Barossa Grounds. https://www.barossawine.com/barossa/barossa-chapters/barossa-chapters-barossa-grounds/ (Accessed November 14, 2019).

Bouzas-Cid Y, Falqué E, Orriols I, Mirás-Avalos JM. 2018. Effects of irrigation over three years on the amino acid composition of Treixadura (*Vitis vinifera* L.) musts and wines, and on the aromatic composition and sensory profiles of its wines. *Food Chem* **240**: 707–716.

Boyko A, Kovalchuk I. 2010. Transgenerational response to stress in *Arabidopsis thaliana*. *Plant Signal Behav* **5**: 995–998.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**: 2633–2635.

Bushnell B, Rood J, Singer E. 2017. BBMerge - accurate paired shotgun read merging via overlap. *PLoS One* **12**: e0185056.

Carbonell-Bejerano P, Diago M-P, Martínez-Abaigar J, Martínez-Zapater JM, Tardáguila J, Núñez-Olivera E. 2014. Solar ultraviolet radiation is necessary to enhance grapevine fruit ripening transcriptional and phenolic responses. *BMC Plant Biol* **14**: 183.

Chinnusamy V, Gong Z, Zhu J-K. 2008. Abscisic acid-mediated epigenetic processes in plant development and stress responses. *J Integr Plant Biol* **50**: 1187–1195.

Conde C, Silva P, Fontes N, Dias ACP, Tavares RM, Sousa MJ, Agasse A, Delrot S, Geros H, Conde C, et al. 2007. Biochemical changes throughout grape berry development and fruit and wine quality. *Food* **1**: 1–22.

Coombe BG. 1995. Growth Stages of the Grapevine: Adoption of a system for identifying grapevine growth stages. *Aust J Grape Wine Res* **1**: 104–110.

Cramer GR, Cochetel N, Ghan R, Destrac-Irvine A, Delrot S. 2020. A sense of place: transcriptomics identifies environmental signatures in Cabernet Sauvignon berry skins in the late stages of ripening. *BMC Plant Biol* **20**: 41

Daccord N, Celton J-M, Linsmith G, Becker C, Choisne N, Schijlen E, van de Geest H, Bianco L, Micheletti D, Velasco R, et al. 2017. High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nat Genet* **49**: 1099–1106.

Dal Santo S, Tornielli GB, Zenoni S, Fasoli M, Farina L, Anesi A, Guzzo F, Delledonne M, Pezzotti M. 2013. The plasticity of the grapevine berry transcriptome. *Genome Biol* **14**: r54.

Dal Santo S, Zenoni S, Sandri M, De Lorenzis G, Magris G, De Paoli E, Di Gaspero G, Del Fabbro C, Morgante M, Brancadoro L, et al. 2018. Grapevine field experiments reveal the contribution of genotype, the influence of environment and the effect of their interaction (G×E) on the berry transcriptome. *Plant J* **93**: 1143–1159.

Diaz M, Sanchez-Barrena MJ, Gonzalez-Rubio JM, Rodriguez L, Fernandez D, Antoni R, Yunta C, Belda-Palazon B, Gonzalez-Guzman M, Peirats-Llobet M, et al. 2016. Calcium-dependent oligomerization of CAR proteins at cell membrane modulates ABA signaling. *Proc Natl Acad Sci USA* **113**: E396–405.

Du Z, Zhou X, Ling Y, Zhang Z, Su Z. 2010. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res* **38**: W64–70.

Eichten SR, Schmitz RJ, Springer NM. 2014. Epigenetics: beyond chromatin modifications and complex genetic regulation. *Plant Physiol* **165**: 933–947.

Enot DP, Lin W, Beckmann M, Parker D, Overy DP, Draper J. 2008. Preprocessing, classification modeling and feature selection using flow injection electrospray mass spectrometry metabolite fingerprint data. *Nat Protoc* **3**: 446–470.

Farquhar W. 2005. Which Shiraz clones are best? *Aust NZ Grapegrow Winema*k **495**: 31–33.

Feil R, Fraga MF. 2012. Epigenetics and the environment: emerging patterns and implications. *Nat Rev Genet* **13**: 97–109.

Fonseca Lira-Medeiros C, Parisod C, Fernandes RA, Mata CS, Cardoso MA, Ferreira PC. 2010. Epigenetic variation in mangrove plants occurring in contrasting natural environment. *PLoS One* **5**: e10326.

Futschik A, Schlötterer C. 2010. The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics* **186**: 207–218.

Gallusci P, Hodgman C, Teyssier E, Seymour GB. 2016. DNA methylation and chromatin regulation during fleshy fruit development and ripening. *Front Plant Sci* **7**: 807.

Giovannoni J, Nguyen C, Ampofo B, Zhong S, Fei Z. 2017. The epigenome and transcriptional dynamics of fruit ripening. *Annu Rev Plant Biol* **68**: 61–84.

Gladstones J. 2011. Introduction and definition of terroir. In *Wine, Terroir & Climate Change*, pp. 1–4, Wakefield Press, Kent Town, Australia.

Grigg D, Methven D, de Bei R, Rodríguez López CM, Dry P, Collins C. 2017. Effect of vine age on vine performance of Shiraz in the Barossa Valley, Australia. *Aust J Grape Wine Res* **24**: 75–87.

Grigg DP. 2017. An investigation into the effect of grapevine age on vine performance, grape and wine composition, sensory evaluation and epigenetic characterisation. *University of Adelaide*.

Herten K, Hestand MS, Vermeesch JR, Van Houdt JKJ. 2015. GBSX: a toolkit for experimental design and demultiplexing genotyping by sequencing experiments. *BMC Bioinformatics* **16**: 73.

Imazio S, Labra M, Grassi F, Winfield M, Bardini M, Scienza A. 2002. Molecular tools for clone identification: the case of the grapevine cultivar Traminer. *Plant Breeding* **121**: 531–535.

Intrigliolo DS, Castel JR. 2011. Interactive effects of deficit irrigation and shoot and cluster thinning on grapevine cv. Tempranillo. Water relations, vine performance and berry and wine composition. *Irrig Sci* **29**: 443–454.

Intrigliolo DS, Castel JR. 2010. Response of grapevine cv. “Tempranillo” to timing and amount of irrigation: water relations, vine growth, yield and berry and wine composition. *Irrig Sci* **28**: 113–125.

Jackson DI, Lombard PB. 1993. Environmental and Management Practices Affecting Grape Composition and Wine Quality - A Review. *Am J Enol Vitic* **44**: 409–460.

Jaenisch R, Bird A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33 Suppl**: 245–254.

Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, et al. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**: 463–467.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* **11**: 94.

Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**: 357–360.

Kim SY, Li Y, Guo Y, Li R, Holmkvist J, Hansen T, Pedersen O, Wang J, Nielsen R. 2010. Design of association studies with pooled or un-pooled next-generation sequencing data. *Genet Epidemiol* **34**: 479–491.

Kitimu SR, Taylor J, March TJ, Tairo F, Wilkinson MJ, Rodríguez López CM. 2015. Meristem micropropagation of cassava (*Manihot esculenta*) evokes genome-wide changes in DNA methylation. *Front Plant Sci* **6**: 590.

Kumar G, Rattan UK, Singh AK. 2016. Chilling-mediated DNA methylation changes during dormancy and its release reveal the importance of epigenetic regulation during winter dormancy in apple (*Malus x domestica* Borkh.). *PLoS One* **11**: e0149934.

Lämke J, Bäurle I. 2017. Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. *Genome Biol* **18**: 124

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

Laucou V, Launay A, Bacilieri R, Lacombe T, Adam-Blondon A-F, Bérard A, Chauveau A, de Andrés MT, Hausmann L, Ibáñez J, et al. 2018. Extended diversity analysis of cultivated grapevine *Vitis vinifera* with 10K genome-wide SNPs. *PLoS One* **13**: e0192540.

Law CW, Chen Y, Shi W, Smyth GK. 2014. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**: R29.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Li X, Zhu J, Hu F, Ge S, Ye M, Xiang H, Zhang G, Zheng X, Zhang H, Zhang S, et al. 2012. Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. *BMC Genomics* **13**: 300.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930.

Magris G, Di Gaspero G, Marroni F, Zenoni S, Tornielli GB, Celii M, De Paoli E, Pezzotti M, Conte F, Paci P, et al. 2019. Genetic, epigenetic and genomic effects on variation of gene expression among grape varieties. *Plant J* **99**: 895–909.

Marfil C, Ibañez V, Alonso R, Varela A, Bottini R, Masuelli R, Fontana A, Berli F. 2019. Changes in grapevine DNA methylation and polyphenols content induced by solar ultraviolet-B radiation, water deficit and abscisic acid spray treatments. *Plant Physiol Biochem* **135**: 287–294.

Mayne BT, Leemaqz SY, Buckberry S, Rodriguez Lopez CM, Roberts CT, Bianco-Miotto T, Breen J. 2018. msgbsR: An R package for analysing methylation-sensitive restriction enzyme sequencing data. *Sci Rep* **8**: 2190.

McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**: 4288–4297.

Monteuuis O, Baurens FC, Goh DKS, Quimado M, Doulbeau S, Verdeil JL. 2009. DNA methylation in *Acacia mangium in vitro* and *ex-vitro* buds, in relation to their within-shoot position, age and leaf morphology of the shoots. *Silvae genetica* **58**.

Monteuuis O, Doulbeau S, Verdeil J-L. 2008. DNA methylation in different origin clonal offspring from a mature *Sequoiadendron giganteum* genotype. *Trees* **22**: 779–784.

Ocaña J, Walter B, Schellenbaum P. 2013. Stable MSAP markers for the distinction of *Vitis vinifera* cv Pinot noir clones. *Mol Biotechnol* **55**: 236–248.

Ogneva ZV, Dubrovina AS, Kiselev KV. 2016. Age-associated alterations in DNA methylation and expression of methyltransferase and demethylase genes in *Arabidopsis thaliana*. *Biol Plant* **60**: 628–634.

Ortega-Heras M, Pérez-Magariño S, Del-Villar-Garrachón V, González-Huerta C, Moro Gonzalez LC, Guadarrama Rodríguez A, Villanueva Sanchez S, Gallo González R, Martín de la Helguera S. 2014. Study of the effect of vintage, maturity degree, and irrigation on the amino acid and biogenic amine content of a white wine from the Verdejo variety. *J Sci Food Agric* **94**: 2073–2082.

Pasaniuc B, Rohland N, McLaren PJ, Garimella K, Zaitlen N, Li H, Gupta N, Neale BM, Daly MJ, Sklar P, et al. 2012. Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. *Nat Genet* **44**: 631–635.

Pikaard CS, Mittelsten Scheid O. 2014. Epigenetic regulation in plants. *Cold Spring Harb Perspect Biol* **6**: a019315.

Plongthongkum N, Diep DH, Zhang K. 2014. Advances in the profiling of DNA modifications: cytosine methylation and beyond. *Nat Rev Genet* **15**: 647 – 661.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

R Core Team. 2014. R: A language and environment for statistical' ' computing. *The R Project for Statistical Computing*. https://www.r-project.org/ (Accessed February 25, 2019).

Raj S, Bräutigam K, Hamanishi ET, Wilkins O, Thomas BR, Schroeder W, Mansfield SD, Plant AL, Campbell MM. 2011. Clone history shapes *Populus* drought responses. *Proc Natl Acad Sci USA* **108**: 12521–12526.

Reynolds AG, Pearson EG, De Savigny C, Coventry J, Strommer J. 2008. Interactions of vine age and reflective mulch upon berry, must, and wine composition of five *Vitis vinifera* cultivars. *Int J Fruit Sci* **7**: 85–119.

Rigal M, Becker C, Pélissier T, Pogorelcnik R, Devos J, Ikeda Y, Weigel D, Mathieu O. 2016. Epigenome confrontation triggers immediate reprogramming of DNA methylation and transposon silencing in *Arabidopsis thaliana* F1 epihybrids. *Proc Natl Acad Sci USA* **113**: E2083–92.

Robinson AL, Adams DO, Boss PK, Heymann H, Solomon PS, Trengove RD. 2012. Influence of geographic origin on the sensory characteristics and wine composition of *Vitis vinifera* cv. Cabernet Sauvignon wines from Australia. *Am J Enol Vitic* **63**: 467–476.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Robinson MD, Smyth GK. 2007. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* **23**: 2881–2887.

Robinson S, Sandercock N. 2014. *An analysis of climate, soil and topographic information to aid the understanding of barossa terroir*. SA: Primary Industries and Regions South Australia, Adelaide, Australia.

Rodrigues JA, Zilberman D. 2015. Evolution and function of genomic imprinting in plants. *Genes Dev* **29**: 2517–2531.

Roullier-Gall C, Lucio M, Noret L, Schmitt-Kopplin P, Gougeon RD. 2014. How subtle is the "terroir" effect? Chemistry-related signatures of two "climats de Bourgogne". *PLoS One* **9**: e97615.

Ruiz-García L, Cervera MT, Martínez-Zapater JM. 2005. DNA methylation increases throughout *Arabidopsis* development. *Planta* **222**: 301–306.

Schubert M, Lindgreen S, Orlando L. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes* **9**: 88.

Sheldon CC, Burn JE, Perez PP, Metzger J, Edwards JA, Peacock WJ, Dennis ES. 1999. The FLF MADS box gene: a repressor of flowering in *Arabidopsis* regulated by vernalization and methylation. *Plant Cell* **11**: 445–458.

Son H-S, Hwang G-S, Kim KM, Ahn H-J, Park W-M, Van Den Berg F, Hong Y-S, Lee C-H. 2009. Metabolomic studies on geographical grapes and their wines using 1H NMR analysis coupled with multivariate statistics. *J Agric Food Chem* **57**: 1481–1490.

Sultan SE. 2000. Phenotypic plasticity for plant development, function and life history. *Trends Plant Sci* **5**: 537–542.

Sun R, He F, Lan Y, Xing R, Liu R, Pan Q, Wang J, Duan C. 2015. Transcriptome comparison of Cabernet Sauvignon grape berries from two regions with distinct climate. *J Plant Physiol* **178**: 43–54.

Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**: e21800.

Tarr PT, Dreyer ML, Athanas M, Shahgholi M, Saarloos K, Second TP. 2013. A metabolomics based approach for understanding the influence of terroir in *Vitis Vinifera* L. *Metabolomics* **9**: 170–177.

This P, Lacombe T, Thomas MR. 2006. Historical origins and genetic diversity of wine grapes. *Trends Genet* **22**: 511–519.

Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z. 2017. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res* **45**: W122–W129.

Tricker PJ, López CMR, Gibbings G, Hadley P, Wilkinson MJ. 2013. Transgenerational, dynamic methylation of stomata genes in response to low relative humidity. *Int J Mol Sci* **14**: 6674–6689.

Trigo-Córdoba E, Bouzas-Cid Y, Orriols-Fernández I, Mirás-Avalos JM. 2015. Effects of deficit irrigation on the performance of grapevine (*Vitis vinifera* L.) cv. "Godello" and "Treixadura" in Ribeiro, NW Spain. *Agricultural Water Management* **161**: 20–30.

van Leeuwen C, Friant P, Choné X, Tregoat O, Koundouras S, Dubourdieu D, van Leeuwen C, Friant P, Choné X, Tregoat O, et al. 2004. Influence of climate, soil, and cultivar on *terroir*. *Am J Enol Vitic* **55**: 207-217.

Vaughn MW, Tanurdzić M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A, et al. 2007. Epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Biol* **5**: e174.

Verhoeven KJF, Jansen JJ, van Dijk PJ, Biere A. 2010. Stress-induced DNA methylation changes and their heritability in asexual dandelions. *New Phytol* **185**: 1108–1118.

Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Maechler M, Magnusson A, Moeller S, et al. 2009. Gplots: Various R Programming Tools for Plotting Data. *Gplots: Various R Programming Tools for Plotting Data*. https://cran.r-project.org/web/packages/gplots/index.htmlhttps://www.rdocumentation.org/packages/gplots/versions/3.0.1.1 (Accessed February 25, 2019).

Williams LE, Araujo FJ. 2002. Correlations among predawn leaf, midday leaf, and midday stem water potential and their correlations with other measures of soil and plant water status in *Vitis vinifera*. *J Am S Hort Sci* **127**: 448–454.

Whiting J. 2003. *Selection of grapevine rootstocks and clones for greater Victoria*. Victoria Department of Primary Industries, Melbourne, Australia

Xie H, Konate M, Sai N, Tesfamicael KG, Cavagnaro T, Gilliham M, Breen J, Metcalfe A, Stephen JR, De Bei R, et al. 2017. Global DNA methylation patterns can play a role in defining *terroir* in grapevine (*Vitis vinifera* cv. Shiraz). *Front Plant Sci* **8**: 1860.

Yaish MW, Colasanti J, Rothstein SJ. 2011. The role of epigenetic processes in controlling flowering time in plants exposed to stress. *J Exp Bot* **62**: 3727–3735.

Yuan J-L, Sun H-M, Guo G-P, Yue J-J, Gu X-P. 2014. Correlation between DNA methylation and chronological age of Moso bamboo (*Phyllostachys heterocycla* var. pubescens). *Bot Stud* **55**: 4.

Zhang H, Lang Z, Zhu J-K. 2018. Dynamics and function of DNA methylation in plants. *Nat Rev Mol Cell Biol* **19**: 489–506.

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW-L, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, et al. 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *arabidopsis*. *Cell* **126**: 1189–1201.

Zilberman D. 2017. An evolutionary case for functional gene body methylation in plants and animals. *Genome Biol* **18**: 87.

Chapter 4: Dicistronic tRNA-mRNA transcripts in grapevine (*Vitis vinifera*) display distinct, regional expression patterns that correlate with tRNA expression

Statement of Authorship

Title of Paper	Dicistronic transfer RNA-messenger RNA transcripts in grapevine (*Vitis vinifera*) cv Shiraz display geographically-distinct/regional expression patterns that correlates with transfer RNA expression
Publication Status	☐ Published ☐ Accepted for Publication ☒ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Pastor Jullian Fabres, Na Sai, Steve Pederson, Fei Zheng, , James Breen, , Matthew Gilliham, , Penny Tricker, Carlos M. Rodriguez López, Rakesh David This is an experimental manuscript about identification of dicistronic tRNA-mRNA transcripts in grapevine (*Vitis vinifera*), using a novel pipeline (DiRT) developed to identify dicistronic transcripts from high-throughput RNA sequencing data. Dicistronic tRNA-mRNAs were identified in grapevine leaf and berry samples from 22 commercial vineyards covering six sub-regions of the Barossa Wine growing region, Australia. Findings of our study suggest that the presence and abundance of dicistronic molecules is tissue specific, and that each the sub-region displays a distinct expression profile for dicistronic tRNA-mRNA transcripts.

Principal Author

Name of Principal Author (Candidate)	Pastor Jullian Fabres
Contribution to the Paper	Data generation and analysis, interpretation and wrote manuscript.
Overall percentage (%)	60%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 26-11-19

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Na Sai
Contribution to the Paper	Development of the DiRT pipeline. Analysed subset of RNA-seq data. I hereby certify that the statement of the contribution is accurate.
Signature	Date 11-12-19

Name of Co-Author	Steve Pederson
Contribution to the Paper	Supervise the development of the DiRT pipeline. I hereby certify that the statement of the contribution is accurate.
Signature	Date 13/12/19

Please cut and paste additional co-author panels here as required

Name of Co-Author	Fei Zheng		
Contribution to the Paper	Performed the RT-PCR for candidate's dicistronic tRNA-mRNA genes. I hereby certify that the statement of the contribution is accurate.		
Signature	FeiZheng	Date	10/12/2019

Name of Co-Author	James Breen		
Contribution to the Paper	Reviewed and edited the manuscript and provided feedback for the development of the project. I hereby certify that the statement of the contribution is accurate.		
Signature		Date	02/12/2019

Name of Co-Author	Matthew Gilliham		
Contribution to the Paper	Reviewed and edited the manuscript and provided feedback for the development of the project. I hereby certify that the statement of the contribution is accurate.		
Signature		Date	6/12/19

Name of Co-Author	Penny Tricker		
Contribution to the Paper	Review and edited the manuscript. I hereby certify that the statement of the contribution is accurate.		
Signature		Date	26/11/19.

Name of Co-Author	Carlos M. Rodríguez López		
Contribution to the Paper	Conceived and supervise the design of the DiRT pipeline. Supervised development of the project. Reviewed and edited the manuscript and acted as the corresponding author. I hereby certify that the statement of the contribution is accurate.		
Signature		Date	12/11/2019

Name of Co-Author	Rakesh David		
Contribution to the Paper	Conceived and supervise the design of the DiRT pipeline. Supervised development of the project. Reviewed and edited the manuscript. I hereby certify that the statement of the contribution is accurate.		
Signature		Date	26/11/2019

Dicistronic tRNA-mRNA transcripts in grapevine (*Vitis vinifera*) display distinct, regional expression patterns that correlate with tRNA expression.

Pastor Jullian Fabres[1], Na Sai[1,2], Stephen Pederson[3], Fei Zheng[1,2], James Breen[3], Matthew Gilliham[1,2,4], Penny Tricker[1], Carlos M. Rodríguez López[a,1,5], Rakesh David[a,1,2]

[1] School of Agriculture, Food and Wine, University of Adelaide, Adelaide, Australia

[2] ARC Centre of Excellence in Plant Energy Biology, Waite Research Institute & School of Agriculture, Food and Wine, University of Adelaide, Adelaide, Australia

[3] Bioinformatics Hub, School of Biological Sciences, University of Adelaide, Adelaide, SA, Australia.

