Sequence-function relationships of the *Vitis vinifera* L. terpene synthase (*VviTPS*) family towards understanding the grapevine flower volatilome

by

Samuel Jacobus Smit

Dissertation presented for the degree of
Doctor of Philosophy
Agricultural Sciences

at
Stellenbosch University
Department of Viticulture and Oenology, Institute for Wine Biotechnology,
Faculty of AgriSciences

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF.

Supervisor: Dr. Philip Richard Young
Co-supervisor: Prof. Melané Alethea Vivier

March 2020

Stellenbosch University https://scholar.sun.ac.za

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated) that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: March 2020

Copyright © 2020 Stellenbosch University
All rights reserved

Stellenbosch University https://scholar.sun.ac.za

Summary

Terpenes are ubiquitous to plants and represent the most diverse class of natural products. More than 50 000 terpenes have been described in nature with the enzymes involved in their biosynthesis, namely terpene synthases (TPS), facilitating some of the most complex catalytic activities observed in nature. Terpenes are hydrocarbons that are made of five carbon building blocks. The coupling and prenylation of these building blocks result in different substrates that are characteristic to the biochemical properties of the synthesised terpene. The C_{10} monoterpene and C_{15} sesquiterpene compounds are extensively studied due to their volatility. These terpene classes are often associated with the pleasant aromas emitted by flowers, fulfilling important roles as volatile attractants. Terpenes are also known to have strong antimicrobial and insecticidal activities. These metabolites are important defence compounds and are therefore also considered to be specialised metabolites due to the fitness advantage associated with their ecophysiological activities.

Grapevine, *Vitis vinifera* L. (Vitaceae), has one of the largest TPS families, with the majority of these genes involved in mono- and sesquiterpene biosynthesis. These terpene classes are extensively studied for their organoleptic properties in grapes and wine, most often associated with aromatic wines that have floral, Muscat or pepper aromas. The genetic potential of grapevine TPS genes is largely underappreciated seeing that most studies focus on a select few terpene classes that are relevant to wine flavour and aroma. Limited studies have provided some insight into the *in planta* emission of grapevine terpenes. Grapevine flowers have been identified is a promising organ for terpene biosynthesis due to extensive upregulation of the TPS genes and a concomitant emission of volatile terpenes. Grapevine flowers were therefore characterised for their volatile profiles to identify chemotypic differences. This resulted in the identification of unique chemotypes, with cultivar-specific major sesquiterpene volatiles observed.

The genetic factors that contribute to the volatile differences were analysed through functional and computational characterisation. This resulted in functional characterisation of numerous sesquiterpene synthases with aberrant mutations rendering more than half of the isolated genes non-functional. Furthermore, a novel sesquiterpene synthase involved in the unique chemotype of the cultivar Muscat of Alexandria was characterised. This enzyme showed a unique enzyme active site that was linked to the biosynthesis of (E)-β-farnesene. In addition to functional characterisation of genes was the annotation and computational characterisation of the TPS gene families for three diploid grapevine genomes. This allowed for new and fundamentally important insight into how this gene family differs between genotypes.

The TPS gene family of grapevine is of great importance from an ecophysiological and economic perspective. By studying the genetic and chemotypic variations in multiple genotypes it was possible to characterise the grapevine TPS landscape. The insights gained from this study

Stellenbosch University https://scholar.sun.ac.za

provided important fundamental knowledge that furthers our understanding of the complex biochemical and genetic processes involved in grapevine terpene biosynthesis.

Stellenbosch University https://scholar.sun.ac.za

Opsomming

Terpene is alomteenwoordig in plante en word beskou as die mees diverse klas natuurlike produkte. Meer as 50 000 terpene is al beskryf; hul biosintese word gedryf deur terpeen sintases (TPS), 'n groep ensieme wat van die mees komplekse reaksies in die natuur kataliseer. Terpene bestaan uit saamgestelde 5-koolstof (C_5) eenhede. Hierdie eenhede word eerstens deur preniel transferases verbind, wat aanleiding gee tot die verlengde kenmerkende substrate vir TPS ensieme. Die 10-koolstof (C_{10}) monoterpene en 15-koolstof (C_{15}) sesquiterpene is baie vlugtig en word dus wyd bestudeer vir hul chemiese rol in plant kommunikasie. Terpene is ook belangrike metaboliete in plant verdediging weens sterk antimikrobiese en insekwerende kenmerke. As gevolg van dié funksies word terpene beskou as gespesialiseerde metaboliete weens die natuurlike voordele wat dit tot die plant toevoeg.