[4]ARC Industrial Transformation Training Centre in Innovative Wine Production, Waite Research Institute & School of Agriculture, Food and Wine, University of Adelaide, Adelaide, Australia

[5] Environmental Epigenetics and Genetics Group, Department of Horticulture, College of Agriculture, Food and Environment, University of Kentucky, Lexington KY, U.S.A

[a] Joint senior authorship

Corresponding author: Carlos M. Rodríguez López

Email: carlos.rodriguezlopez@uky.edu

Running title: Detection of dicistronic tRNA-mRNAs in *V. vinifera*

Keywords: tRNA-mRNA dicistronic transcripts, tRNA, grapevine, Mobile RNAs, syntenic genomic arrangement, *Vitis vinifera* L., RNA-seq, transcriptomics

Abstract

Transfer RNAs (tRNA) are crucial adaptor molecules between messenger RNA (mRNA) and amino acids. Recent evidence in plants suggests that dicistronic tRNA-like structures can also act as mobile signals for mRNA transcripts to move between distant tissues. Co-transcription is not a common feature in the plant nuclear genome and, in the few cases where polycistronic transcripts have been found, they include the expression of non-coding RNA species such as small nucleolar RNAs and microRNA clusters. It is not known, however, the extent to which dicistronic transcripts of tRNA and mRNAs are expressed in field-grown plants, or the factors contributing to their expression. To address these questions, we analysed tRNA-mRNA dicistronic transcripts in the major horticultural crop grapevine (*Vitis vinifera*) using a novel pipeline developed to identify dicistronic transcripts from high-throughput RNA sequencing data. We identified dicistronic tRNA-mRNA in grapevine leaf and berry samples from 22 commercial vineyards covering six sub-regions of the Barossa wine growing region, Australia. Of the 124 tRNA genes that were expressed in both tissues, 18 tRNA were expressed forming part of 19 dicistronic tRNA-mRNA molecules. The presence and the abundance of dicistronic molecules were tissue and geographic sub-region specific. In leaf tissue, the expression patterns of dicistronic tRNA-mRNAs significantly correlated with tRNA expression, suggesting that transcriptional regulation of their expression might be linked. We also found evidence of evolutionary conservation of dicistronic candidates in grapevine, and previously reported dicistronic transcripts in *Arabidopsis*, indicating a syntenic genomic arrangement of tRNAs and protein coding genes between species.

Introduction

Polycistronic mRNAs are RNA molecules that contain two or more open reading frames (ORFs). These are usually found in viruses, bacteria, archaea, protozoans and invertebrates (Karginov et al. 2017). Polycistronic transcripts are synthesized when multiple genes forming an operon are coexpressed from a single promoter. These transcripts are then translated into protein from two or more translation initiation sites. This strategy has been described as an efficient mechanism to coordinate gene expression (Karginov et al. 2017). Although polycistronic transcripts are less common in plants, several chloroplast genes are organized in clusters and are co-transcribed in polycistronic primary transcripts and subsequently processed to form mature RNAs (Sugita and Sugiura 1996), reflecting their prokaryotic ancestry (Barkan 2011). The majority of nuclear-encoded genes in plants are monocistronic with a few exceptions, such as certain classes of polycistronic microRNAs (miRNAs) (Merchan et al. 2009) and small nucleolar RNAs (snoRNAs), which are organized in genomic clusters and are transcribed from a common promoter (Kruszka et al. 2003; Leader et al. 1997). These precursor transcripts are processed to mature snoRNA and miRNA molecules. There are also a few reports of dicistronic transcripts encoding genes that are not functionally related to each other such as tRNAs-snoRNA, snoRNA-miRNAs and tRNA-mRNA in some plant species (Kruszka et al. 2003; Michaud et al. 2011; Qu et al. 2015; Zhang et al. 2016); however, the molecular and physiological significance of co-transcription for many of these transcripts is still poorly understood.

Recent work in model plants, *Arabidopsis thaliana* and tobacco, has shed light on the function of dicistronic tRNAs-mRNAs (Zhang et al. 2016). Using transgenic lines, Zhang et al. (2016) demonstrated that tRNA-like structures (TLSs), when co-transcribed with mRNA transcripts, could act as mobility signals, triggering the systemic movement of the mRNA between roots and

shoots. Notably, the mRNA components of the dicistronic transcripts were also shown to be translated into functional proteins. Endogenously produced tRNA-mRNA dicistronic transcripts have also been detected in *A. thaliana* suggesting that functional tRNA and tRNA-like structures could act as non-autonomous signals in plants able to deliver functional mRNAs to distantly located tissues. Beyond their canonical role in protein translation, tRNAs have been also demonstrated to function in other chemical transformations, for example, delivering amino acids during lipid modification and antibiotic biosynthesis (Banerjee et al. 2010; Francklyn and Minajigi 2010).

In grapevine (*Vitis vinifera*), the effect of growth environment on gene expression has been extensively studied (Dal Santo et al. 2013; Liu et al. 2012; Pontin et al. 2010; Sun et al. 2015a). Several studies have identified small non-coding RNAs (sRNAs) in grapevine that can influence development in response to environmental stimuli. Among these sRNAS, miRNAs respond to low temperature treatment (Sun et al. 2015b), application of exogenous gibberellin (Han et al. 2014) and viral infection (Alabi et al. 2012). In addition, studies have shown that miRNAs present tissue specificity in grapevine (Pantaleo et al. 2010; Wang et al. 2011). Bester et al. (2017) identified sRNA species in grapevine phloem. Notably, this study also showed the non-random manner in which tRNA-derived sRNAs originated (Bester et al. 2017). A study looking at the effect of grafting in grapevine identified more than 3000 genes transporting mRNA across graph junctions including genes for response to abiotic stress and signal transduction (Yang et al. 2015). Moreover, there is evidence of both passive and selective mechanisms of mRNA trafficking in grapevine (Yang et al. 2015) and Zhang et al. (2016) confirmed that 11% of the mobile mRNA also had TLS motifs in their coding sequence or 3’ UTR.

We hypothesized that dicistronic tRNA-mRNA transcripts would be transcribed differentially between different grapevine tissues and in growing regions with different environments. As a first step towards identifying such transcripts, we present DiRT (**Di**cistronic **R**NA **T**ranscripts), a computational pipeline to detect dicistronic transcripts from short-read RNA-seq data that can be adapted for use in any organism. Using this pipeline, we analysed dicistronic tRNA-mRNA transcripts in commercial, field-grown grapevine and assessed the effect of regionality on their expression profile.

Results

RNA-sequencing of *Vitis Vinifera* cv. Shiraz

To identify tRNA-mRNA dicistronic transcripts in *Vitis Vinifera* cv. Shiraz, we performed RNA-seq of libraries from two different tissues, leaf and berry, collected at budburst (E-L 7) and veraison (E-L 35) (Coombe 1995) respectively, from 22 vineyards from the Barossa wine growing region, South Australia, Australia (Figure 1). The region divides into six sub-regions, each with unique combinations of growing environments with factors such as temperature, rainfall, soil type and elevation contributing to differences in plant growth, berry composition and wine characteristics (Robinson and Sandercock 2014).

Sequencing reads were aligned to the *V. vinifera* reference genome (IGGP_12X from EnsemblPlants) with an average mapping percentage of 90% for leaf and 87% for berry samples. We obtained an average of 23 million and 21 million paired-end (2x75 nucleotide) Illumina reads for each leaf and berry sample (three plants per sample, three samples per vineyard) (Supplemental Table S1).

Identification of putative dicistronic tRNA-mRNA transcripts

We searched for combinations of tRNA and adjacently located protein coding mRNA genes that were expressed forming one continuous transcript. With that objective, we developed DiRT, a bioinformatic pipeline to systematically analyse high-throughput, short read-based RNA-sequencing data for actively co-transcribed tRNA-mRNA loci (Figure 2). The pipeline takes into consideration reads mapping in the tRNA, mRNA and the intervening intergenic region to predict dicistronic tRNA-mRNA candidates. Biological replicates were used to estimate background noise and improve the accuracy of the predictions.

The Genomic tRNA Database predicts 609 tRNA genes in the *V. vinifera* genome based on the tRNAscan-SE tool (Chan and Lowe 2016). From these, 116 tRNA genes overlapped with protein coding genes (PCGs) (5′ or 3′ untranslated regions or introns) and were removed from further analysis since such reads could not be unambiguously assigned to either the tRNA or the PCG.

Using DiRT, we detected 124 and 90 transcribed tRNA genes (read count ≥ 1) in leaf and berry tissue samples, respectively, across all sub-regions. Individual tRNA genes displayed a wide-range of transcript abundances covering the 20 isoacceptor families in both leaves and berries, showing a distinct tRNA expression profile across the six regions analysed (Supplemental Fig S1 and Supplemental Table S2). We assembled combinations of tRNA-PCGs and identified 81 expressed tRNA-mRNA combinations (Figure 3A) in leaves and 50 in berries. As the intergenic region between the transcribed tRNA and mRNA for sequence reads would be indicative of co-transcription, we tested the significance of reads in the intergenic region to eliminate background noise attributable to DNA contamination or spurious transcription events that would not be observed in biological replicates. tRNA-mRNA combinations were selected for further analysis only if the coverage of their intergenic region was significantly higher (t-test, FDR < 0.05) than

reads detected in the two closest introns (Figure 3B). Finally, candidates that passed both tests were tested for continuous read coverage in the intergenic region indicating transcriptional read-through of the region between the tRNA and the mRNA (Figure 3C). DiRT identified 16 dicistronic tRNA-mRNA transcripts in leaves and nine in berries, of which six were present in both tissues (Table 1) across 13 of the 19 *V. vinifera* chromosomes. Sequencing coverage was significantly higher (t-test, FDR < 0.05) in intergenic regions than in the first two introns of dicistronic tRNA-mRNA pairs. Conversely, no significant difference in coverage was observed for tRNA-mRNA pairs deemed non-dicistronic (Supplemental Fig S2).

In total, 19 individual tRNA genes, representing 13 isoacceptor families were found to be dicistronic with the neighbouring protein coding genes, among which, glycine tRNA genes were the most common. We validated, through RT-PCR, two randomly selected tRNA-mRNA dicistronic candidates (i.e. $tRNA^{ValCAC}$-VIT_15s0046g02860 and $tRNA^{ProTGG2.9}$-VIT_18s0001g09050) identified using DiRT. Primers were design to perform PCR amplification from the tRNA to the intergenic region from cDNA of leaf samples (Figure 4A). For the two candidates tested, a single band of the expected product size was obtained (Figure 4B). Sanger sequencing of the PCR product confirmed the amplification of the intergenic regions (Figure 4C, Supplemental Table S3.1 and Supplemental Table S3.2).

Characteristics of grapevine dicistronic tRNA-mRNA candidates

We searched for combinations of tRNA and adjacently located protein coding mRNA genes that were expressed forming one continuous transcript. With that objective, we developed DiRT, a bioinformatic pipeline to systematically analyse high-throughput, short read-based RNA-sequencing data for actively co-transcribed tRNA-mRNA loci (Figure 2). The pipeline takes into consideration reads mapping in the tRNA, mRNA and the intervening intergenic region to

predict dicistronic tRNA-mRNA candidates. Biological replicates were used to estimate background noise and improve the accuracy of the predictions.

The Genomic tRNA Database predicts 609 tRNA genes in the *V. vinifera* genome based on the tRNAscan-SE tool (Chan and Lowe 2016). From these, 116 tRNA genes overlapped with protein coding genes (PCGs) (5′ or 3′ untranslated regions or introns) and were removed from further analysis since such reads could not be unambiguously assigned to either the tRNA or the PCG.

Using DiRT, we detected 124 and 90 transcribed tRNA genes (read count ≥ 1) in leaf and berry tissue samples, respectively, across all sub-regions. Individual tRNA genes displayed a wide-range of transcript abundances covering the 20 isoacceptor families in both leaves and berries, showing a distinct tRNA expression profile across the six regions analysed (Supplemental Fig S1 and Supplemental Table S2). We assembled combinations of tRNA-PCGs and identified 81 expressed tRNA-mRNA combinations (Figure 3A) in leaves and 50 in berries. As the intergenic region between the transcribed tRNA and mRNA for sequence reads would be indicative of co-transcription, we tested the significance of reads in the intergenic region to eliminate background noise attributable to DNA contamination or spurious transcription events that would not be observed in biological replicates. tRNA-mRNA combinations were selected for further analysis only if the coverage of their intergenic region was significantly higher (t-test, FDR < 0.05) than reads detected in the two closest introns (Figure 3B). Finally, candidates that passed both tests were tested for continuous read coverage in the intergenic region indicating transcriptional read-through of the region between the tRNA and the mRNA (Figure 3C). DiRT identified 16 dicistronic tRNA-mRNA transcripts in leaves and nine in berries, of which six were present in both tissues (Table 1) across 13 of the 19 *V. vinifera* chromosomes. Sequencing coverage was significantly higher (t-test, FDR < 0.05) in intergenic regions than in the first two introns of

dicistronic tRNA-mRNA pairs. Conversely, no significant difference in coverage was observed for tRNA-mRNA pairs deemed non-dicistronic (Supplemental Fig S2).

In total, 19 individual tRNA genes, representing 13 isoacceptor families were found to be dicistronic with the neighbouring protein coding genes, among which, glycine tRNA genes were the most common. We validated, through RT-PCR, two randomly selected tRNA-mRNA dicistronic candidates (i.e. $tRNA^{ValCAC}$-VIT_15s0046g02860 and $tRNA^{ProTGG2.9}$-VIT_18s0001g09050) identified using DiRT. Primers were design to perform PCR amplification from the tRNA to the intergenic region from cDNA of leaf samples (Figure 4A). For the two candidates tested, a single band of the expected product size was obtained (Figure 4B). Sanger sequencing of the PCR product confirmed the amplification of the intergenic regions (Figure 4C, Supplemental Table S3.1 and Supplemental Table S3.2).

Regional patterns of dicistronic expression

We next assessed if the geographical origin of the samples had an effect on the expression of dicistronic transcripts. We first analysed the expression of all tRNAs expressed in each tissue and we used hierarchical clustering to group sub-regions according to their tRNA expression patterns in leaves and berries. Both tissues presented two main clusters containing three sub-regions each (Figure 6). The tRNA expression in Eastern Edge and Northern Grounds clustered together in both tissues, while the clustering of the four other subregions were tissue dependent. We then analysed the expression of the tRNA genes, the intergenic regions and PCGs forming dicistronic transcripts independently. We used RNA-seq reads mapping specifically to the intergenic region as a proxy to estimate expression patterns of dicistronic candidates, as reads mapping to the flanking tRNA and PCG loci could originate from both monocistronic and dicistronic transcripts (Figure 6). Sub-regional clusters for tRNAs forming part of dicistronic

constructs were similar to those observed for all expressed tRNAs in both tissues (Figure 6). In leaf, one of the main clusters (SG, EE and NG) was the same for all expressed tRNAs and tRNAs that were part of dicistronic constructs. While in berry, EE/NG and CG/WR clusters were the same in all expressed tRNAs and dicistronic tRNAs. EE/NG and CG/WR clustered together in both tissues, while SG and EV clustering was tissue dependant. When the expression of the intergenic regions and dicistronic PCGs was used rather than tRNA, sub-regional clustering was tissue and dicistronic construct component specific (intergenic region or PCG) (Figure 6).

Local Fisher Discriminant Analysis (LFDA) on expression patterns of all expressed tRNAs and the dicistronic construct components (i.e. dicistronic tRNAs, PCG and intergenic regions), showed that PCGs' expression patterns occupying and a unique eigen space, while dicistronic tRNAs and intergenic regions shared the eigen space occupied by all expressed tRNAs (Supplemental Fig S5). Consistent with this observation, correlation analyses of the expression of the different part of the candidates dicistronic transcripts showed that the absolute values of Pearson correlation coefficients were generally higher between the expression of dicistronic tRNAs and the expression of the intergenic region than between the expression of PCGs and the expression of the intergenic regions on both tissues (Supplemental Table S5). These correlations were only significant (Pearson correlation, p-value < 0.05) between dicistronic tRNAs and intergenic regions in leaves (Supplemental Table S5).

Discussion

In this study, using an RNA-seq approach, we found that 15.3% (19/124) of all expressed tRNAs in grapevine leaf and berry samples were putatively expressed in a dicistronic manner, with neighbouring protein coding genes. We developed DiRT, a customised, computational pipeline to specifically detect co-transcribed tRNA and mRNA candidates using stringent criteria. Using

DiRT we were able to identify dicistronic transcripts in two different grapevine tissues (i.e. leaf and berry) sampled from commercial vineyards. Validation of the pipeline to accurately predict dicistronic candidates was confirmed through RT-PCR detection and Sanger sequencing of dicistronic candidates in leaf samples.

Interestingly, of the 12 tRNA isoacceptor families (representing 15 distinct anticodons) found to be dicistronic in *Vitis vinifera*, 11 tRNA families have also found to be dicistronic in *A. thaliana*, suggesting conservation amongst plant tRNAs that form dicistronic transcripts. Furthermore, the 11 tRNA families in *A. thaliana* were dicistronically associated with transcripts that were demonstrated to be mobile between roots and shoots (Thieme et al. 2015; Zhang et al. 2016). Among these tRNA coding for Gly^{GCC} and Met^{CAT} were able to mobilise mRNA transcripts to different tissues as part of a fusion construct and translate into functional proteins in grafted *A. thaliana* plants indicating that these tRNA were able to confer mobility to these transcripts. (Zhang et al. 2016). This suggests a non-autonomous role for dicistronic tRNAs in delivering mRNA transcripts to distantly located tissues. A recent study also revealed that mobile RNA transcripts are enriched in the modified base 5-methylcytosine (m^5C), indicating a role of RNA cytosine methylation in systemic RNA movement (Yang et al. 2019). In plants, tRNA and mRNA m^5C methylation is mediated by the methyltransferases DNMT2 and TRM4B (Burgess et al. 2015; Cui et al. 2017; David et al. 2017) and loss of these enzymes was demonstrated to impair transcript mobility (Yang et al. 2019). Future studies will need to be undertaken to determine if the dicistronic tRNAs identified in this study also confer mRNA mobility and to assess the role of cytosine methylation in mRNA transport in grapevine.

Interestingly, for four of the 19 dicistronic candidates we also observed sequence conservation between *A. thaliana* and *V. vinifera* for the protein coding gene and the adjacently co-transcribed

tRNA genes. The dicistronic activity at these conserved loci may provide an explanation of why such syntenic clusters are conserved through evolution and suggests that these transcripts may have an important functional role.

Of the 19 dicistronic tRNA genes identified in the *Vitis vinifera* genome, 18 were located fewer than 1000 base pairs from the co-transcribed protein coding gene (median distance 133 bp). Our findings suggest that tRNA genes and protein coding genes need to be closely positioned in the genome in order to form dicistronic transcripts. Similar observations were obtained in *A. thaliana*, where the majority of the previously identified PCGs forming part of mobile dicistronic transcripts were located less than 200 bp from their partner tRNA (Zhang et al. 2016). This genomic proximity rule has been found to extend to the dicistronic tRNA-snoRNA genes that have been identified in higher plants (*A. thaliana, M. truncatula, P. trichocarpa, O. sativa and B. distachyon*) in which the intergenic region ranged between 1 to 16 base pair (Michaud et al. 2011).

Previous studies have indicated that a large proportion of mobile transcripts are also highly abundant (Calderwood et al. 2016; Thieme et al. 2015). This suggests that passive diffusion of these transcripts through the phloem may contribute to their mobility. A significant proportion (11.4%) of these transcripts was subsequently shown to be dicistronically associated with tRNA (Zhang et al. 2016). However, when we assessed the expression levels of mRNA and tRNA that formed dicistronic transcripts in grapevine, we did not observe higher abundance of these transcripts in either tissue analysed. Thus, in our study, the expression level of the tRNA and mRNA was not a good indicator of the formation of dicistronic transcripts.

In eukaryotes, tRNA and mRNA are transcribed by different types of RNA polymerase. RNA polymerase II (Pol II) transcribes protein coding genes and RNA polymerase III (Pol III) for a

variety of genes that generally encode for RNAs with catalytic activity such as tRNA (Schramm and Hernandez 2002). Results from Kruszka et al. (2003) suggested that, in *A. thaliana*, dicistronic tRNA-snoRNA are transcribed by Pol III from the tRNA gene promoter. However, Pol III transcribes genes shorter than 400 base pairs (Schramm and Hernandez 2002) and the dicistronic transcripts identified in our study were considerably longer (between 1486 to 6002 bp) suggesting Pol III may not be co-transcribing these transcripts. A comparative analysis of flowering species showed a poly-T stretch immediately downstream of ≥ 90% of tRNA genes (Michaud et al. 2011). Additionally, this study reported that the few tRNAs lacking poly-Ts were capable of forming dicistronic transcript with snoRNAs. The authors hypothesized that the lack of the poly T transcriptional termination signal could be a possible explanation for why these transcripts were transcribed as a single unit by Pol III. Sequence analysis of the upstream and downstream sequences of the dicistronic tRNAs identified in our study revealed canonical elements previously associated with transcription start and termination (Michaud et al. 2011; Yukawa et al. 2000). In particular, all dicistronic tRNA transcripts we identified had a poly-T termination signal suggesting the transcriptional read-through mechanism for these transcripts is different from tRNA-snoRNAs. It remains to be determined whether one of, or both polymerases are involved in transcribing tRNA-mRNA dicistronic transcripts.

The expression patterns of all genomic features studied (i.e. tRNAs, PCGs, and intergenic regions (considered a proxy for dicistronic tRNA-mRNA transcripts) were found to be organ specific and sensitive to regional environmental differences. The effect of organ and environment on PCG and tRNA gene expression has been extensively studied before in grapevine (Bester et al. 2017; Dal Santo et al. 2013; Liu et al. 2012; Pontin et al. 2010; Sun et al. 2015a). Special effort has been put into deciphering the effect that the growing environment has

on fruit quality traits associated with wine regionality (Jackson and Lombard 1993; van Leeuwen 2009; Zsófi et al. 2011). However, the effect that the environment and tissue have on tRNA expression and on dicistronic transcript expression has not been previously described. Our results show that the expression patterns of dicistronic transcript-forming tRNA genes mimic those of all expressed tRNA (Supplemental Fig S5). We also found that the expression of dicistronic tRNA-mRNAs, measured as the expression of the intergenic region, showed a higher correlation with that of dicistronic tRNA than with that of dicistronic PCGs in both tissues. Although this correlation was only statistically significant in leaves, the lack of statistical significance in berry samples could be due to the low number of dicistronic transcripts identified in berries compared to leaves (9 vs 16 respectively). Taken collectively, our results suggest that environmentally induced dicistronic tRNA-mRNA expression is, at least partially, directed by the mechanisms regulating tRNA expression.

Prior to this work, dicistronic transcripts had been found in a handful of plant species grown under laboratory conditions. Comparative analysis of dicistronic transcripts in grapevine and *A. thaliana* indicate that the genomic clustering of the tRNA and protein coding gene is evolutionarily conserved between both species. A better understanding of their prevalence across the plant kingdom is needed to begin to understand the biological roles that these types of transcripts may play in plants. We provide here the first evidence for dicistronic transcription in a commercially important crop species grown in field conditions, and highlight that differential tRNA expression could be relevant for differential dicistronic transcript formation. Our approach can be utilized to survey distribution patterns of dicistronic transcripts in different plant species and tissues which will help elucidate the molecular function of these transcripts in plant signalling.

Material and methods

Sampling material

Tissues were sampled from own-rooted grapevines (*Vitis vinifera*) cv. Shiraz from 22 commercial vineyards located in the Barossa wine zone (South Australia, Australia). Vineyards were selected as part of a larger study of Barossa Terroir (Xie et al. 2017). Vineyards were chosen to be representative of the climate, soil and management practices that are used in the different Barossa sub-regions. These sub-regions are the Eden Valley (EV) (3 vineyards), Northern Grounds (NG) (4 vineyards), Central Grounds (CG) (4 vineyards), Southern Grounds (SG) (3 vineyards), Eastern Edge (EE) (4 vineyards) and Western Ridge (WR) (4 vineyards).

Leaf and a berry bunches were collected from nine plants in each of three rows in each vineyard (total of 198 plants) during the 2016 growing season. The first fully expanded leaf at budburst (E-L 7) (Coombe 1995) was collected from three nodes per plant and pooled into a single sample per plant. Berries were collected at veraison (E-L 35) (Coombe 1995) from three different bunches per plant (i.e. three berries per bunch). All samples were taken before dawn (between 10:00 pm and sunrise) to minimise variability associated with differences in plant water status (Williams and Araujo 2002). Samples were snap-frozen in liquid nitrogen in the vineyards and stored at -80◦C.

RNA extraction and RNA-seq library preparation

Total RNA was extracted from each sample using the Spectrum Plant Total RNA kit (Sigma-Aldrich) following the manufacturer's instructions and including DNase treatment. Three samples per vineyard were generated by pooling 2 µg of total RNA from three plants from the same row in the vineyard for a total of 66 pools. Ribosome was depleted in 6 µg of RNA from

each pool using the Dynabeads mRNA purification kit (Ambion, Invitrogen) following the manufacturer's instructions. Ribosomal depleted RNA (25 ng per pool) was used as input for library preparation using the NEBNext Ultra RNA Library Prep Kit for Illimina (New England Biolabs Inc). Libraries were sequenced using Illumina NextSeq High Output 75 bp pair-end (Illumina Inc., San Diego, CA, United States) at the Australian Genome Research Facility (Adelaide, SA, Australia). Reads were trimmed and filtered using AdapterRemoval v2.2.1 (Schubert et al. 2016) using default settings. Alignment of trimmed reads to the *Vitis vinifera* genome reference IGGP_12X obtained from EnsemblPlants 36 was performed using Hisat2 v2.1.0 (Kim et al. 2015). BAM files from samples from the same vineyard were merged, sorted and indexed using SAMtools v1.8 (Li et al. 2009). Mapped reads were counted to genomic features using featureCounts v1.5.2 (Liao et al. 2014), with the minimum mapping quality score for a read to be assigned to a feature was set to 10 (Liao et al. 2013). The merged, sorted and indexed BAM files were then directly input into the R environment in order to identify the dicistronic tRNA-mRNA transcripts.

In silico detection of dicistronic tRNA-mRNA transcripts

DiRT (**Di**cistronic **R**NA **T**ranscripts) is a custom pipeline implemented in the R environment and source codes are available at GitHub (https://github.com/CharlotteSai/DiRT). While the pipeline was developed for analysing grapevine RNA-seq data, it can be adapted for use in other species provided a genomic tRNA annotation is available. Firstly, protein coding gene (PCG) information and coordinates were downloaded from Ensembl Plants (release 45) (http://plants.ensembl.org/Vitis_vinifera/Info/Index) and the chromosomal coordinates of tRNA genes were extracted from the Genomic tRNA database using tRNAscan-SE based on predicted structure analysis (http://gtrnadb.ucsc.edu/GtRNAdb2/genomes/eukaryota/Vvini/). We used

BEDTools version 2.25 (Quinlan and Hall 2010) to determine the relative location of PCGs in relation to all tRNA genes. Predicted tRNAs overlapping with PCGs were discarded for further analysis. tRNAs expressed (i.e. above 1 raw read) in leaf and berry samples were identified using GenomicRanges (Lawrence et al. 2013). To infer putative co-transcription, first we filtered the RNA-seq data for genomic regions where both tRNA loci and closest neighbouring gene were transcribed (Raw read >= 1 for tRNAs and raw read >= 10 in PCGs), independently of their DNA-strand. In order to identify region-specific putative dicistronic transcripts, each Barossa sub-region was interrogated separately.

The selected candidate tRNA-mRNA transcripts were then scanned for dicistronic transcripts. We demanded that, first; the sequencing coverage of the intergenic region must be significantly higher than the intron closest to the intergenic region and the second intron closest to intergenic region. To achieve this, reads for each base of the intergenic region, the closest intron and the second closest intron were counted by the coverage method from the *GenomicRanges* package (Lawrence et al. 2013) using merged BAM files for each region to obtain total coverage for each region. Then significant differences in average coverage between the intergenic region and the two closest introns were determined by a t-test including all regions as biological replicates. The complete set of p-values were adjusted using the Benjamini-Hochberg false discovery rate (FDR) (Benjamini and Hochberg 1995) and intergenic regions with higher mean coverage than both introns, and an FDR-adjusted p-value < 0.05 were included for subsequent steps of the pipeline.

tRNA-mRNA transcripts passing the first condition were further filtered for those with uninterrupted sequencing coverage spanning the tRNA, the intergenic region and the mRNA by selecting candidates with at least one count for every base in the intergenic region. This

condition was implemented to make sure that at least one entire molecule of the dicistronic transcript had been potentially produced.

Finally, dicistronic candidates with continuous coverage in the intergenic region were manually inspected using IGV (Robinson et al. 2011) for visual validation of continuous coverage. The candidates passing both of the t-tests and continuously coverage examination were deemed putative tRNA-mRNA dicistronic transcripts. Multidimensional scaling (MDS) were analysed in the R environment using the function *plotMDS* from the limma package (Ritchie et al. 2015).

RT-PCR confirmation

Complementary DNA was synthesized from the same total RNA used for the RNA-seq using SuperScript IV first strand synthesis system (Invitrogen, 18091050) following the manufacturer's instructions. Complementary DNA was synthesised using gene specific reverse primers that aligned to the second exon for each candidate gene (15s0046g02860-Reverse and 18s0001g09050-Reverse) (+RT) and the reverse primer of the gene Elongation Factor 1-alpha (EF1a) as a positive control. Negative controls for the cDNA synthesis (-RT) in which reverse transcriptase enzyme was omitted were included for each of the dicistronic candidate. Resulting cDNA was diluted 1:10 and 2 µl was used for RT-PCR. The RT-PCR reaction was conducted using Kappa Taq PCR Kit (Kappa Biosystem, KK1020) following manufacture instructions. The amplification program used was 95°C for 3 min, 37 cycles at 95°C for 30 sec followed by 60°C for 30 sec and 72°C for 50 sec and finally 72°C per 2 min. For the candidate $tRNA^{ValCAC1.7}$-VIT_15s0046g02860 we used primers tRNA-Val-forward and Intergenic-tRNAVal-reverse (376 bp). For candidate $tRNA^{ProTGG2.9}$- VIT_18s0001g09050 primers tRNA-Pro-forward and Intergenic-tRNAPro-reverse (172 bp) were used. Negative control for the PCR reaction (-Ctr) contained all components for the reaction except the cDNA template. RT-PCR products were

analysed by agarose gel electrophoresis and SYBR Safe DNA gel staining (ThermoFisher Scientific, S33102). RT-PCR products were purified using PCR Clean-up (Macherey-Nagel, 740609.250) following the manufacturer's instructions. Sanger sequencing was performed at the Australian Genome Research Facility. Oligonucleotides used for RT-PCR are listed in Supplemental Table S6. Matching of the sequencing results for both putative dicistronic pairs and the expected sequence of each locus was confirmed using BLAST (blastn) with default settings (Zhang et al. 2000).

Regional effect on the expression of dicistronic tRNA:mRNAs

To determine the effect of the region of origin on dicistronic tRNA:mRNAs, we independently compared the expression levels of all three components of the identified dicistronic transcripts (i.e. tRNAs, intergenic regions and PCGs). Similarly, the regional expression levels of all expressed tRNAs were compared. Briefly, mapped reads for each selected genomic feature obtained from featureCounts, were analysed in the R environment for plotting the gene expression through heatmaps. The expression of all tRNA, dicistronic tRNAs, intergenic region and the dicistronic genes (logCPM) were plotted using the pheatmap (Kolde 2015). Local Fisher LFDA was performed in the R environment using the package *lfda* (Sugiyama 2006) to the expression (logCPM) values of all tRNA, dicistronic tRNAs, intergenic region and the dicistronic genes. In order to determine whether primary mRNA or the tRNA is driving the expression of candidate dicistronic tRNA-mRNA transcripts, we carried out Pearson correlation analyses between the expression of tRNA vs intergenic region and PCGs vs intergenic region for each dicistronic candidate identified. Pearson correlation analysis (p-value < 0.05) was performed using the R function *cor.test*(). Read counts of the intergenic regions were used as a proxy to define the expression of dicistronic transcripts. The rationale behind this lies on the

assumption that reads mapping to the intergenic region can only be the result of the sequencing of a dicistronic RNA molecule, while reads mapping to tRNA genes and PCGs could result both from the expression of monocistronic and dicistronic transcripts (i.e. tRNA genes and PCGs pairs forming two independent RNA molecules or a single RNA molecules respectively) (Figure 6).

Gene analyses

A bar plot was made to represent the distance (bp) between the tRNA and its proximal gene. Pairs of expressed tRNA-mRNA were split in two groups depending if they formed dicistronic or monocistronic transcripts. A non-overlapping sliding window approach (200bp) was used to count the number of pairs of genes of each type. Expression of the dicistronic genes and dicistronic tRNA was assessed by plotting their expression values against the distribution of the total gene expression for each tissue from the RNA-seq data. Gene annotation for dicistronic genes was obtained from EnsemblPlants release 45 (http://plants.ensembl.org/Vitis_vinifera/Info/Index), BLASTP search from NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi), from the grapevine V1 annotation of PN40024 (Jaillon et al. 2007) and from the Additional file 1 of Cramer et al. (2020). Protein information and gene ontology terms were obtained from UniProt (https://www.uniprot.org/uniprot). GO enrichment analysis was performed from Gene Ontology Consortium (http://geneontology.org/) with default settings.

Motif analysis

Upstream and downstream sequence from the dicistronic tRNA was obtained from Genomic tRNA data and analysed in Weblogo (Crooks et al. 2004) for sequence analysis using default settings.

Acknowledgments

This study was funded through a Pilot Program in Genomic Applications in Agriculture and Environment Sectors jointly supported by the University of Adelaide and the Australian Genome Research Facility Ltd. PJF was supported by Graduate Research Scholarships from Wine Australia (PH1503) and the University of Adelaide. NS was supported by a summer scholarship from the ARC Centre of Excellence in Plant Energy Biology (CE1400008). Dr Lopez is currently partially supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, Hatch Program number 2352987000. We thank the Barossa Grounds Project and the growers who allowed us to collect samples and supplied information about their vineyards. We thank Kendall Corbin for performing the DNA extraction of the leaf samples. We thank Cassandra Collins for the experimental design and collection of the plant material. We thank Roberta DeBei, Sandra Milena Mantilla, Annette James, and Valentin Olek who helped with the sample collection. We thank Stephen Tyerman for his contribution in the development of DiRT. We thank Timothy Cavagnaro and Andrew Metcalfe for their contribution in the experimental design.

Availability of data and materials

The full sequencing data are available on SRA database under the accession number PRJNA591273. The full sequencing data are available on SRA database under the accession number PRJNA591273.

Table

Table 1. Dicistronic tRNA-mRNA candidates identified from RNA-seq data in leaves and berries of grapevine.

tRNA ID	Gene ID (Ensembl accession number)	Tissue(s)
tRNA-Phe-GAA-1-4	VIT_07s0005g02200	Leaf
tRNA-Ala-AGC-1-7	VIT_14s0066g02600	Leaf
tRNA-Pro-CGG-2-2	VIT_14s0060g01370	Leaf
tRNA-Arg-TCG-2-2	VIT_09s0002g04750	Leaf
tRNA-Gly-GCC-1-1	VIT_02s0154g00160	Leaf
tRNA-Gly-GCC-1-5	VIT_07s0005g02990	Leaf
tRNA-Gly-GCC-1-6	VIT_08s0058g00460	Leaf
tRNA-Val-CAC-1-7	VIT_15s0046g02860	Leaf
tRNA-Asn-GTT-2-2	VIT_18s0001g12620	Leaf
tRNA-Met-CAT-1-3	VIT_07s0129g00230	Leaf
tRNA-His-GTG-8-1	VIT_17s0000g06990	Leaf and Berry
tRNA-Gly-TCC-1-6	VIT_13s0064g00200	Leaf and Berry
tRNA-Thr-AGT-1-4	VIT_00s0322g00020	Leaf and Berry
tRNA-Ile-AAT-3-1	VIT_04s0023g03700	Leaf and Berry
tRNA-Pro-TGG-2-9	VIT_18s0001g09050	Leaf and Berry
tRNA-Phe-GAA-1-4	VIT_07s0005g02210	Leaf and Berry
tRNA-Leu-TAA-2-3	VIT_08s0007g03950	Berry
tRNA-Gly-CCC-1-3	VIT_19s0177g00220	Berry
tRNA-Thr-AGT-2-1	VIT_05s0077g01490	Berry

Figure Legends

Figure 1. Geographical location of grapevine tissue samples analysed in this study. Leaf and berry samples were harvested from selected vineyards from the Barossa wine region, Australia. Northern Grounds (NG, blue) Central Grounds (CG, green), Southern Grounds (SG, yellow), Western Ridge (WR, purple), Eastern Edge (EE, red) and Even Valley (EV, orange).

Figure 2. Schematic workflow of the DiRT pipeline to identify dicistronic candidates from RNA-seq data. tRNA gene and protein coding gene coordinates were retrieved from GtRNAdb and Ensembl Plants respectively. tRNA-protein coding gene pairs that occupied contiguous spaces in the genome were assembled and selected for subsequent analysis if they were transcribed based on RNA-seq reads (Raw read >= 1 for tRNAs and raw read >= 10 in PCGs). Dicistronic tRNA-mRNAs are identified by assessing active transcription of the intergenic region. Candidate tRNA-mRNA are selected if the intergenic regions had significantly higher expression (FDR < 0.05) than the closest two introns. Pairs of tRNA-gene are classified as putatively dicistronic tRNA-mRNA if the intergenic region showed continuous sequencing coverage between the tRNA and the mRNA of the protein coding gene.

Figure 3: Assessing active transcription of intergenic region of putative dicistronic tRNA-mRNAs transcripts from RNA-seq data. A) Gene model of dicistronic tRNA-mRNA transcript in which the intergenic region was expressed B) Average read coverage of the intergenic region versus the closest two introns of a tRNA-mRNA combination that passed B) ($tRNA^{Gly\text{-}GCC\text{-}1\text{-}6}$ VIT_08s0058g00460) and failed C) ($tRNA^{Tyr\text{-}GTA\text{-}4\text{-}1}$ VIT_00s0505g00030) the t-test (p-value < 0.05). C) Genome browser view of a candidate dicistronic tRNA-mRNA formed by $tRNA^{Gly\text{-}TCC\text{-}1\text{-}6}$ and VIT_13s0064g00200 identified using the DiRT pipeline.

Figure 4. RT-PCR confirmation of identified dicistronic transcripts. Model of putative dicistronic tRNA-mRNA transcript showing primers used for cDNA synthesis (cDNA rev.) and for the PCR reaction (RT-PCR For and RT-PCR Rev). Confirmation of actively transcribed intergenic region through RT-PCR for candidates A) tRNA ^{ValCAC}-VIT_15s0046g02860 (376 bp) and B) tRNA^{ProTGG}-VIT_18s0001g09050 (172 bp). +RT: cDNA as template, gDNA: genomic DNA was used as a control, -RT: RT-PCR negative control, EF1a: Elongation Factor 1-alpha was used as a positive control (150 bp), -Ctr: PCR negative control. Alignment of the sequenced PCR product for candidate C) tRNA^{ValCAC}-VIT_15s0046g02860 and D) tRNA^{ProTGG}-VIT_18s0001g09050 to the expected PCR product confirmed active transcription of the intergenic region.

Figure 5. Genomic distance between protein coding genes (PGCs) and tRNA for dicistronic and monocistronic transcripts. Vertical bars show the number of neighbouring tRNAs and PCGs pairs forming putative dicistronic tRNA-mRNA transcripts (red), or monocistronic transcripts (blue).

Figure 6. Effect of region of origin on the expression of dicistronic tRNA-mRNAs. Top panel: Schematic representation of RNA-seq reads mapping when originated from a dicistronic transcript (purple bars), a monocistronic tRNA transcript (blue bars) and a monocistronic protein coding gene (red bars). Bottom panel: Heatmap of the expression (logCPM) of all expressed tRNAS, dicistronic tRNAs, intergenic region and dicistronic protein coding genes (rows) for each sub-region from the Barossa Wine growing region (columns) for leaf and berry samples.

Dendrograms represent the hierarchical clustering analysis of the sub-regions according to each genomic feature expression pattern.

Figures

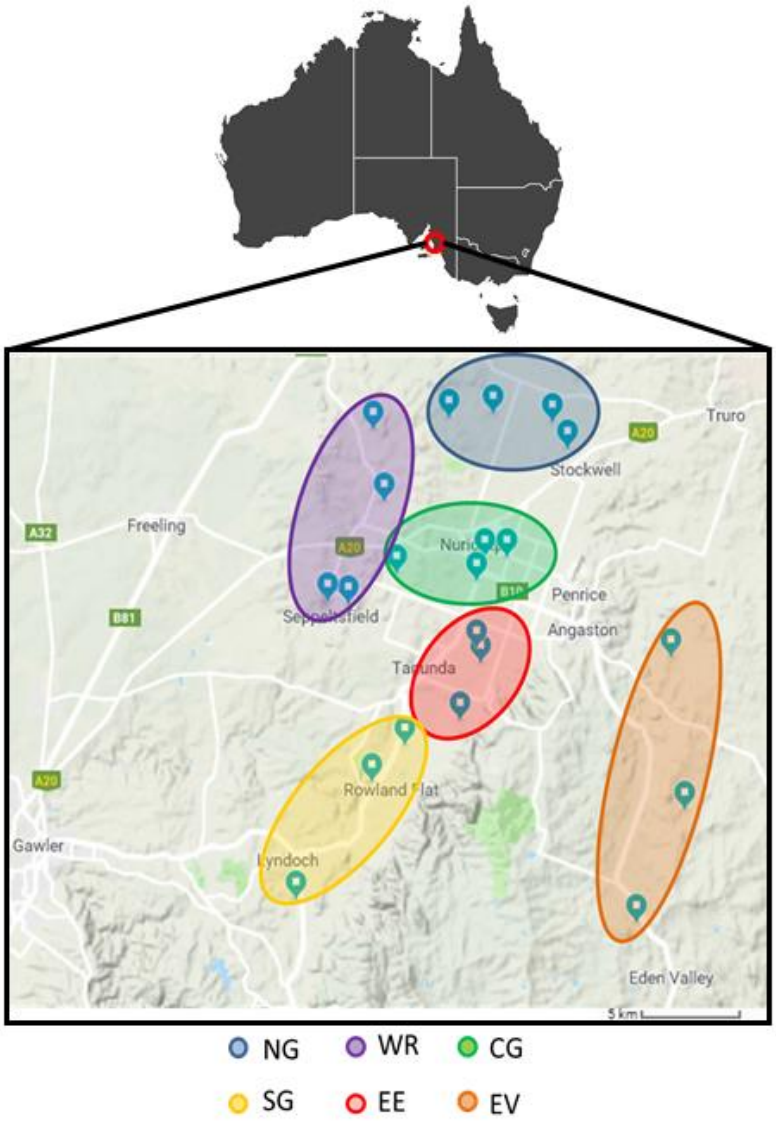


Figure 1. Geographical location of grapevine tissue samples analysed in this study. Leaf and berry samples were harvested from selected vineyards from the Barossa wine region, Australia. Northern Grounds (NG, blue) Central Grounds (CG, green), Southern Grounds (SG, yellow), Western Ridge (WR, purple), Eastern Edge (EE, red) and Even Valley (EV, orange).