Druifplante (*Vitis vinifera* L.) het een van die grootste TPS families (hierna as *VviTPS* verwys) met die oorhoofse meerderheid betrokke in die vorming van mono- en sesquiterpene. Hierdie klasse van terpene dra by tot die unieke kultivar-spesifieke aromas van sekere wyne, soos bv. blom-, muskaat- of peper aromas wat aanleiding gee dat terpene uit hierdie oogpunt goed bestudeer is. Die volle genetiese potensiaal van *VviTPS* gene is egter nog grootliks onbekend, aangesien die fokus sover so sterk geval het op slegs die koppelinge met wyn aroma en reuk. Die beperkte hoeveelheid studies wat wel die rolle van terpene in die plantliggaam bestudeer het, sluit die analise van terpeen produksie vanaf wingerd blomme in. Geenuitdrukking en die produksie van vlugtige terpene vanaf die blomme van 'n paar kultivars gee 'n aanduiding dat blom terpeenprofiele kenmerkend kan wees van 'n kultivar. Die meerderheid van terpene wat in wingerdblomme gemeet is val in die sesquiterpeen groep.

In hierdie studie was die genetiese en metaboliet verskille in wingerd blomme bestudeer om terpeen biosintese beter te verstaan. *VviTPS* gene was geïsoleer vanaf menige kultivars en bestudeer om hul funksie te bepaal. Meer as die helfte van dié gene was nie funksioneel nie weens mutasies uniek tot 'n kultivar. Verder was 'n unieke geen geïsoleer wat betrokke is by die vorming van (*E*)-β-farneseen, 'n terpeen wat betrokke is by die unieke blomprofiel van die kultivar Muskat van Aleksandrië. Isolasie en karakterisering van gene was opgevolg met die toepassing van bioinformatika om die *VviTPS* geen families van drie diploïede kultivars te annoteer. Hierdie annotasies was uiters insiggewend en het gelei tot nuwe kennis rakende die *VviTPS* genetiese verskille tussen kultivar soorte.

Die *VviTPS* familie is van groot belang uit die oogpunt van ekofisiologiese aanpassings van druifplante en ook vir die ekonomiese belang van die terpene ten opsigte van wynaroma. Deur die genetiese en metaboliese verskille in 'n verskeidenheid van kultivars te bestudeer kon die kompleksiteit van die *VviTPS* familie verder ontrafel word. Hierdie studie voeg dus nuwe kennis tot die wingerd terpeen veld, veral rakende die genetiese, metaboliese en biochemiese prosesse betrokke by terpeen biosintese.

Stellenbosch University https://scholar.sun.ac.za

To Sarah, Mackenzie and Hazel
for reminding me to enjoy the journey.

Stellenbosch University https://scholar.sun.ac.za

Biographical sketch

Samuel Jacobus (Cobus) Smit was born in Polokwane, South Africa on 5 January 1990 and was raised in Kuilsriver. He matriculated at Paul Roos Gymnasium on 2008 and commenced his studies at the University of Stellenbosch in 2009 where he enrolled for a BSc-degree in Molecular Biology and Biotechnology. After graduating in 2012, he pursued postgraduate studies, obtaining a HonsBSc-degree in Wine Biotechnology Cum Laude in 2013 and a MSc in Wine biotechnology Cum Laude in 2016.

Stellenbosch University https://scholar.sun.ac.za

Acknowledgements

The last four years have been an incredible journey which would not have been possible without the help, support and encouragement from numerous people.

I would like to thank Philip for his patience and guidance. I started with you back in 2013 and can truly say that we have become good friends over the years. Thanks for all the coffee and extensive critical discussions about terpenes. You have been an amazing mentor and teacher.

To Melané, thank you for the opportunity to be a member of your lab. You have always been constructively critical and often pulled Philip and I out of a rabbit hole. Thank you for helping me stay on track and providing me with opportunities to better myself as an aspiring academic.

To my wife Sarah, you have been a constant stream of love and support. Thank you for believing in me and taking this journey with me. Without you this dream would never have come true. I am incredibly thankful for you!

To Mackenzie and Hazel, the two most amazing little girls a father could ask for. You remind me every day to laugh and be silly.

To my parents, for your encouragement, support and constant prayers.

To Jan, Frances, Carl, Meg, Ivan and Nandi for pretending to understand what I do and always listening when a talk science. You have all been the most amazing supporters!