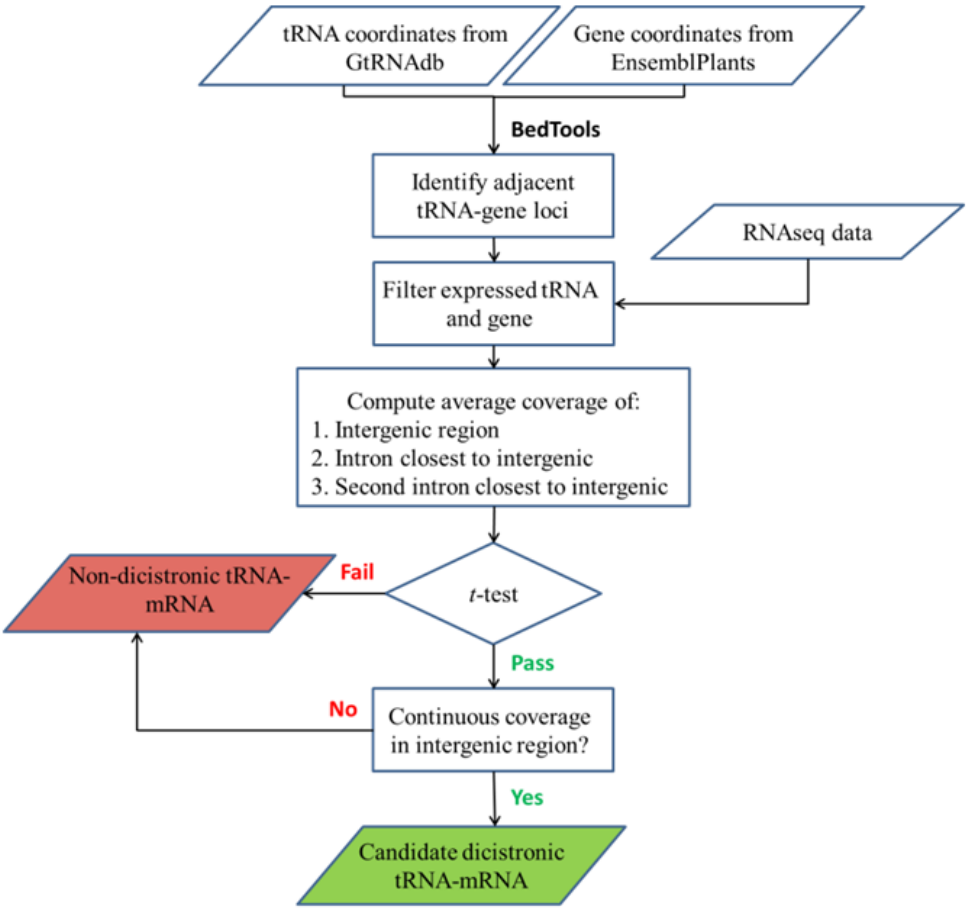


Figure 2. Schematic workflow of the DiRT pipeline to identify dicistronic candidates from RNA-seq data. tRNA gene and protein coding gene coordinates were retrieved from GtRNAdb and Ensembl Plants respectively. tRNA-protein coding gene pairs that occupied contiguous spaces in the genome were assembled and selected for subsequent analysis if they were transcribed based on RNA-seq reads (Raw read >= 1 for tRNAs and raw read >= 10 in PCGs). Dicistronic tRNA-mRNAs are identified by assessing active transcription of the intergenic region. Candidate tRNA-mRNA are selected if the intergenic regions had significantly higher expression (FDR < 0.05) than the closest two introns. Pairs of tRNA-gene are classified as

putatively dicistronic tRNA-mRNA if the intergenic region showed continuous sequencing coverage between the tRNA and the mRNA of the protein coding gene.

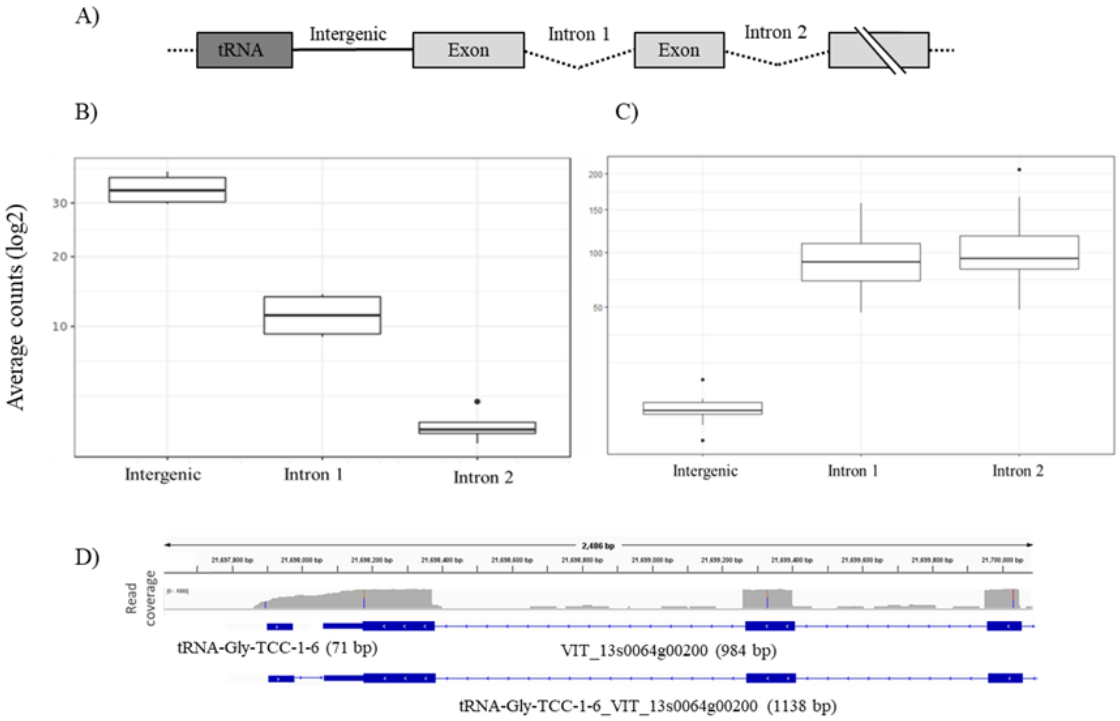


Figure 3: Assessing active transcription of intergenic region of putative dicistronic tRNA-mRNAs transcripts from RNA-seq data. A) Gene model of dicistronic tRNA-mRNA transcript in which the intergenic region was expressed B) Average read coverage of the intergenic region versus the closest two introns of a tRNA-mRNA combination that passed B) ($tRNA^{Gly\text{-}GCC\text{-}1\text{-}6}$ VIT_08s0058g00460) and failed C) ($tRNA^{Tyr\text{-}GTA\text{-}4\text{-}1}$ VIT_00s0505g00030) the t-test (p-value < 0.05). C) Genome browser view of a candidate dicistronic tRNA-mRNA formed by $tRNA^{Gly\text{-}TCC\text{-}1\text{-}6}$ and VIT_13s0064g00200 identified using the DiRT pipeline.

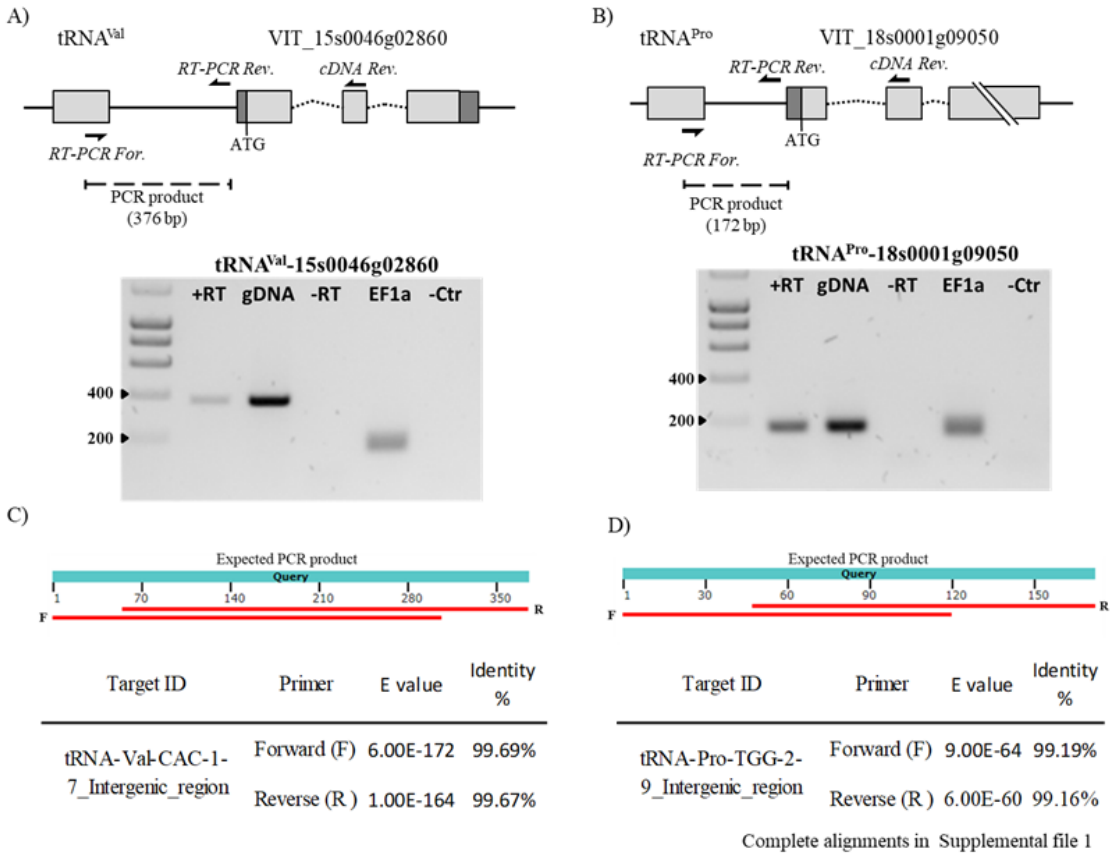


Figure 4. RT-PCR confirmation of identified dicistronic transcripts. Model of putative dicistronic tRNA-mRNA transcript showing primers used for cDNA synthesis (cDNA rev.) and for the PCR reaction (RT-PCR For and RT-PCR Rev). Confirmation of actively transcribed intergenic region through RT-PCR for candidates A) tRNA ^{ValCAC}-VIT_15s0046g02860 (376 bp) and B) tRNA^{ProTGG}-VIT_18s0001g09050 (172 bp). +RT: cDNA as template, gDNA: genomic DNA was used as a control, -RT: RT-PCR negative control, EF1a: Elongation Factor 1-alpha was used as a positive control (150 bp), -Ctr: PCR negative control. Alignment of the sequenced PCR product for candidate C) tRNA^{ValCAC}-VIT_15s0046g02860 and D) tRNA^{ProTGG}-VIT_18s0001g09050 to the expected PCR product confirmed active transcription of the intergenic region.

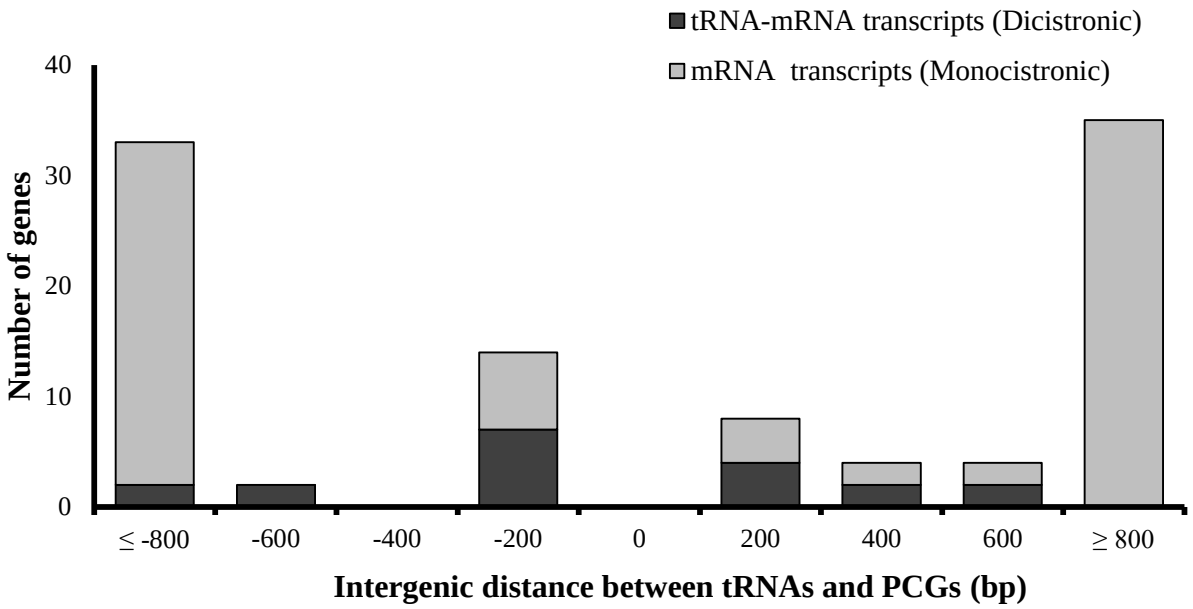


Figure 5. Genomic distance between protein coding genes (PGCs) and tRNA for dicistronic and monocistronic transcripts. Vertical bars show the number of neighbouring tRNAs and PCGs pairs forming putative dicistronic tRNA-mRNA transcripts (black), or monocistronic transcripts (grey).

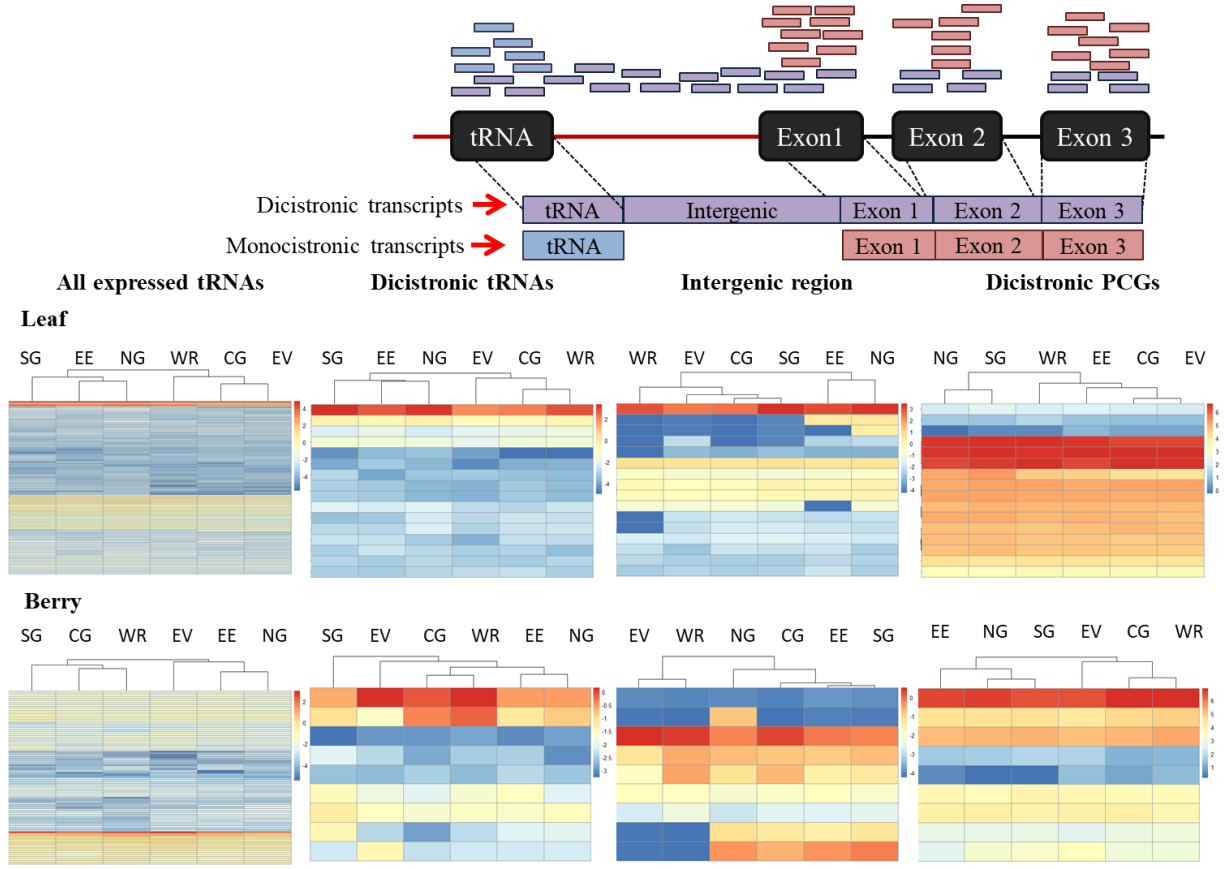
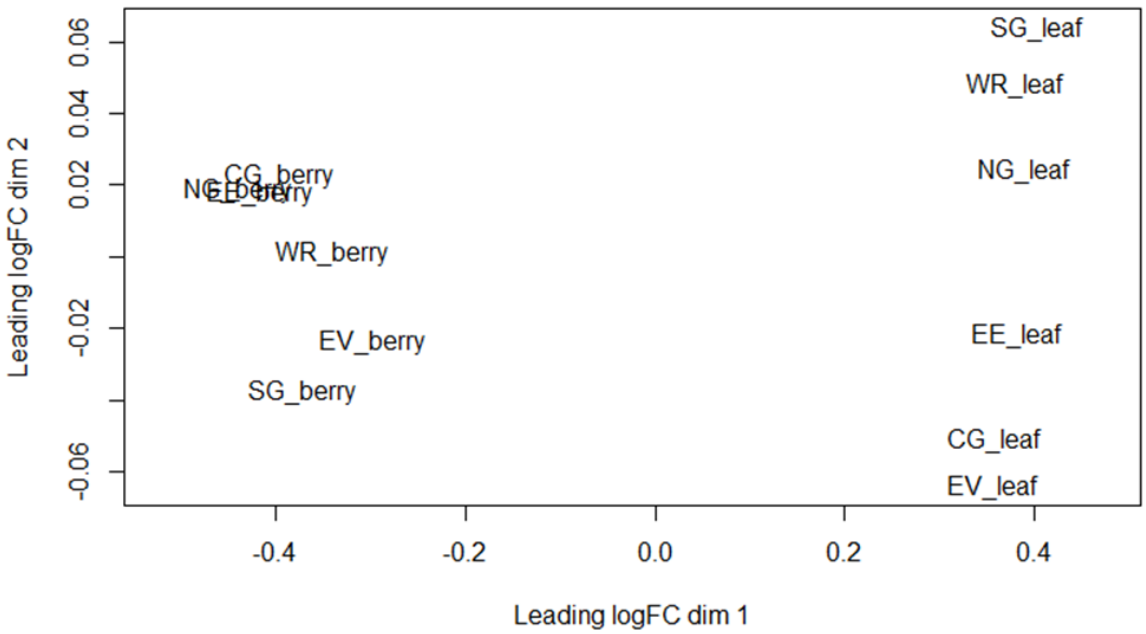
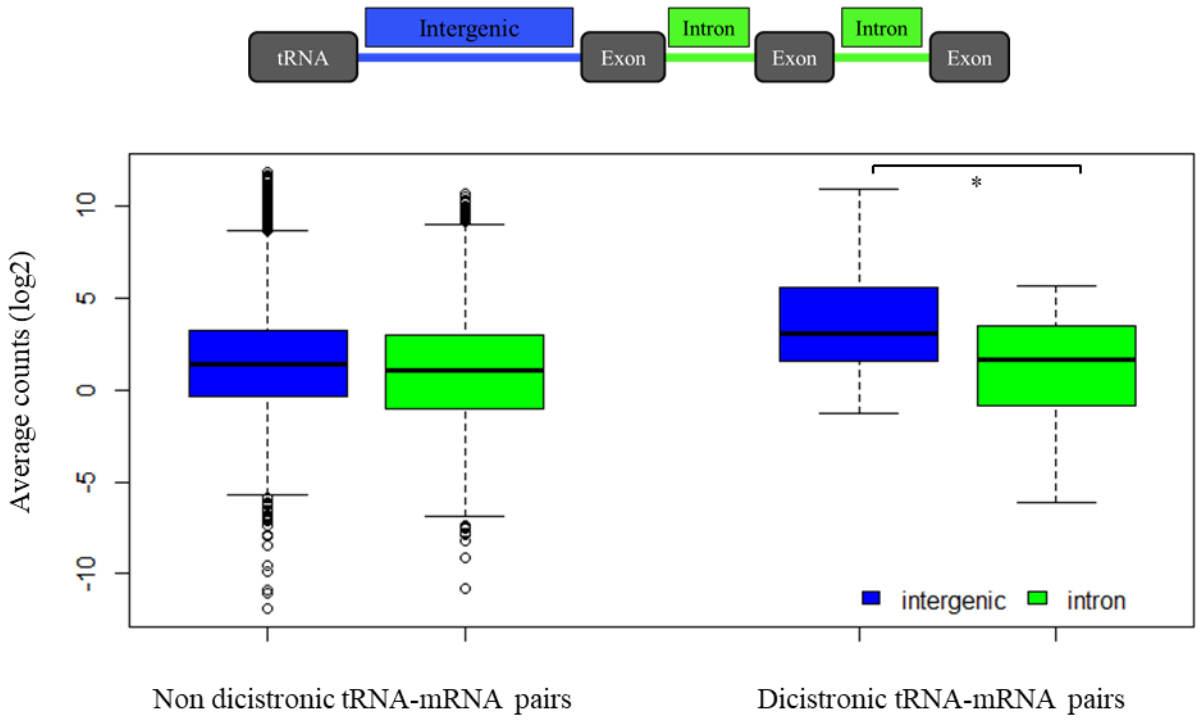


Figure 6. Effect of region of origin on the expression of dicistronic tRNA-mRNAs. Top panel: Schematic representation of RNA-seq reads mapping when originated from a dicistronic transcript (purple bars), a monocistronic tRNA transcript (blue bars) and a monocistronic protein coding gene (red bars). Bottom panel: Heatmap of the expression (logCPM) of all expressed tRNAS, dicistronic tRNAs, intergenic region and dicistronic protein coding genes (rows) for each sub-region from the Barossa Wine growing region (columns) for leaf and berry samples. Dendrograms represent the hierarchical clustering analysis of the sub-regions according to each genomic feature expression pattern.

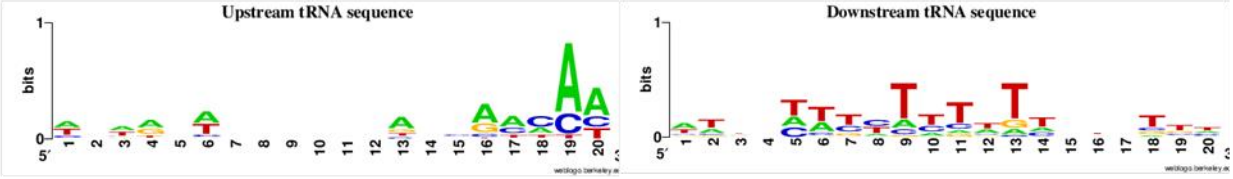


Supplemental Fig S1: MDS plots of total raw counts of all tRNAs expressed in leaf and berry samples for each sub-region.



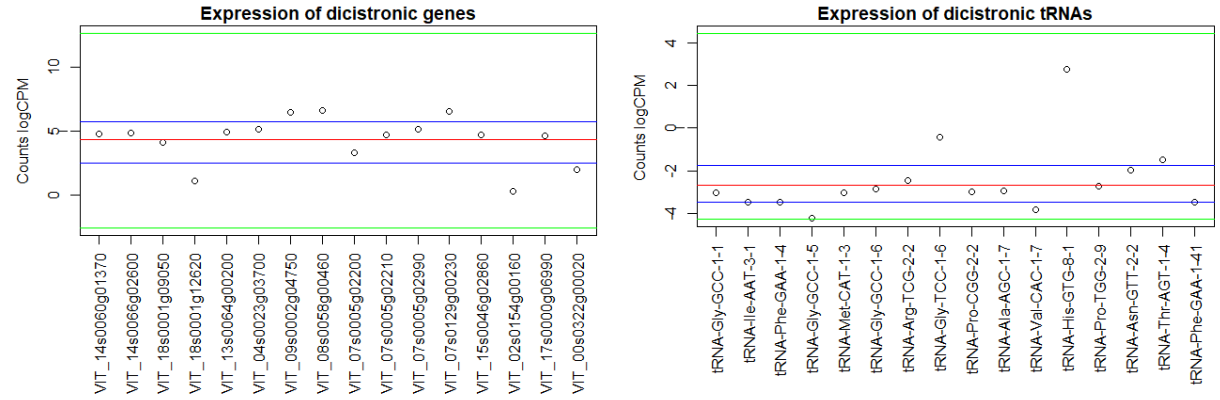
Supplemental Fig S2: Comparative analysis of expression levels of tRNA-mRNA intergenic regions (blue) and closest two introns (green) in non-dicistronic (left) and putative dicistronic tRNA-mRNA pairs (right) in leaf tissue. Average expression of intergenic region was significantly higher than the two closest introns in dicistronic tRNA-mRNA pairs (t-test, p-value < 0.05).

A B

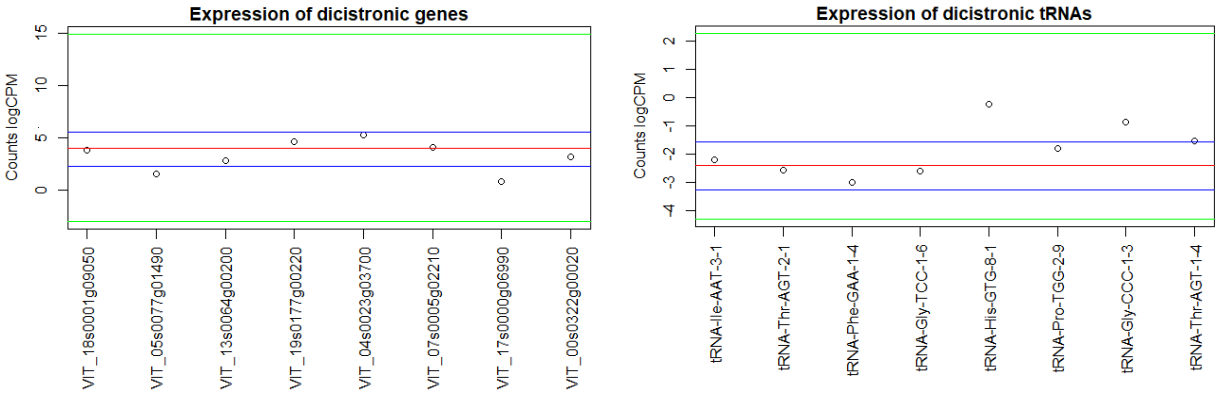


Supplemental Fig S3: Weblogo sequence analysis of the first 20 bp A) upstream; B) downstream of the 19 candidates dicistronic tRNAs.

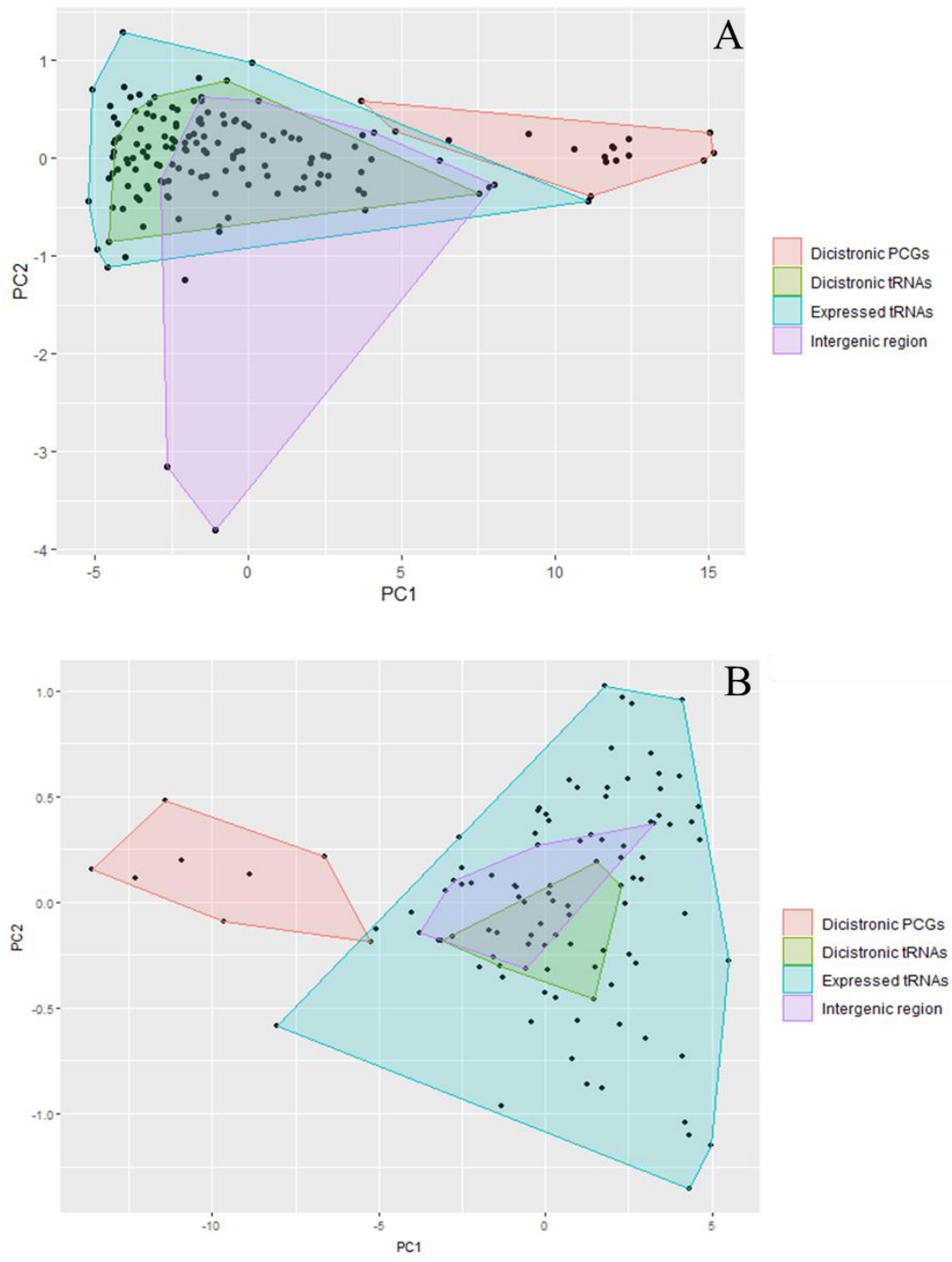
Leaf



Berry



Supplemental Fig S4: Expression of the candidate dicistronic genes and tRNA with respect of the total gene and tRNAs expressed from the RNA-seq data for both tissues. Expression of candidates dicistronic genes compared with the expression of 18,698 genes in leaf and 17,160 genes in berry samples. Expression of candidates dicistronic tRNAs were compared against the expression of total tRNA expressed in leaf (124) and berry (90) samples. Red line represents the median of the total expression. Blue lines correspond to the 25% and 75% quartile of the distribution of the total gene expression. Green lines are the minimum and maximum of gene expression.



Supplemental Fig S5: Local Fisher Discriminant Analysis of the expression (logCPM) of all expressed tRNAS, dicistronic tRNAs, intergenic region and dicistronic genes for each sub-region from the Barossa Wine growing region for A) leaf and B) berry samples. Clustering of the samples allows to clearly differentiating the expression of the PCGs from the tRNAs and intergenic regions.

Supplementary Table S1: Mapped RNA-seq data for leaf and berry samples to the *V. vinifera* to the reference genome PN40024.

Samples	Reads leaf samples	Reads berry samples	Merging, sorting and indexing of samples	Vineyard	Reads per vineyard leaf samples	Reads per vineyard berry samples
CG1.1	25,743,097	25,531,723				
CG1.2	27,225,555	24,411,207		CG1	79,352,399	73,630,662
CG1.3	26,383,747	23,687,732				
CG2.1	20,726,878	23,885,740				
CG2.2	37,326,602	26,059,204		CG2	78,384,092	75,369,795
CG2.3	20,330,612	25,424,851				
CG3.1	25,863,809	24,002,370				
CG3.2	20,856,760	20,203,746		CG3	66,286,754	62,900,663
CG3.3	19,566,185	18,694,547				
CG4.1	25,342,050	20,911,435				
CG4.2	22,221,312	22,605,405		CG4	69,661,221	69,523,235
CG4.3	22,097,859	26,006,395				
EV1.1	23,056,320	20,946,939				
EV1.2	25,524,736	23,149,751		EV1	66,187,032	68,002,341
EV1.3	17,605,976	23,905,651				
EV2.1	25,353,854	16,541,723				
EV2.2	13,559,924	16,291,835		EV2	60,076,277	54,009,007
EV2.3	21,162,499	21,175,449				
EV3.1	10,817,673	20,664,463				
EV3.2	21,138,475	22,523,299		EV3	57,632,497	65,759,052
EV3.3	25,676,349	22,571,290				
NG1.1	29,778,866	20,227,972				
NG1.2	24,798,843	18,273,130		NG1	108,814,365	61,059,571
NG1.3	54,236,656	22,558,469				
NG2.1	24,638,270	19,482,965				
NG2.2	29,024,880	16,554,581		NG2	72,024,100	57,528,418
NG2.3	18,360,950	21,490,872				
NG3.1	32,656,908	18,957,727				
NG3.2	15,254,356	18,703,259		NG3	69,753,767	60,136,835
NG3.3	21,842,503	22,475,849				
NG4.1	18,310,992	23,146,019		NG4	65,089,556	71,503,774

NG4.2	27,310,080	29,280,200			
NG4.3	19,468,484	19,077,555			
EE1.1	31,012,663	18,810,900			
EE1.2	30,189,945	16,909,798	EE1	88,188,361	56,037,489
EE1.3	26,985,753	20,316,791			
EE2.1	42,939,482	19,603,339			
EE2.2	29,521,962	20,177,138	EE2	97,607,231	61,555,159
EE2.3	25,145,787	21,774,682			
EE3.1	29,216,096	22,752,201			
EE3.2	18,263,688	21,445,703	EE3	69,208,157	64,480,260
EE3.3	21,728,373	20,282,356			
EE4.1	11,557,040	19,026,973			
EE4.2	13,484,751	16,911,470	EE4	33,915,023	54,361,862
EE4.3	8,873,232	18,423,419			
SG1.1	15,484,201	21,718,901			
SG1.2	17,223,812	19,702,658	SG1	49,870,401	59,074,343
SG1.3	17,162,388	17,652,784			
SG2.1	16,276,961	18,821,186			
SG2.2	18,392,528	22,273,698	SG2	50,753,611	62,118,686
SG2.3	16,084,122	21,023,802			
SG3.1	14,614,325	16,408,686			
SG3.2	17,965,039	20,178,538	SG3	56,385,714	57,281,924
SG3.3	23,806,350	20,694,700			
WR1.1	16,338,124	21,886,857			
WR1.2	20,531,631	21,246,529	WR1	79,394,381	67,822,176
WR1.3	42,524,626	24,688,790			
WR2.1	21,867,390	21,933,643			
WR2.2	21,709,202	27,810,725	WR2	61,070,617	72,039,790
WR2.3	17,494,025	22,295,422			
WR3.1	21,713,180	23,717,426			
WR3.2	18,808,425	19,326,702	WR3	64,182,265	67,154,196
WR3.3	23,660,660	24,110,068			
WR4.1	20,540,628	21,293,512			
WR4.2	44,712,982	23,626,566	WR4	83,079,538	69,561,686
WR4.3	17,825,928	24,641,608			

Supplementary table S2: Percentage of each tRNA isotype expressed in leaf and berry samples.

Isotype	Leaf %	Berry %
Pro	16%	18%
Ala	9%	9%
Arg	8%	8%
Gly	8%	6%
Gln	6%	4%
Asn	5%	4%
Thr	5%	7%
Met	5%	4%
Val	5%	6%
Leu	5%	2%
His	4%	3%
Ile	4%	4%
Cys	3%	4%
Phe	3%	3%
Tyr	3%	3%
Ser	3%	4%
Glu	3%	4%
Asp	2%	1%
Lys	1%	1%
Trp	0%	1%

Supplementary table S3.1: Sanger sequencing results of the RT-PCR product from the tRNA to the intergenic region of two dicistronic candidates ($tRNA^{ProTGG}$-Vv18s0001g09 and $tRNA^{ValCAC}$-Vv15s0046g02860).

Name	primer	Sequencing results
$tRNA^{ProTGG}$	Forward	>1_ProC_F_B09 TTTCTTTGGCTGGGTTTTGGTTTTACTTCACCATAAACCTCAAAAAAGCCCTCTTATGCT CTTCTGCAAATTTCATTTGTGTTATTGGTACTGAAACTCCGAGGCGGTGGCAGGCAAGGA AACAG
Intergenic_$tRNA^{ProTGG}$-Vv18s0001g09	Reverse	>2_ProC_R_B10 TGAATTTGCAGAGAGCATAAGAGGGCTTTTTTGAGGTTTATGGTGAAGTAAAACCAAAAC CCAGCGAAAGAAAATAAAAAGATTGGGGCATTCCGAGAATCGAACTCGGGACCTCTCGCA
$tRNA^{ValCAC}$	Forward	>5_ValC_F_C01 ATTGCCAC-AGTCTTCCATTTCTGTTGGGAGTCTCCCAGGGTCAGAGTATCAACGACACT CAGTGCCACAATCTCATTTCCATTTCTGCTAGGAGTCTCCCAGATTCTCAATATCAACAA CACTCACCCAGATTTTTAACATTTTCTCATCTGGATGTTCATCAATTAGTCAAACAATGC AGATTCAGCCACACCCACTTCCAAACTATAGTCTTAGGTCACCAATTTTCTCACCTTCGA CCCATTTTCTCATACCTTTCTCATCATCATTCCAACCTACTAAACTTGTCTCTATTCCAA ATTTCCCATCTGGGTCTTGCG
Intergenic_$tRNA^{ValCAC}$-Vv15s0046g02860	Reverse	>6_ValC_R_C02 ATGAGAAAATGGGTCGAAGGTGAGAAAATTGGTGACCTAAGACTATAGTTTGGAAGTGGG TGTGGCTGAATCTGCATTGTTTGACTAATTGATGAACATCCAGATGAGAAAATGTTAAAA ATCTGGGTGAGTGTTGTTGATATTGAGAATCTGGGAGACTCCTAGCAGAAATGGAAATGA GATTGTGGCACTGAGTGTCGTTGATACTCTGACCCTGGGAGACTCCCAACAGAAATGGAA GACTGTGGCAATAATAAAATAAAAATGCAAATGTCTGAGCCCGGGTTCGAACCGGGGACC TCTAGTG

Supplementary table S3.2: Sequenced PCR products were align to the query sequence in Blast2sequence with default settings.

Query ID	Sequence length	Sequenced PCR product	Primer	Max Score	Total Score	Query Cover	E value	Per. Ident
tRNA-Pro-TGG-2-9_Intergenic_region	172	$tRNA^{ProTGG}$	Forward	224	224	0.72	9E-64	99.2%
		Intergenic_$tRNA^{ProTGG}$-Vv18s0001g09	Reverse	213	213	0.69	6E-60	99.2%
tRNA-Val-CAC-1-7_Intergenic_region	376	$tRNA^{ValCAC}$	Forward	586	586	0.85	6E-172	99.7%
		Intergenic_$tRNA^{ValCAC}$-Vv15s0046g02860	Reverse	562	562	0.81	1E-164	99.7%

Supplementary Table S4: Gene ontology terms and annotated function of dicistronic genes identified in grapevine leaf, berry and common between both tissues. Function of the genes was determined by gene ontology motif that the genes presented (BLASTP) and from Cramer et al. (2020). Closest Arabidopsis match to the grapevine PCG candidates (based on protein sequence BLAST search) and evidence of dicistronic transcription and mobility.

Gene ID	RefSeq	Gene function	GO terms	Description of the GO term	BLAST top *Arabidopsis* match (Ensembl BLAST results)	Description *A. thaliana* (source Araport11)	Dicistronic in *Arabidopsis* (Zhang et al. 2016)	Common dicistronic tRNA in Arabidopsis and grapevine	Experimental evidence for mobility (PubMed ID)
VIT_09s0002g04750	XP_010654677.1	MAR-binding filament-like protein 1-1 isoform	GO:0006259 GO:0005622 GO:0006950	DNA metabolic process Intracellular Response to stress	AT1G05320 (19.4% AA ID, 0.0011 E-val)	PICC-LIKE, PICL myosin heavy chain, embryonic smooth protein	No	-	-
VIT_08s0058g00460	XP_002263502.1	Uncharacterized protein	No GO terms	No GO terms	AT1G02816 (27.8% AA ID, 2.7E-8 E-val)	pectinesterase	No	-	-
VIT_07s0129g00230	XP_002279597.1	Glucan endo-1,3-beta-glucosidase 1	GO:0004553 GO:0005975	hydrolase activity, hydrolyzing O-glycosyl compounds carbohydrate metabolic process	AT5G67460 (76.9% AA ID, 1.1E-48 E-val)	O-Glycosyl hydrolases family 17 protein	Yes	tRNA-Met(CAT)	-
VIT_14s0060g01370	XP_019080365.1	Zinc finger matrin-type protein 2 isoform	GO:0003676 GO:0008270 GO:0000398	nucleic acid binding zinc ion binding mRNA splicing, via spliceosome	AT3G05760 (87.9% AA ID, 5.0E-62 E-val)	C2H2 and C2HC zinc fingers superfamily protein	No	-	EDTA-exudation (184850610

Gene ID	RefSeq	Gene function	GO terms	Description of the GO term	BLAST top *Arabidopsis* match (Ensembl BLAST results)	Description *A. thaliana* (source Araport11)	Dicistronic in *Arabidopsis* (Zhang et al. 2016)	Common dicistronic tRNA in Arabidopsis and grapevine	Experimental evidence for mobility (PubMed ID)
VIT_18s0001g12620	XP_010664706.1	RING-type E3 ubiquitin transferase	GO:0004842	ubiquitin-protein transferase activity	AT1G29340 (45.9% AA ID, 3.5E-21 E-val)	ATPUB17 Encodes a protein containing a UND, a U-box, and an ARM domain	No	-	grafting (27247031)
VIT_02s0154g00160	XP_010662302.1	Flavin-containing monooxygenase FMO GS-OX-like 3	GO:0050660 GO:0004497 GO:0004499 GO:0050661	flavin adenine dinucleotide binding monooxygenase activity N,N-dimethylaniline monooxygenase activity NADP binding	AT1G62540 (60.6 % AA ID, 1.7E-39 E-val)	belongs to the flavin-monooxygenase (FMO) family, encodes a glucosinolate S-oxygenase that catalyzes the conversion of methylthioalkyl glucosinolates to methylsulfinylalkyl glucosinolates	No	-	grafting (27247031)
VIT_07s0005g02200	XP_002280035.1	Autophagy-related 8i-like	GO:0006914 GO:0006995	autophagy cellular response to nitrogen starvation	AT3G15580 (78.3% AA ID, 3.8E-59 E-val)	APG8H, Encodes APG8, a component of autophagy conjugation pathway. Delivered to the lumens of vacuole under nitrogen-starvation condition.	Yes	tRNA-Phe(GAA)	LMPC-exudation, EDTA-exudation (18485061)

Gene ID	RefSeq	Gene function	GO terms	Description of the GO term	BLAST top *Arabidopsis* match (Ensembl BLAST results)	Description *A. thaliana* (source Araport11)	Dicistronic in *Arabidopsis* (Zhang et al. 2016)	Common dicistronic tRNA in Arabidopsis and grapevine	Experimental evidence for mobility (PubMed ID)
VIT_15s0046g02860	XP_002273879.2	Riboflavin biosynthesis protein PYRD, chloroplastic	GO:0008835	diaminohydroxyphosphoribosylaminopyrimidine deaminase activity	AT3G47390 (36% AA ID, 2.3E-4 E-val)	PHOTOSENSITIVE 1 Encodes a protein that is believed to function as a pyrimidine reductase involved in riboflavin and FAD biosynthesis.	No	-	-
			GO:0008270	zinc ion binding					
			GO:0055114	oxidation-reduction process					
			GO:0009231	riboflavin biosynthetic process					
VIT_07s0005g02990	XP_002262985.1	Bet1-like SNARE 1-1	GO:0005484	SNAP receptor activity	AT3G58170 (75% AA ID, 4.7E-48 E-val)	ARABIDOPSIS THALIANA BET1P/SFT1P-LIKE PROTEIN 14A Encodes a Bet1/Sft1-like SNARE protein which fully suppresses the temperature-sensitive growth defect in sft1-1 yeast cells; however, it cannot support the deletion of the yeast BET1 gene	Yes	tRNA-Gly(CCC)	-
			GO:0015031	protein transport					

Gene ID	RefSeq	Gene function	GO terms	Description of the GO term	BLAST top *Arabidopsis* match (Ensembl BLAST results)	Description *A. thaliana* (source Araport11)	Dicistronic in *Arabidopsis* (Zhang et al. 2016)	Common dicistronic tRNA in Arabidopsis and grapevine	Experimental evidence for mobility (PubMed ID)
VIT_14s0066g02600	XP_019080906.1	Protein arginine methyltransferase NDUFAF7 homolog, mitochondrial isoform	GO:0035243 GO:0032981	protein-arginine omega-N symmetric methyltransferase activity mitochondrial respiratory chain complex I assembly	AT3G28700 (75% AA ID, 2.5E-142 E-val)	NADH dehydrogenase ubiquinone complex I, assembly factor-like protein (DUF185)	No	-	-
VIT_13s0064g00200	XP_002274669.1	Uncharacterized protein LOC100267159	No GO terms	No GO terms	AT4G19390 (59.2% AA ID, 3.3E-25 E-val)	Uncharacterized protein family (UPF0114)	No	-	EDTA-exudation (184850610)
VIT_17s0000g06990	XP_019081977.1	NAC transcription factor 32-like	GO:0003677 GO:0003735 GO:0006355 GO:0006412	DNA binding structural constituent of ribosome regulation of transcription, DNA-templated translation	AT1G33060 (48.9% AA ID, 4.6E-7 E-val)	ANAC014, NAC 014, NAC014	No	-	-
VIT_18s0001g09050	XP_010664507.1	GPN-loop GTPase 2	GO:0003924 GO:0005525	GTPase activity GTP binding	AT4G12790 (51.8% AA ID, 8.9E-26 E-val)	GPN3 GPN GTPase involved in selective nuclear import of RNA polymerase II.	No	-	EDTA-exudation (18485061)

Gene ID	RefSeq	Gene function	GO terms	Description of the GO term	BLAST top *Arabidopsis* match (Ensembl BLAST results)	Description *A. thaliana* (source Araport11)	Dicistronic in *Arabidopsis* (Zhang et al. 2016)	Common dicistronic tRNA in Arabidopsis and grapevine	Experimental evidence for mobility (PubMed ID)
VIT_00s0322g00020	XP_010646812.1	Heptahelical transmembrane protein 4	GO:0038023 GO:0009725	signalling receptor activity response to hormone	AT2G40710 (86% AA ID, 5.2E-55 E-val)	hemolysin-III related integral membrane protein	No	-	LMPC-exudation, EDTA-exudation (18485061)
VIT_04s0023g03700	XP_002263940.1	Uncharacterized protein LOC100267639	No GO terms	No GO terms	AT4G35980 (77.6 AA ID, 1.2E-38 E-val)	hypothetical protein	No	-	LMPC-exudation (18485061)
VIT_07s0005g02210	XP_010652305.1	Uncharacterized protein LOC100257191	GO:0005515	protein binding	AT4G01860 (59.2% AA ID, 4.5E-97 E-val)	Transducin family protein / WD-40 repeat family protein	Yes	tRNA-Phe(GAA)	-
VIT_05s0077g01490	XR_002030041.1	Uncharacterized protein LOC104879211	GO:0003676	nucleic acid binding	AT5G26940 (40% AA ID, 6.0E-5 E-val)	DEFECTIVE IN POLLEN ORGANELLE DNA DEGRADATION1, DPD1 Encodes a Mg2+-dependent DNA exonuclease that degrades organelle DNA during Arabidopsis pollen development	No	-	-

Gene ID	RefSeq	Gene function	GO terms	Description of the GO term	BLAST top *Arabidopsis* match (Ensembl BLAST results)	Description *A. thaliana* (source Araport11)	Dicistronic in *Arabidopsis* (Zhang et al. 2016)	Common dicistronic tRNA in Arabidopsis and grapevine	Experimental evidence for mobility (PubMed ID)
VIT_19s0177g00220	XP_002281637.1	Electron transfer flavoprotein subunit alpha, mitochondrial	GO:0005507 GO:0009055 GO:0050660 GO:0016491 GO:0033539	copper ion binding electron transfer activity flavin adenine dinucleotide binding oxidoreductase activity fatty acid beta-oxidation using acyl-CoA dehydrogenase	AT1G50940 (75.7 AA ID, 3.8E-160 E-val)	ELECTRON TRANSFER FLAVOPROTEIN ALPHA, ETFALPHA Encodes the electron transfer flavoprotein ETF alpha, a putative subunit of the mitochondrial electron transfer flavoprotein complex (ETF beta is At5g43430.1) in Arabidopsis. Mutations of the ETF beta gene results in accelerated senescence and early death compared to wild-type during extended darkness.	No	-	-

Gene ID	RefSeq	Gene function	GO terms	Description of the GO term	BLAST top *Arabidopsis* match (Ensembl BLAST results)	Description *A. thaliana* (source Araport11)	Dicistronic in *Arabidopsis* (Zhang et al. 2016)	Common dicistronic tRNA in Arabidopsis and grapevine	Experimental evidence for mobility (PubMed ID)
VIT_08s0007g03950	XP_010654034.1	Splicing factor U2af small subunit B	GO:0046872 GO:0030628 GO:0000398	metal ion binding pre-mRNA 3'-splice site binding mRNA splicing, via spliceosome	AT1G30480 (34.1% AA ID, 0.07 E-val)	DNA-DAMAGE-REPAIR/TOLERATION PROTEIN 111, DRT111, REQUIRED FOR SNC4-1D 2, RSN2, SFPS, SPLICING FACTOR FOR PHYTOCHROME SIGNALING recombination and DNA-damage resistance protein	No	-	-

Supplemental Table S5: Pearson correlation of the expression of tRNA and gene against intergenic region.

Leaf	Correlation analysis (Pearson) of the expression of:			
	Dicistronic tRNAs vs intergenic region		Dicistronic genes vs intergenic region	
region	correlation coefficient	p-value	correlation coefficient	p-value
CG	0.76	0.0006241	0.217	0.4187
EE	0.638	0.007714	0.275	0.302
EV	0.769	0.0004872	0.258	0.3337
NG	0.663	0.005119	0.49	0.05378
SG	0.765	0.0005456	0.317	0.2303
WR	0.779	0.0003724	0.243	0.364

Berry	Correlation analysis (Pearson) of the expression of:			
	Dicistronic tRNAs vs intergenic region		Dicistronic genes vs intergenic region	
region	correlation coefficient	p-value	correlation coefficient	p-value
CG	0.666	0.07122	-0.467	0.2432
EE	0.591	0.1224	-0.445	0.2686
EV	0.679	0.06355	-0.556	0.1518
NG	0.376	0.3577	-0.445	0.2682
SG	0.703	0.0515	-0.281	0.5
WR	0.69	0.05807	-0.286	0.4915

Supplemental Table S6: List of primers used for cDNA synthesis and RT-PCR

Name	primer	Sequence	Comments	Use for:		
				cDNA synthesis	RT-PCR	Sanger sequencing
$tRNA^{ProTGG}$	Forward	TGCGAGAGGTCCCGAGTTCGATT	PCR product is 172 bp			
Intergenic_$tRNA^{ProTGG}$-Vv18s0001g09	Reverse	CTGTTTCCTTGCCTGCCACC				
Vv18s0001g09	Reverse	TGCATCATTGGCAGGATCCA				
$tRNA^{ValCAC}$	Forward	CACTAGAGGTCCCCGGTTCGAA	PCR product is 376 bp			
Intergenic_$tRNA^{ValCAC}$-Vv15s0046g02860	Reverse	CCGCAAGACCCAGATGGGAA				
Vv15s0046g02860	Reverse	CCACCCCCTTTGAAGCCACA				
VvEF1-a_F	Forward	GAACTGGGTGCTTGATAGGC	PCR product is 150 bp			
VvEF1-a_R	Reverse	AACCAAAATATCCGGAGTAAAAGA				

References

Alabi OJ, Zheng Y, Jagadeeswaran G, Sunkar R, Naidu RA. 2012. High-throughput sequence analysis of small RNAs in grapevine (*Vitis vinifera* L.) affected by grapevine leafroll disease. *Mol Plant Pathol* **13**: 1060–1076.

Banerjee R, Chen S, Dare K, Gilreath M, Praetorius-Ibba M, Raina M, Reynolds NM, Rogers T, Roy H, Yadavalli SS, et al. 2010. tRNAs: cellular barcodes for amino acids. *FEBS Lett* **584**: 387–395.

Barkan A. 2011. Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. *Plant Physiol* **155**: 1520–1532.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Statist Soc* **57**: 239–300.

Bester R, Burger JT, Maree HJ. 2017. The small RNA repertoire in phloem tissue of three *Vitis vinifera* cultivars. *Plant Gene* **10**: 60–73.

Burgess AL, David R, Searle IR. 2015. Conservation of tRNA and rRNA 5-methylcytosine in the kingdom Plantae. *BMC Plant Biol* **15**: 199.

Calderwood A, Kopriva S, Morris RJ. 2016. Transcript Abundance Explains mRNA Mobility Data in Arabidopsis thaliana. *Plant Cell* **28**: 610–615.

Chan PP, Lowe TM. 2016. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res* **44**: D184–9.

Coombe BG. 1995. Growth Stages of the Grapevine: Adoption of a system for identifying grapevine growth stages. *Aust J Grape Wine Res* **1**: 104–110.

Cramer GR, Cochetel N, Ghan R, Destrac-Irvine A, Delrot S. 2020. A sense of place: transcriptomics identifies environmental signatures in Cabernet Sauvignon berry skins in the late stages of ripening. *BMC Plant Biol* **20**: 41.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res* **14**: 1188–1190.

Cui X, Liang Z, Shen L, Zhang Q, Bao S, Geng Y, Zhang B, Leo V, Vardy LA, Lu T, et al. 2017. 5-Methylcytosine RNA Methylation in Arabidopsis Thaliana. *Mol Plant* **10**: 1387–1399.

Dal Santo S, Tornielli GB, Zenoni S, Fasoli M, Farina L, Anesi A, Guzzo F, Delledonne M, Pezzotti M. 2013. The plasticity of the grapevine berry transcriptome. *Genome Biol* **14**: r54.

David R, Burgess A, Parker B, Li J, Pulsford K, Sibbritt T, Preiss T, Searle IR. 2017. Transcriptome-Wide Mapping of RNA 5-Methylcytosine in Arabidopsis mRNAs and Noncoding RNAs. *Plant Cell* **29**: 445–460.

Francklyn CS, Minajigi A. 2010. tRNA as an active chemical scaffold for diverse chemical transformations. *FEBS Lett* **584**: 366–375.

Guan D, Yan B, Thieme C, Hua J, Zhu H, Boheler KR, Zhao Z, Kragler F, Xia Y, Zhang S. 2017. PlaMoM: a comprehensive database compiles plant mobile macromolecules. *Nucleic Acids Res* **45**: D1021–D1028.

Han J, Fang J, Wang C, Yin Y, Sun X, Leng X, Song C. 2014. Grapevine microRNAs responsive to exogenous gibberellin. *BMC Genomics* **15**: 111.

Jackson DI, Lombard PB. 1993. Environmental and Management Practices Affecting Grape Composition and Wine Quality - A Review. *Am J Enol Vitic* **44**: 409-430Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, et al. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**: 463–467.

Karginov TA, Pastor DPH, Semler BL, Gomez CM. 2017. Mammalian Polycistronic mRNAs and Disease. *Trends Genet* **33**: 129–142.

Kersey PJ, Allen JE, Allot A, Barba M, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Grabmueller C, et al. 2018. Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Res* **46**: D802–D808.

Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**: 357–360.

Kolde R. 2015. pheatmap: Pretty Heatmaps. R package version 1.0.8.

Kruszka K, Barneche F, Guyot R, Ailhas J, Meneau I, Schiffer S, Marchfelder A, Echeverría M. 2003. Plant dicistronic tRNA-snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z. *EMBO J* **22**: 621–632.

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. 2013. Software for computing and annotating genomic ranges. *PLoS Comput Biol* **9**: e1003118.

Leader DJ, Clark GP, Watters J, Beven AF, Shaw PJ, Brown JW. 1997. Clusters of multiple different small nucleolar RNA genes in plants are expressed as and processed from polycistronic pre-snoRNAs. *EMBO J* **16**: 5742–5751.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930.

Liao Y, Smyth GK, Shi W. 2013. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* **41**: e108.

Liu G-T, Wang J-F, Cramer G, Dai Z-W, Duan W, Xu H-G, Wu B-H, Fan P-G, Wang L-J, Li S-H. 2012. Transcriptomic analysis of grape (*Vitis vinifera* L.) leaves during and after recovery from heat stress. *BMC Plant Biol* **12**: 174.

Merchan F, Boualem A, Crespi M, Frugier F. 2009. Plant polycistronic precursors containing non-homologous microRNAs target transcripts encoding functionally related proteins. *Genome Biol* **10**: R136.

Michaud M, Cognat V, Duchêne A-M, Maréchal-Drouard L. 2011. A global picture of tRNA genes in plant genomes. *Plant J* **66**: 80–93.

Pantaleo V, Szittya G, Moxon S, Miozzi L, Moulton V, Dalmay T, Burgyan J. 2010. Identification of grapevine microRNAs and their targets using high-throughput sequencing and degradome analysis. *Plant J* **62**: 960–976.

Pontin MA, Piccoli PN, Francisco R, Bottini R, Martinez-Zapater JM, Lijavetzky D. 2010. Transcriptome changes in grapevine (*Vitis vinifera* L.) cv. Malbec leaves induced by ultraviolet-B radiation. *BMC Plant Biol* **10**: 224.

Qu G, Kruszka K, Plewka P, Yang S-Y, Chiou T-J, Jarmolowski A, Szweykowska-Kulinska Z, Echeverria M, Karlowski WM. 2015. Promoter-based identification of novel non-coding RNAs reveals the presence of dicistronic snoRNA-miRNA genes in Arabidopsis thaliana. *BMC Genomics* **16**: 1009.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**: e47.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**: 24–26.

Robinson S, Sandercock N. 2014. An analysis of climate, soil and topographic information to aid the understanding of Barossa terroir – A Barossa Grounds Project, PIRSA Spatial Information Services, Government of South Australia.

Schramm L, Hernandez N. 2002. Recruitment of RNA polymerase III to its target promoters. *Genes Dev* **16**: 2593–2620.

Schubert M, Lindgreen S, Orlando L. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes* **9**: 88.

Sugita M, Sugiura M. 1996. Regulation of gene expression in chloroplasts of higher plants. *Plant Mol Biol* **32**: 315–326.

Sugiyama M. 2006. Local Fisher discriminant analysis for supervised dimensionality reduction. In *Proceedings of the 23rd international conference on Machine learning' ' - ICML '06*, pp. 905–912, ACM Press, New York, New York, USA.

Sun R, He F, Lan Y, Xing R, Liu R, Pan Q, Wang J, Duan C. 2015a. Transcriptome comparison of Cabernet Sauvignon grape berries from two regions with distinct climate. *J Plant Physiol* **178**: 43–54.

Sun X, Fan G, Su L, Wang W, Liang Z, Li S, Xin H. 2015b. Identification of cold-inducible microRNAs in grapevine. *Front Plant Sci* **6**: 595.

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. 2008. Synteny and collinearity in plant genomes. *Science* **320**: 486–488.

Thieme CJ, Rojas-Triana M, Stecyk E, Schudoma C, Zhang W, Yang L, Miñambres M, Walther D, Schulze WX, Paz-Ares J, et al. 2015. Endogenous Arabidopsis messenger RNAs transported to distant tissues. *Nat Plants* **1**: 15025.

Van Leeuwen C. 2009. Soils and Terroir Expression in Wines. In *Soil and Culture* (eds. E.R. Landa and C. Feller), pp. 453–465, Springer Netherlands, Dordrecht.

Wang C, Wang X, Kibet NK, Song C, Zhang C, Li X, Han J, Fang J. 2011. Deep sequencing of grapevine flower and berry short RNA library for discovery of novel microRNAs and validation of precise sequences of grapevine microRNAs deposited in miRBase. *Physiol Plant* **143**: 64–81.

Wickham H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. 2, illustrated. Springer-Verlag New York.

Williams LE, Araujo FJ. 2002. Correlations among Predawn Leaf, Midday Leaf, and Midday Stem Water Potential and their Correlations with other Measures of Soil and Plant Water Status in *Vitis vinifera*. *J Amer Soc Hort Sci* **127**: 448–454.

Xie H, Konate M, Sai N, Tesfamicael KG, Cavagnaro T, Gilliham M, Breen J, Metcalfe A, Stephen JR, De Bei R, et al. 2017. Global DNA Methylation Patterns Can Play a Role in Defining Terroir in Grapevine (*Vitis vinifera* cv. Shiraz). *Front Plant Sci* **8**: 1860.

Yang Y, Mao L, Jittayasothorn Y, Kang Y, Jiao C, Fei Z, Zhong G-Y. 2015. Messenger RNA exchange between scions and rootstocks in grafted grapevines. *BMC Plant Biol* **15**: 251.

Yang L, Perrera V, Saplaoura E, Apelt F, Bahin M, Kramdi A, Olas J, Mueller-Roeber B, Sokolowska E, Zhang W, et al. 2019. m5C Methylation Guides Systemic Transport of Messenger RNA over Graft Junctions in Plants. *Curr Biol* **29**: 2465–2476.e5.

Yukawa Y, Sugita M, Choisne N, Small I, Sugiura M. 2000. The TATA motif, the CAA motif and the poly(T) transcription termination motif are all important for transcription re-initiation on plant tRNA genes. *Plant J* **22**: 439–447.

Zhang W, Thieme CJ, Kollwig G, Apelt F, Yang L, Winter N, Andresen N, Walther D, Kragler F. 2016. tRNA-Related Sequences Trigger Systemic mRNA Transport in Plants. *Plant Cell* **28**: 1237–1249.

Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. *J Comput Biol* **7**: 203–214.

Zsófi Z, Tóth E, Rusjan D, Bálo B. 2011. Terroir aspects of grape quality in a cool climate wine region: Relationship between water deficit, vegetative growth and berry sugar concentration. *Sci Hortic (Amsterdam)* **127**: 494–499.

Chapter 5: General discussion

Using a multi-omics approach we generated new knowledge about DNA methylation and gene expression response to regionality in samples from twenty-two vineyards of the Barossa Wine Zone. We showed that the methylome of Shiraz leaf samples was differentiated between sub-regions of the Barossa Wine Zone. Gene expression also varied with vineyard regionality. These kinds of molecular differences could potentially underlie wines' *terroir*. When we analysed genetic variability inferred from the SNPs, we did not find a similar clustering of the samples. Although genetic differences between samples might contribute to the differences in DNA methylation and gene expression between vineyards, direct measurements of DNA methylation and gene expression were sufficient to explore regionality. We expected little genetic diversity in these samples as they were all from the same variety. DNA methylation diversity was high enough to group samples from the same sub-region together. This result suggests that by analyzing DNA methylation diversity we are able to detect epigenetic marks in the genome that are produced by (subtle) environmental differences and global methylation could reflect the aggregate elements of the plant's age and growing conditions.

A previous study of leaf samples from the same twenty-two vineyards analysed here from the Barossa Wine Zone, using global patterns of DNA methylation obtained through MSAPs, revealed that vineyards grouped by their area of provenance (Xie et al. 2017). We expanded these results by analyzing DNA methylation of grapevine leaves using msGBS. This technique relies on methylation–sensitive restriction enzymes digestion followed by sequencing. It has the advantage of generating a higher number of methylation markers than MSAPs and also generates the sequence information for each marker (Kitimu et al. 2015). We also expanded the previous report by analyzing the gene expression profile of our leaf samples and identified differentially expressed genes. Using a multi-omics approach we were able to identify genes that were both differentially expressed and methylated. These findings of the environmental impact on DNA

methylation and gene expression are novel in grapevine and should provide new opportunities to study the effect of *terroir*.

In our results, as in Xie et al. (2017), we saw that differences in grapevine DNA methylation associated with the provenance of the sample. By contrast, Dal Santo et al. (2018) did not find convincing evidence that methylation was affected by the environment, suggesting that DNA methylation in grapevine remained stable regardless of the growing conditions. In this previous study, researchers used reduced representation bisulfite sequencing (RRBS) to compare the DNA methylation from samples coming from three locations (Dal Santo et al. 2018). Although they obtained higher coverage per sample than in our results (15 million reads vs 2.4 million reads), Dal Santo et al. (2018) identified 4,696 CG sites, 4,737 CHG sites and 14,179 CHH sites among 48 samples, whereas we detected 450,501 methylated markers (CHG) in 198 samples. The higher number of samples analysed and markers identified have provided us a greater resolution to detect methylation changes driven by the environmental conditions.

Previous epigenetic analyses of *V. vinifera* have used whole genome bisulfite sequencing (WGBS) (Magris et al. 2019) and reduced representation bisulfite sequencing (RRBS) (Dal Santo et al. 2018). WGBS has the advantage of targeting the whole genome returning a single-base resolution and the highest sequence coverage of the genome. However, it is currently a high cost method (Sun et al. 2015). An alternative to WGBS is RRBS, where only a fraction of the genome is sequenced to a single-base resolution (Wreczycka et al. 2017). In RRBS, enrichment of CpG-rich regions is achieved by the isolation of short fragments after *MspI* digestion (*TaqI* in Dal Santo et al., 2018) followed by bisulfite conversion and library preparation as performed for WGBS (Kurdyukov and Bullock 2016). A more cost-effective alternative is msGBS, also known as methylation-sensitive restriction enzyme sequencing (MRE-seq), where, in the same way as RRBS, restriction enzymes are used to reduce the amount of DNA to be sequenced (Li et al.

2015). In msGBS there is no bisulfite treatment of the libraries, the method relying only on the recognition site of the restriction enzymes for the identification of methylation polymorphisms. By combining results from msGBS (MRE-seq) and methylation dependent immunoprecipitation followed by sequencing (MeDIP-seq), Stevens et al. (2013) were able to produce DNA methylomes of all 28 million CpGs in the human genome with equivalent quantification, coverage and resolution for a fraction (<10%) of the cost of whole-genome bisulfite sequencing methods. The previous studies in grapevine used three libraries for WGBS (Magris et al. 2019) and 48 samples in RRBS (Dal Santo et al. 2018). Using the msGBS method we were able to analyse the epigenetic status of 198 samples. With a larger number of samples, we were able to better dissect the environmental effect on DNA methylation and gene expression of the samples. However, msGBS delivers lower sequence coverage of the genome than WGBS or RRBS, influencing the ability to analyse the association between DNA methylation and gene expression. In a large population analysis, msGBS can be very useful to epigenotype, but it would require increasing the sequencing coverage of the whole genome to fully assess the DNA methylation effect on gene expression. Alternative methods would also allow us to investigate the methylation status in the other two methylation sequence contexts in plants (CG and CHH). From all DMMs identified here, only 10% were associated to a gene, suggesting that the majority of DMMs must be in untranscribed regions of the genome. Studies using genome-wide DNA methylation have found that gene body methylation is common in plants as distant as ferns to angiosperms, but mainly in the CG context (Takuno et al. 2016; Zemach et al. 2010), while CHG and CHH methylation are usually dedicated to transposon silencing (Kato et al. 2003). It remains to be assessed at which sequence context methylation is more stable in grapevine and which is more influenced by the environment, soil properties and plant age.

As grape berries are the organ of interest for the wine industry, it would be crucial to repeat the analyses in berries. It is important to assess to what extent the results from one tissue can be applied to another. For example, does DNA methylation of berry reflect its origin in the same way as in leaves? Preliminary results from analysis of gene expression of berries at veraison from the twenty-two vineyards (Appendices: Supporting Figure 1) support what we previously found in leaves, as samples clustered by the sub-regions of the Barossa Wine Zone. In addition, vintage effect, or season effect, is the strongest factor modifying grape berry composition (Anesi et al. 2015). The *terroi*r effect should present molecular elements that are constant between different growing seasons. To better separate the *terroir* effect from the seasonal effect, it would be necessary to repeat these experiments with samples from several years.

RNA molecules are emerging as an important intercellular signalling molecule with regulatory functions in distant tissues (Kehr and Kragler 2018). We analysed gene expression data to identify molecular signatures of long distance RNA movement. tRNA-related sequences have been shown to mobilize mRNA through the vascular system (Zhang et al. 2016c), and could function as signaling molecules between tissues and organs to coordinate plant development, as well as respond to environmental cues. In this work, we were able to identify nineteen dicistronic tRNA-mRNA transcripts in leaf and berry samples. Our data demonstrated for the first time that the identification of such transcripts in field grown plants and that the expression was sub-region dependent and tissue specific.

It still needs to be assessed if these transcripts are able to mobilize through the vascular system in grapevine. As grafting is a common practice in viticulture, we could potentially detect the movement of the dicistronic tRNA-mRNA transcripts between a scion and rootstock of different varieties. A similar strategy was employed to detect movement of other small RNA classes in *A. thaliana* (Molnar et al. 2010). It remains for future research first to investigate if dicistronic

mRNAs are also transcribing in the destination tissue; second, to what extent the dicistronic transcripts are mechanisms of communication between different tissues in grapevine; and finally to study what role the environment could play in dicistronic transcription.

As grapevines are usually vegetatively propagated, further studies are needed to determine if the epigenetic status of new plants is similar to the maternal plants. Are there stable epigenetic markers that get fixed in the genome, regardless of the growing conditions? Can epigenetic memory accumulate to modify berry metabolite composition? Evidence in stressed *in vitro* cultivated grapevines revealed that after the stress was terminated 40% of the acquired methylation gradually reverted to a similar state found in the maternal plants. However, the remaining 60% of the observed DNA methylation diversity was still maintained one year after the plant ceased to be exposed to the *in vitro* conditions (Baránek et al. 2015). If epigenetic changes in grapevine could modify berry metabolite composition, epigenetic memory might play a role in *terroir* expression. The stability of epigenetic memory in grapevines still needs to be investigated.

Studies such as the ones described in this thesis help us to better understand the molecular mechanisms involved in a plant's response to the growing conditions. This understanding can potentially inform and improve management of the plants and, thereby, wine quality and vineyard sustainability in the grape and wine industry. For example, modelling the impacts of climate change on grapevine phenology estimates that harvest will occur earlier in the year in Australia due to warmer temperatures (Webb et al. 2007). A shorter, warmer growing season has implications not only for berry composition, but also for the logistics of harvest and wine making. If we better understand the epigenetic mechanisms underlying grapevine's response to the environment and its effects on gene expression, we can potentially induce epigenetic variation for desirable characteristics without changes in the sequence of the genome. We could

deliberately manipulate and select the epigenetic states of grapevine to better withstand adverse environmental conditions, and at the same time maintain the genotype of the cultivar, which is important in traditional wine regions.

In this work we found that grapevine DNA methylation reflected the regional growth conditions, responding to changes in annual rainfall and with the plant age. We also detected the presence of dicistronic tRNA-mRNAs, which might act as signaling molecules between distant tissues. These types of responses are especially relevant in long-lived woody plants, potentially giving them the plasticity they need to adapt to local conditions, providing, at the same time, a unique type of wine according to their *terroir*.

Appendices

Chapter 4: Supplemental_file1: Alignment of Sanger sequencing results from RT-PCR product for two dicistronic tRNA-mRNA candidates to the expected PCR product.

```
Query: tRNA-Pro-TGG-2-9_Intergenic_region Query ID: lcl|Query_54328 Length:
171

>1_ProC_F_B09

Sequence ID: Query_54330 Length: 125

Range 1: 1 to 124

Score:224 bits(121), Expect:9e-64,

Identities:123/124(99%),  Gaps:0/124(0%), Strand: Plus/Plus


Query  48   TTTCTTTCGCTGGGTTTTGGTTTTACTTCACCATAAACCTCAAAAAAGCCCTCTTATGCT  107
            ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    TTTCTTTGGCTGGGTTTTGGTTTTACTTCACCATAAACCTCAAAAAAGCCCTCTTATGCT  60


Query  108  CTTCTGCAAATTTCATTTGTGTTATTGGTACTGAAACTCCGAGGCGGTGGCAGGCAAGGA  167
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   CTTCTGCAAATTTCATTTGTGTTATTGGTACTGAAACTCCGAGGCGGTGGCAGGCAAGGA  120


Query  168  AACA  171
            ||||
Sbjct  121  AACA  124
```

Query: tRNA-Pro-TGG-2-9_Intergenic_region Query ID: lcl|Query_54328 Length: 171

>2_ProC_R_B10

Sequence ID: Query_54331 Length: 120

Range 1: 3 to 120

Score:213 bits(115), Expect:2e-60,

Identities:118/119(99%), Gaps:1/119(0%), Strand: Plus/Minus

```
Query  1    TGCGAGAGGTCCCGAGTTCGATTCTCGGAATGCCCCAATCTTTTTATTTTCTTTCGCTGG  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  120  TGCGAGAGGTCCCGAGTTCGATTCTCGGAATGCCCCAATCTTTTTATTTTCTTTCGCTGG  61

Query  61   GTTTTGGTTTTACTTCACCATAAACCTCAAAAAAGCCCTCTTATGCTCTTCTGCAAATT  119
            ||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Sbjct  60   GTTTTGGTTTTACTTCACCATAAACCTCAAAAAAGCCCTCTTATGCTCT-CTGCAAATT  3
```

Query: Val_intergenic_PCR_Product Query ID: lcl|Query_16750 Length: 376

>5_ValC_F_C01

Sequence ID: Query_16752 Length: 320

Range 1: 1 to 320

Score:586 bits(317), Expect:6e-172,

Identities:319/320(99%), Gaps:0/320(0%), Strand: Plus/Plus

```
Query  56   ATTGCCAGAGTCTTCCATTTCTGTTGGGAGTCTCCCAGGGTCAGAGTATCAACGACACTC  115
            ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATTGCCACAGTCTTCCATTTCTGTTGGGAGTCTCCCAGGGTCAGAGTATCAACGACACTC  60

Query  116  AGTGCCACAATCTCATTTCCATTTCTGCTAGGAGTCTCCCAGATTCTCAATATCAACAAC  175
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   AGTGCCACAATCTCATTTCCATTTCTGCTAGGAGTCTCCCAGATTCTCAATATCAACAAC  120

Query  176  ACTCACCCAGATTTTTAACATTTTCTCATCTGGATGTTCATCAATTAGTCAAACAATGCA  235
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ACTCACCCAGATTTTTAACATTTTCTCATCTGGATGTTCATCAATTAGTCAAACAATGCA  180

Query  236  GATTCAGCCACACCCACTTCCAAACTATAGTCTTAGGTCACCAATTTTCTCACCTTCGAC  295
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GATTCAGCCACACCCACTTCCAAACTATAGTCTTAGGTCACCAATTTTCTCACCTTCGAC  240

Query  296  CCATTTTCTCATACCTTTCTCATCATCATTCCAACCTACTAAACTTGTCTCTATTCCAAA  355
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  CCATTTTCTCATACCTTTCTCATCATCATTCCAACCTACTAAACTTGTCTCTATTCCAAA  300

Query  356  TTTCCCATCTGGGTCTTGCG  375
            ||||||||||||||||||||
Sbjct  301  TTTCCCATCTGGGTCTTGCG  320
```

Query: Val_intergenic_PCR_Product Query ID: lcl|Query_16750 Length: 376

>6_ValC_R_C02

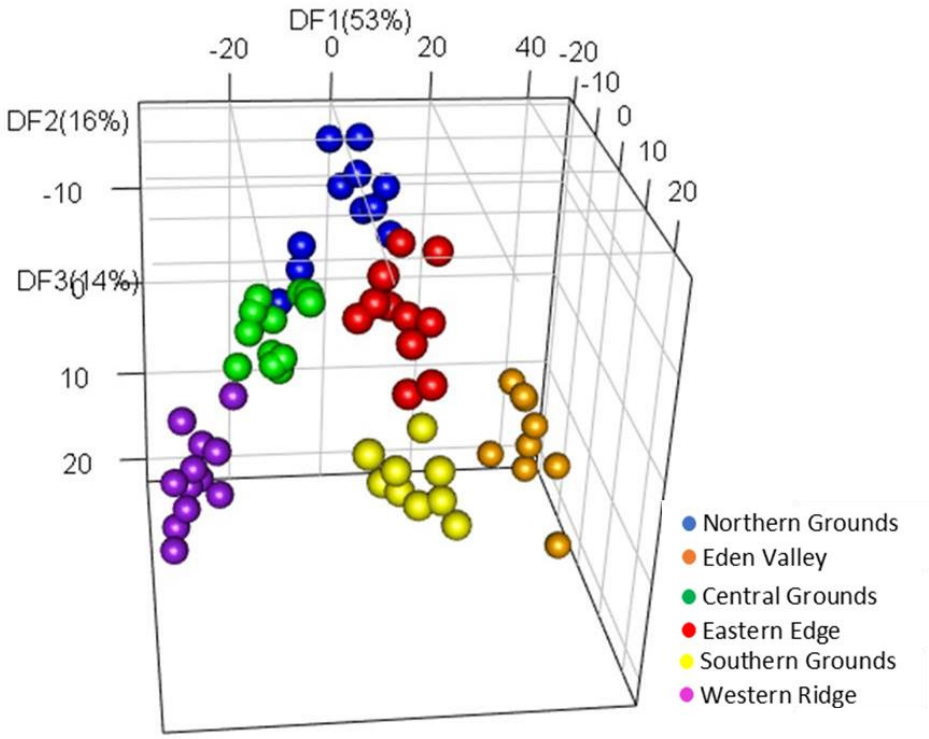
Sequence ID: Query_16753 Length: 307

Range 1: 1 to 307

Score:562 bits(304), Expect:1e-164,

Identities:306/307(99%), Gaps:0/307(0%), Strand: Plus/Minus

```
Query  1    CACTAGAGGTCCCCGGTTCGAACCCGGGCTCAGACATTTGCATTTTTATTTTATTATTGC  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  307  CACTAGAGGTCCCCGGTTCGAACCCGGGCTCAGACATTTGCATTTTTATTTTATTATTGC  248

Query  61   CAGAGTCTTCCATTTCTGTTGGGAGTCTCCCAGGGTCAGAGTATCAACGACACTCAGTGC  120
            || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  247  CACAGTCTTCCATTTCTGTTGGGAGTCTCCCAGGGTCAGAGTATCAACGACACTCAGTGC  188

Query  121  CACAATCTCATTTCCATTTCTGCTAGGAGTCTCCCAGATTCTCAATATCAACAACACTCA  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  187  CACAATCTCATTTCCATTTCTGCTAGGAGTCTCCCAGATTCTCAATATCAACAACACTCA  128

Query  181  CCCAGATTTTTAACATTTTCTCATCTGGATGTTCATCAATTAGTCAAACAATGCAGATTC  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  127  CCCAGATTTTTAACATTTTCTCATCTGGATGTTCATCAATTAGTCAAACAATGCAGATTC  68

Query  241  AGCCACACCCACTTCCAAACTATAGTCTTAGGTCACCAATTTTCTCACCTTCGACCCATT  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  67   AGCCACACCCACTTCCAAACTATAGTCTTAGGTCACCAATTTTCTCACCTTCGACCCATT  8

Query  301  TTCTCAT  307
            |||||||
Sbjct  7    TTCTCAT  1
```



Chapter 5: Supporting figure 1. Visualization of gene expression diversity between berry samples from six sub-regions of the Barossa Wine Zone. Three-dimensional plot of PC-LDA of the gene expression profile of 18,952 genes of 66 samples on the three axis between-group variation (discriminant function, DF).

References

(Excluding Chapters 2, 3 and 4)

Agudelo-Romero P, Erban A, Rego C, Carbonell-Bejerano P, Nascimento T, Sousa L, Martínez-Zapater JM, Kopka J, Fortes AM. 2015. Transcriptome and metabolome reprogramming in *Vitis vinifera* cv. Trincadeira berries upon infection with *Botrytis cinerea*. *J Exp Bot* **66**: 1769–1785.

Amoah S, Kurup S, Rodriguez Lopez CM, Welham SJ, Powers SJ, Hopkins CJ, Wilkinson MJ, King GJ. 2012. A hypomethylated population of Brassica rapa for forward and reverse epi-genetics. *BMC Plant Biol* **12**: 193.

Anesi A, Stocchero M, Dal Santo S, Commisso M, Zenoni S, Ceoldo S, Tornielli GB, Siebert TE, Herderich M, Pezzotti M, et al. 2015. Towards a scientific interpretation of the *terroir* concept: plasticity of the grape berry metabolome. *BMC Plant Biol* **15**: 191.

Antolín MC, Ayari M, Sánchez-Díaz M. 2006. Effects of partial rootzone drying on yield, ripening and berry ABA in potted Tempranillo grapevines with split roots. *Aust J Grape Wine Res* **12**: 13–20.

Aziz MH, Kumar R, Ahmad N. 2003. Cancer chemoprevention by resveratrol: in vitro and in vivo studies and the underlying mechanisms (review). *Int J Oncol* **23**: 17–28.

Azuma A, Yakushiji H, Koshita Y, Kobayashi S. 2012. Flavonoid biosynthesis-related genes in grape skin are differentially regulated by temperature and light conditions. *Planta* **236**: 1067–1080.

Baali-Cherif D, Besnard G. 2005. High genetic diversity and clonal growth in relict populations of *Olea europaea* subsp. *laperrinei* (Oleaceae) from Hoggar, Algeria. *Ann Bot* **96**: 823–830.

Banerjee AK, Chatterjee M, Yu Y, Suh S-G, Miller WA, Hannapel DJ. 2006. Dynamics of a mobile RNA of potato involved in a long-distance signaling pathway. *Plant Cell* **18**: 3443–3457.

Baránek M, Čechová J, Raddová J, Holleinová V, Ondrušíková E, Pidra M. 2015. Dynamics and reversibility of the DNA methylation landscape of grapevine plants (*Vitis vinifera*) stressed by *in vitro* cultivation and thermotherapy. *PLoS One* **10**: e0126638.

Barends S, Rudinger-Thirion J, Florentz C, Giegé R, Pleij CWA, Kraal B. 2004. tRNA-like structure regulates translation of Brome mosaic virus RNA. *J Virol* **78**: 4003–4010.

Barossa Grape and Wine Association. Barossa Chapters: Barossa Grounds. https://www.barossawine.com/barossa/barossa-chapters/barossa-chapters-barossa-grounds/ (Accessed November 14, 2019a).

Barossa Grape and Wine Association. Barossa Chapters: Shiraz. https://www.barossawine.com/barossa/barossa-chapters/barossa-chapters-shiraz/ (Accessed November 14, 2019b).

Barossa Grape and Wine Association. Barossa Valley - Wine has been a way of life in Barossa since 1842. https://www.barossawine.com/barossa/ (Accessed November 18, 2019c).

Baulcombe DC, Dean C. 2014. Epigenetic regulation in plant responses to the environment. *Cold Spring Harb Perspect Biol* **6**: a019471.

Bavaresco L, Giachino E, Pezzutto S. 2003. Grapevine rootstock effects on lime-induced chlorosis, nutrient uptake, and source–sink relationships. *J Plant Nutr* **26**: 1451–1465.

Berdeja M, Nicolas P, Kappel C, Dai ZW, Hilbert G, Peccoux A, Lafontaine M, Ollat N, Gomès E, Delrot S. 2015. Water limitation and rootstock genotype interact to alter grape berry metabolism through transcriptome reprogramming. *Hortic Res* **2**: 15012.

Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A. 2009. An operational definition of epigenetics. *Genes Dev* **23**: 781–783.

Bergqvist J, Dokoozlian N, 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. *Am J Enol Vitic* **52**: 1–7.

Bester R, Burger JT, Maree HJ. 2017. The small RNA repertoire in phloem tissue of three *Vitis vinifera* cultivars. *Plant Gene* **10**: 60–73.

Bräutigam K, Vining KJ, Lafon-Placette C, Fossdal CG, Mirouze M, Marcos JG, Fluch S, Fraga MF, Guevara MÁ, Abarca D, et al. 2013. Epigenetic regulation of adaptive responses of forest tree species to the environment. *Ecol Evol* **3**: 399–415.

Burn JE, Bagnall DJ, Metzger JD, Dennis ES, Peacock WJ. 1993. DNA methylation, vernalization, and the initiation of flowering. *Proc Natl Acad Sci USA* **90**: 287–291.

Camps C, Kappel C, Lecomte P, Léon C, Gomès E, Coutos-Thévenot P, Delrot S. 2010. A transcriptomic study of grapevine (*Vitis vinifera* cv. Cabernet-Sauvignon) interaction with the vascular ascomycete fungus *Eutypa lata*. *J Exp Bot* **61**: 1719–1737.

Carbonell-Bejerano P, Diago M-P, Martínez-Abaigar J, Martínez-Zapater JM, Tardáguila J, Núñez-Olivera E. 2014. Solar ultraviolet radiation is necessary to enhance grapevine fruit ripening transcriptional and phenolic responses. *BMC Plant Biol* **14**: 183.

Carbonell-Bejerano P, Santa María E, Torres-Pérez R, Royo C, Lijavetzky D, Bravo G, Aguirreolea J, Sánchez-Díaz M, Antolín MC, Martínez-Zapater JM. 2013. Thermotolerance responses in ripening berries of *Vitis vinifera* L. cv Muscat Hamburg. *Plant Cell Physiol* **54**: 1200–1216.

Carbonneau A. 1985. The Early Selection of Grapevine Rootstocks for Resistance to Drought Conditions. *Am J Enol Vitic* **36**: 195–198.

Chone X, Van Leeuwen C, Chery P, Ribereau-Gayon P. 2001. *Terroir* influence on water status and nitrogen status of non-irrigated' ' Cabernet Sauvignon (*Vitis vinifera*). Vegetative development, must and wine composition (Example of a Medoc top estate vineyard, Saint Julien area, Bordeaux, 1997). *S Afr J Enol Vitic* **22**: 8–26.

Choudhuri S. 2011. From Waddington's epigenetic landscape to small noncoding RNA: some important milestones in the history of epigenetics research. *Toxicol Mech Methods* **21**: 252–274.

Cohen SD, Tarara JM, Kennedy JA. 2008. Assessing the impact of temperature on grape phenolic metabolism. *Anal Chim Acta* **621**: 57–67.

Cohen SD, Tarara JM, Kennedy JA. 2012. Diurnal temperature range compression hastens berry development and modifies flavonoid partitioning in grapes. *Am J Enol Vitic* **63**: 112–120.

Conde C, Silva P, Fontes N, Dias ACP, Tavares RM, Sousa MJ, Agasse A, Delrot S, Geros H, Conde C, et al. 2007. Biochemical changes throughout grape berry development and fruit and wine quality. *Food* **1**: 1–22.

Cookson SJ, Ollat N. 2013. Grafting with rootstocks induces extensive transcriptional re-programming in the shoot apical meristem of grapevine. *BMC Plant Biol* **13**: 147.

Coombe BG. 1992. Research on development and ripening of the grape berry. *Am J Enol Vitic* **43**: 101–110.

Costantini L, Battilana J, Lamaj F, Fanizza G, Grando MS. 2008. Berry and phenology-related traits in grapevine (*Vitis vinifera* L.): from quantitative trait loci to underlying genes. *BMC Plant Biol* **8**: 38.

Cramer GR, Cochetel N, Ghan R, Destrac-Irvine A, Delrot S. 2019. A sense of place: transcriptomics identifies environmental signatures in cabernet sauvignon berry skins in the late stages of ripening. *BioRxiv*.

Cramer GR, Ghan R, Schlauch KA, Tillett RL, Heymann H, Ferrarini A, Delledonne M, Zenoni S, Fasoli M, Pezzotti M. 2014. Transcriptomic analysis of the late stages of grapevine (*Vitis vinifera* cv. Cabernet Sauvignon) berry ripening reveals significant induction of ethylene signaling and flavor pathways in the skin. *BMC Plant Biol* **14**: 370.

Dal Santo S, Fasoli M, Negri S, D'Incà E, Vicenzi N, Guzzo F, Tornielli GB, Pezzotti M, Zenoni S. 2016a. Plasticity of the berry ripening program in a white grape variety. *Front Plant Sci* **7**: 970.

Dal Santo S, Palliotti A, Zenoni S, Tornielli GB, Fasoli M, Paci P, Tombesi S, Frioni T, Silvestroni O, Bellincontro A, et al. 2016b. Distinct transcriptome responses to water limitation in isohydric and anisohydric grapevine cultivars. *BMC Genomics* **17**: 815.

Dal Santo S, Tornielli GB, Zenoni S, Fasoli M, Farina L, Anesi A, Guzzo F, Delledonne M, Pezzotti M. 2013. The plasticity of the grapevine berry transcriptome. *Genome Biol* **14**: r54.

Dal Santo S, Zenoni S, Sandri M, De Lorenzis G, Magris G, De Paoli E, Di Gaspero G, Del Fabbro C, Morgante M, Brancadoro L, et al. 2018. Grapevine field experiments reveal the contribution of genotype, the influence of environment and the effect of their interaction (G×E) on the berry transcriptome. *Plant J* **93**: 1143–1159.

Degu A, Hochberg U, Sikron N, Venturini L, Buson G, Ghan R, Plaschkes I, Batushansky A, Chalifa-Caspi V, Mattivi F, et al. 2014. Metabolite and transcript profiling of berry skin during fruit development elucidates differential regulation between Cabernet Sauvignon and Shiraz cultivars at branching points in the polyphenol pathway. *BMC Plant Biol* **14**: 188.

Deluc LG, Grimplet J, Wheatley MD, Tillett RL, Quilici DR, Osborne C, Schooley DA, Schlauch KA, Cushman JC, Cramer GR. 2007. Transcriptomic and metabolite analyses of Cabernet Sauvignon grape berry development. *BMC Genomics* **8**: 429.

Des Gachons CP, Leeuwen CV, Tominaga T, Soyer J-P, Gaudillere J-P, Dubourdieu D. 2005. Influence of water and nitrogen deficit on fruit ripening and aroma potential of *Vitis vinifera* L cv Sauvignon blanc in field conditions. *J Sci Food Agric* **85**: 73–85.

Dry PR, Loveys BR. 1998. Factors influencing grapevine vigour and the potential for control with partial rootzone drying. *Aust J Grape Wine Res* **4**: 140–148.

Du Plessis K, Young PR, Eyéghé-Bickong HA, Vivier MA. 2017. The transcriptional responses and metabolic consequences of acclimation to elevated light exposure in grapevine berries. *Front Plant Sci* **8**: 1261.

Duchêne E, Huard F, Dumas V, Schneider C, Merdinoglu D. 2010. The challenge of adapting grapevine varieties to climate change. *Clim Res* **41**: 193–204.

Eichten SR, Schmitz RJ, Springer NM. 2014. Epigenetics: beyond chromatin modifications and complex genetic regulation. *Plant Physiol* **165**: 933–947.

Esmenjaud D, Bouquet A. 2009. Selection and application of resistant germplasm for grapevine nematodes management. In *Integrated Management of Fruit Crops Nematodes* (eds. A. Ciancio and K.G. Mukerji), pp. 195–214, Springer, Dordrecht,' ' Netherlands.

Fasoli M, Richter CL, Zenoni S, Bertini E, Vitulo N, Dal Santo S, Dokoozlian N, Pezzotti M, Tornielli GB. 2018. Timing and order of the molecular events marking the onset of berry ripening in grapevine. *Plant Physiol* **178**: 1187–1206.

Feil R, Fraga MF. 2012. Epigenetics and the environment: emerging patterns and implications. *Nat Rev Genet* **13**: 97–109.

Finnegan EJ, Genger RK, Kovac K, Peacock WJ, Dennis ES. 1998. DNA methylation and the promotion of flowering by vernalization. *Proc Natl Acad Sci USA* **95**: 5824–5829.

Fisarakis I, Nikolaou N, Tsikalas P, Therios I, Stavrakas D. 2005. Effect of salinity and rootstock on concentration of potassium, calcium, magnesium, phosphorus, and nitrate–nitrogen in thompson seedless grapevine. *J Plant Nutr* **27**: 2117–2134.

Fonseca Lira-Medeiros C, Parisod C, Fernandes RA, Mata CS, Cardoso MA, Ferreira PC. 2010. Epigenetic variation in mangrove plants occurring in contrasting natural environment. *PLoS One* **5**: e10326.

Food Agriculture Organization. 2012. FAOSTAT. http://faostat.fao.org/site/339/default.aspx (Accessed August 4, 2016).

Fortes AM, Gallusci P. 2017. Plant stress responses and phenotypic plasticity in the Epigenomics era: perspectives on the grapevine scenario, a model for perennial crop plants. *Front Plant Sci* **8**: 82.

Gallusci P, Hodgman C, Teyssier E, Seymour GB. 2016. DNA methylation and chromatin regulation during fleshy fruit development and ripening. *Front Plant Sci* **7**: 807.

Giovannoni J, Nguyen C, Ampofo B, Zhong S, Fei Z. 2017. The epigenome and transcriptional dynamics of fruit ripening. *Annu Rev Plant Biol* **68**: 61–84.

Gladstones J. 2011a. Introduction and definition of terroir. In *Wine, Terroir & Climate Change*, pp. 1–4, Wakefield Press, Kent Town, Australia.

Gladstones J. 2011b. Vine balances and management. In *Wine, Terroir and Climate Change*, pp. 60–67, Wakefield Press, Kent Town, Australia.

Glissant D, Dédaldéchamp F, Delrot S. 2008. Transcriptomic analysis of grape berry softening during ripening. *OENO One* **42**: 1.

Granett J, Walker MA, Kocsis L, Omer AD. 2001. Biology and management of grape phylloxera. *Annu Rev Entomol* **46**: 387–412.

Greer DH, Weston C. 2010. Heat stress affects flowering, berry growth, sugar accumulation and photosynthesis of *Vitis vinifera* cv. Semillon grapevines grown in a controlled environment. *Functional Plant Biol* **37**: 206.

Guilpart N, Metay A, Gary C. 2014. Grapevine bud fertility and number of berries per bunch are determined by water and nitrogen stress around flowering in the previous year. *European Journal of Agronomy* **54**: 9–20.

Hepner Y, Bravdo B. 1985. Effect of crop level and drip irrigation scheduling on the potassium status of Cabernet Sauvignon and Carignane vines and its influence on must and wine composition and quality. *Am J Enol Vitic* **36**: 140–147.

Herrera CM, Bazaga P. 2013. Epigenetic correlates of plant phenotypic plasticity: DNA methylation differs between prickly and nonprickly leaves in heterophyllous *Ilex aquifolium* (Aquifoliaceae) trees. *Bot J Linn Soc* **171**: 441–452.

Iacono F, Sommer KJ. 1996. Photoinhibition of photosynthesis and photorespiration in *Vitis vinifera* under field conditions — effects of light climate and leaf position. *Aust J Grape Wine Res* **2**: 1–11.

Imazio S, Labra M, Grassi F, Winfield M, Bardini M, Scienza A. 2002. Molecular tools for clone identification: the case of the grapevine cultivar Traminer. *Plant Breeding* **121**: 531–535.

Jackson DI, Lombard PB. 1993. Environmental and Management Practices Affecting Grape Composition and Wine Quality - A Review. *Am J Enol Vitic* **44**: 409–460.

Jones GV, Davis RE. 2000. Climate Influences on Grapevine Phenology, Grape Composition, and Wine Production and Quality for Bordeaux, France. *Am J Enol Vitic* **51**: 249–261.

Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13**: 484–492.

Jullian Fabres P, Collins C, Cavagnaro TR, Rodríguez López CM. 2017. A Concise review on multi-omics data integration for *terroir* analysis in *Vitis vinifera*. *Front Plant Sci* **8**: 1065.

Kalantidis K, Schumacher HT, Alexiadis T, Helm JM. 2008. RNA silencing movement in plants. *Biol Cell* **100**: 13–26.

Kato M, Miura A, Bender J, Jacobsen SE, Kakutani T. 2003. Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis. *Curr Biol* **13**: 421–426.

Kehr J, Kragler F. 2018. Long distance RNA movement. *New Phytol* **218**: 29–40.

Keller M. 2010. Managing grapevines to optimise fruit development in a challenging environment: a climate change primer for viticulturists. *Aust J Grape Wine Res* **16**: 56–69.

Kim M, Canio W, Kessler S, Sinha N. 2001. Developmental changes due to long-distance movement of a homeobox fusion transcript in tomato. *Science* **293**: 287–289.

Kitimu SR, Taylor J, March TJ, Tairo F, Wilkinson MJ, Rodríguez López CM. 2015. Meristem micropropagation of cassava (*Manihot esculenta*) evokes genome-wide changes in DNA methylation. *Front Plant Sci* **6**: 590.

Koundouras S, Marinos V, Gkoulioti A, Kotseridis Y, van Leeuwen C. 2006. Influence of vineyard location and vine water status on fruit maturation of nonirrigated cv. Agiorgitiko

(*Vitis vinifera* L.). Effects on wine phenolic and aroma components. *J Agric Food Chem* **54**: 5077–5086.

Kumar G, Rattan UK, Singh AK. 2016. Chilling-mediated DNA methylation changes during dormancy and its release reveal the importance of epigenetic regulation during winter dormancy in apple (*Malus x domestica* Borkh.). *PLoS One* **11**: e0149934.

Kurdyukov S, Bullock M. 2016. DNA methylation analysis: choosing the right method. *Biology (Basel)* **5**: 3.

Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* **11**: 204–220.

Lecourieux F, Kappel C, Pieri P, Charon J, Pillet J, Hilbert G, Renaud C, Gomès E, Delrot S, Lecourieux D. 2017. Dissecting the biochemical and transcriptomic effects of a locally applied heat treatment on developing cabernet sauvignon grape berries. *Front Plant Sci* **8**: 53.

Lewsey M, Robertson FC, Canto T, Palukaitis P, Carr JP. 2007. Selective targeting of miRNA-regulated plant development by a viral counter-silencing protein. *Plant J* **50**: 240–252.

Lewsey MG, Hardcastle TJ, Melnyk CW, Molnar A, Valli A, Urich MA, Nery JR, Baulcombe DC, Ecker JR. 2016. Mobile small RNAs regulate genome-wide DNA methylation. *Proc Natl Acad Sci USA* **113**: E801–10.

Li D, Zhang B, Xing X, Wang T. 2015. Combining MeDIP-seq and MRE-seq to investigate genome-wide CpG methylation. *Methods* **72**: 29–40.

Liu G-T, Wang J-F, Cramer G, Dai Z-W, Duan W, Xu H-G, Wu B-H, Fan P-G, Wang L-J, Li S-H. 2012. Transcriptomic analysis of grape (*Vitis vinifera* L.) leaves during and after recovery from heat stress. *BMC Plant Biol* **12**: 174.

Liu R, How-Kit A, Stammitti L, Teyssier E, Rolin D, Mortain-Bertrand A, Halle S, Liu M, Kong J, Wu C, et al. 2015. A DEMETER-like DNA demethylase governs tomato fruit ripening. *Proc Natl Acad Sci USA* **112**: 10804–10809.

Magris G, Di Gaspero G, Marroni F, Zenoni S, Tornielli GB, Celii M, De Paoli E, Pezzotti M, Conte F, Paci P, et al. 2019. Genetic, epigenetic and genomic effects on variation of gene expression among grape varieties. *Plant J* **99**: 895–909.

Manning K, Tör M, Poole M, Hong Y, Thompson AJ, King GJ, Giovannoni JJ, Seymour GB. 2006. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat Genet* **38**: 948–952.

Marfil C, Ibañez V, Alonso R, Varela A, Bottini R, Masuelli R, Fontana A, Berli F. 2019. Changes in grapevine DNA methylation and polyphenols content induced by solar

ultraviolet-B radiation, water deficit and abscisic acid spray treatments. *Plant Physiol Biochem* **135**: 287–294.

Martínez-Lüscher J, Sánchez-Díaz M, Delrot S, Aguirreolea J, Pascual I, Gomès E. 2014. Ultraviolet-B radiation and water deficit interact to alter flavonol and anthocyanin profiles in grapevine berries through transcriptomic regulation. *Plant Cell Physiol* **55**: 1925–1936.

Miguel C, Marum L. 2011. An epigenetic view of plant cells cultured in vitro: somaclonal variation and beyond. *J Exp Bot* **62**: 3713–3725.

Molnar A, Melnyk CW, Bassett A, Hardcastle TJ, Dunn R, Baulcombe DC. 2010. Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells. *Science* **328**: 872–875.

Mori K, Goto-Yamamoto N, Kitayama M, Hashizume K. 2007. Loss of anthocyanins in red-wine grape under high temperature. *J Exp Bot* **58**: 1935–1945.

Mori K, Saito H, Goto-yamamoto N, Kitayama M, Kobayashi S, Sugaya S, Gemma H, Hashizume K. 2005. Effects of abscisic acid treatment and night temperatures on anthocyanincomposition in Pinot noir grapes. *Vitis* **44**: 161–165.

Ocaña J, Walter B, Schellenbaum P. 2013. Stable MSAP markers for the distinction of *Vitis vinifera* cv Pinot noir clones. *Mol Biotechnol* **55**: 236–248.

Pereira GE, Gaudillere JP, Pieri P, Hilbert G, Maucourt M, Deborde C, Moing A, Rolin D. 2006. Microclimate influence on mineral and metabolic profiles of grape berries. *J Agric Food Chem* **54**: 6765–6775.

Pontin MA, Piccoli PN, Francisco R, Bottini R, Martinez-Zapater JM, Lijavetzky D. 2010. Transcriptome changes in grapevine (*Vitis vinifera* L.) cv. Malbec leaves induced by ultraviolet-B radiation. *BMC Plant Biol* **10**: 224.

Rankine BC, Fornachon JCM, Boehm EW, Cellier KM. 1971. Influence of grape variety, climate and soil on grape composition and on the composition and quality of table wines. *Vitis* **10**: 33–50.

Reynolds AG. 2010. Viticultural and vineyard management practices and their effects on grape and wine quality. In *Managing wine quality* (ed. A.G. Reynolds), Vol. Volume 1: Viticulture and wine quality of, pp. 365–444, Woodhead Publishing Limited, Cambridge, UK.

Robinson AL, Adams DO, Boss PK, Heymann H, Solomon PS, Trengove RD. 2012. Influence of geographic Ooigin on the sensory characteristics and wine composition of *Vitis vinifera* cv. Cabernet Sauvignon wines from Australia. *Am J Enol Vitic* **63**: 467–476.

Robinson S, Sandercock N. 2014. *An analysis of climate, soil and topographic information to aid the understanding of barossa terroir*. SA: Primary Industries and Regions South Australia, Adelaide, Australia.

Rodríguez López CM, Wilkinson MJ. 2015. Epi-fingerprinting and epi-interventions for improved crop production and food quality. *Front Plant Sci* **6**: 397.

Romero P, Fernandez-Fernandez J, Botia Ordaz P. 2016. Interannual climatic variability effects on yield, berry and wine quality indices in long-term deficit irrigated grapevines, determined by multivariate analysis. *IJWR* **8**: 3–17.

Schellenbaum P, Mohler V, Wenzel G, Walter B. 2008. Variation in DNA methylation patterns of grapevine somaclones (*Vitis vinifera* L.). *BMC Plant Biol* **8**: 78.

Serra I, Strever A, Myburgh PA, Deloire A. 2014. Review: the interaction between rootstocks and cultivars (*Vitis vinifera* L.) to enhance drought tolerance in grapevine. *Aust J Grape Wine Res* **20**: 1–14.

Shankar S, Singh G, Srivastava RK. 2007. Chemoprevention by resveratrol: molecular mechanisms and therapeutic potential. *Front Biosci* **12**: 4839–4854.

Smith S, Codrington I, Robertson M, Smart R. 1988. Viticultural and oenoiogical implications of leaf removal for New Zealand vineyards. In (eds. S. Smith, I. Codrington, M. Robertson, and R. Smart), pp. 127–133, New Zeland Society for Viticulture and Oenology, Auckland.

Stevens M, Cheng JB, Li D, Xie M, Hong C, Maire CL, Ligon KL, Hirst M, Marra MA, Costello JF, et al. 2013. Estimating absolute methylation levels at single-CpG resolution from methylation enrichment and restriction enzyme sequencing methods. *Genome Res* **23**: 1541–1553.

Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE. 2014. Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat Struct Mol Biol* **21**: 64–72.

Sugita M, Sugiura M. 1996. Regulation of gene expression in chloroplasts of higher plants. *Plant Mol Biol* **32**: 315–326.

Sugiura T, Sumida H, Yokoyama S, Ono H. 2012. Overview of recent effects of global warming on agricultural production in Japan. *JARQ* **46**: 7–13.

Sun Z, Cunningham J, Slager S, Kocher J-P. 2015. Base resolution methylome profiling: considerations in platform selection, data preprocessing and analysis. *Epigenomics* **7**: 813–828.

Suzuki M, Nakabayashi R, Ogata Y, Sakurai N, Tokimatsu T, Goto S, Suzuki M, Jasinski M, Martinoia E, Otagaki S, et al. 2015. Multiomics in grape berry skin revealed specific

induction of the stilbene synthetic pathway by ultraviolet-C irradiation. *Plant Physiol* **168**: 47–59.

Suzuki MM, Bird A. 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* **9**: 465–476.

Sweetman C, Wong DC, Ford CM, Drew DP. 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**: 691.

Takuno S, Ran J-H, Gaut BS. 2016. Evolutionary patterns of genic DNA methylation vary across land plants. *Nat Plants* **2**: 15222.

Thieme CJ, Rojas-Triana M, Stecyk E, Schudoma C, Zhang W, Yang L, Miñambres M, Walther D, Schulze WX, Paz-Ares J, et al. 2015. Endogenous Arabidopsis messenger RNAs transported to distant tissues. *Nat Plants* **1**: 15025.

Tyunin AP, Kiselev KV, Karetin YA. 2013. Differences in the methylation patterns of the VaSTS1 and VaSTS10 genes of *Vitis amurensis* Rupr. *Biotechnol Lett* **35**: 1525–1532.

Valat L, Fuchs M, Burrus M. 2006. Transgenic grapevine rootstock clones expressing the coat protein or movement protein genes of Grapevine fanleaf virus: Characterization and reaction to virus infection upon protoplast electroporation. *Plant Sci* **170**: 739–747.

van Leeuwen C. 2010. Terroir: the effect of the physical environment on vine growth, grape ripening and wine sensory attributes. In *Managing Wine Quality* (ed. A.G. Reynolds), pp. 273–315, Woodhead Publishing, Cambridge, UK.

van Leeuwen C, Seguin G. 2006. The concept of terroir in viticulture. *Journal of Wine Research* **17**: 1–10.

Vaughn MW, Tanurdzić M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A, et al. 2007. Epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Biol* **5**: e174.

Verhoeven KJF, Jansen JJ, van Dijk PJ, Biere A. 2010. Stress-induced DNA methylation changes and their heritability in asexual dandelions. *New Phytol* **185**: 1108–1118.

Vrebalov J, Ruezinsky D, Padmanabhan V, White R, Medrano D, Drake R, Schuch W, Giovannoni J. 2002. A MADS-box gene necessary for fruit ripening at the tomato ripening-inhibitor (rin) locus. *Science* **296**: 343–346.

Warschefsky EJ, Klein LL, Frank MH, Chitwood DH, Londo JP, von Wettberg EJB, Miller AJ. 2016. Rootstocks: diversity, domestication, and impacts on shoot phenotypes. *Trends Plant Sci* **21**: 418–437.

Webb LB, Whetton PH, Barlow EWR. 2007. Modelled impact of future climate change on the phenology of winegrapes in Australia. *Aust J Grape Wine Res* **13**: 165–175.

Wine Australia. 2019. Australian wine sector's contribution to the Australian economy grows to $45.5 billion . https://www.wineaustralia.com/news/market-bulletin/issue-178 (Accessed November 14, 2019).

Wolf TK, Pool RM, Mattick LR. 1986. Responses of young chardonnay grapevines to shoot tipping, ethephon, and basal leaf removal. *Am J Enol Vitic* **37**: 263–268.

Wreczycka K, Gosdschan A, Yusuf D, Grüning B, Assenov Y, Akalin A. 2017. Strategies for analyzing bisulfite sequencing data. *J Biotechnol* **261**: 105–115.

Xie H, Konate M, Sai N, Tesfamicael KG, Cavagnaro T, Gilliham M, Breen J, Metcalfe A, Stephen JR, De Bei R, et al. 2017. Global DNA methylation patterns can play a role in defining *terroir* in grapevine (*Vitis vinifera* cv. Shiraz). *Front Plant Sci* **8**: 1860.

Yang Y, Mao L, Jittayasothorn Y, Kang Y, Jiao C, Fei Z, Zhong G-Y. 2015. Messenger RNA exchange between scions and rootstocks in grafted grapevines. *BMC Plant Biol* **15**: 251.

Zamboni A, Di Carli M, Guzzo F, Stocchero M, Zenoni S, Ferrarini A, Tononi P, Toffali K, Desiderio A, Lilley KS, et al. 2010. Identification of putative stage-specific grapevine berry biomarkers and omics data integration into networks. *Plant Physiol* **154**: 1439–1459.

Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**: 916–919.

Zerihun A, McClymont L, Lanyon D, Goodwin I, Gibberd M. 2015. Deconvoluting effects of vine and soil properties on grape berry composition. *J Sci Food Agric* **95**: 193–203.

Zhang B, Tieman DM, Jiao C, Xu Y, Chen K, Fei Z, Giovannoni JJ, Klee HJ. 2016a. Chilling-induced tomato flavor loss is associated with altered volatile synthesis and transient changes in DNA methylation. *Proc Natl Acad Sci USA* **113**: 12580–12585.

Zhang L, Marguerit E, Rossdeutsch L, Ollat N, Gambetta GA. 2016b. The influence of grapevine rootstocks on scion growth and drought resistance. *Theor Exp Plant Physiol* **28**: 143–157.

Zhang W, Thieme CJ, Kollwig G, Apelt F, Yang L, Winter N, Andresen N, Walther D, Kragler F. 2016c. tRNA-related sequences trigger systemic mRNA transport in plants. *Plant Cell* **28**: 1237–1249.

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW-L, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, et al. 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* **126**: 1189–1201.

Zhang Y-Y, Fischer M, Colot V, Bossdorf O. 2013. Epigenetic variation creates potential for evolution of plant phenotypic plasticity. *New Phytol* **197**: 314–322.

Zhong S, Fei Z, Chen Y-R, Zheng Y, Huang M, Vrebalov J, McQuinn R, Gapper N, Liu B, Xiang J, et al. 2013. Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. *Nat Biotechnol* **31**: 154–159.

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. 2007. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* **39**: 61–69.