To my Plant Lab friends, it has been great fun in the lab, and I will miss you all! Thank you to Anke, for being a good friend and always being willing to help and encouraging me to keep going on the bad days.

The Institute for Wine Biotechnology for affording me the opportunity to further my studies and financial support.

The Harry Crossley Foundation, the National Research Foundation and Stellenbosch University for financial assistance

My heavenly Father for the capacity and grace to pursue my dream.

Stellenbosch University https://scholar.sun.ac.za

Preface

This dissertation is presented as a compilation of five chapters. Each chapter is introduced separately. Chapter 1 consists of literature review, as well as the research aims and objectives. This dissertation consists of three research chapters: Chapter 2 consists of a research article that was published in Frontiers in Plant Science. Chapter 3 is written in the style of Frontiers in Plant Science to which it was submitted for publication and is currently pending review. Chapter 4 is written in the style of BMC Plant Sciences to which it will be submitted. The study is concluded in Chapter 5.

Chapter 1 **General introduction, literature overview, and aims and objectives**

Chapter 2 **Research publication**
Linking terpene synthases to sesquiterpene metabolism in grapevine flowers

Chapter 3 **Research results**
The grapevine terpene synthase (VviTPS) compendium: Comparative (within species) genomics to predict catalytic functions for genes involved specialised metabolism

Chapter 4 **Research results**
Seeing the forest through the (phylogenetic) trees: functional characterisation of VviTPS paralogues and orthologues

Chapter 5 **General discussion and conclusions**

Stellenbosch University https://scholar.sun.ac.za

Table of Contents

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

Glossary

Term	Definition
Admixture	The mixing of two or more genetically differentiated populations.
Active site	Catalytic region of a terpene synthase, determined by the position of divalent metal binding motifs (DDxxD, NSE/DTE or DxDD).
Bottleneck	A temporary marked reduction in population size.
Carbocation	Reactive hydrocarbon intermediate through which terpene biosynthesis proceeds.
Domestication syndrome	The suite of traits that marks a crop's divergence from its wild ancestor(s).
FALCON-UNZIP	A haplotype aware sequence assembly algorithm. It is used to generate primary and haplotig assemblies when sequencing diploid genomes.
GWAS	Genome-wide association studies that search for a statistical association between a phenotype and a particular allele by screening loci (most commonly by genotyping SNPs) across the entire genome.
Haplotype	The combination of alleles or genetic markers found on a single chromosome of an individual.
Homolog	A gene related to a second gene by descent from a common ancestral DNA sequence. The term, homolog, may apply to the relationship between genes separated by the event of speciation (see ortholog) or to the relationship between genes separated by the event of genetic duplication (see paralog).
Introgression	The incorporation of genetic material from one population or species into another by hybridization and backcrossing.
Kairomone	A chemical substance emitted by an organism and detected by another of a different species which gains advantage from this, e.g. a parasite seeking a host.
Microarray	A collection of microscopic DNA spots attached to a solid surface. It is used to measure gene expression.
N50	A weighted median statistic such that 50% of the entire assembly is contained in contigs or scaffolds equal to or larger than this value. This value defines the assembly quality in terms of contiguity.
Ortholog	One of a set of homologous genes that have diverged from each other as a consequence of speciation.
Paralog	A pair of genes that derives from the same ancestral gene and now reside at different locations within the same genome.
Read depth	The number of unique reads that include a given nucleotide in the reconstructed sequence.
Selective sweep(s)	Increases in frequency of an allele and closely linked chromosomal segments that are due to positive selection. Sweeps initially reduce variation and subsequently lead to a local excess of rare alleles as new unique mutations accumulate.

Stellenbosch University https://scholar.sun.ac.za

Semiochemical	Pheromone or other chemical that conveys a signal from one organism to another so as to modify the behaviour of the recipient organism.

Abbreviations

CYP450	Cytochrome P450
DMAPP	dimethylallyl diphosphate
DNA	deoxyribonucleic acid
EST	Expressed sequence tags
FPP	Farnesyl diphosphate
GGPP	Geranyl geranyl diphosphate
GPP	Geranyl diphosphate
GWAS	Genome-wide association study
IPP	isopentenyl diphosphate
MEP	2-C-methyl-D-erythritol 4-phosphate pathway
MVA	Mevalonate pathway
RNA	Ribonucleic acid
RNAseq	Ribonucleic acid sequencing
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variation
SSR	Simple sequence repeat
SV	Structural Variation
TPS	Terpene Synthase (Vvi prefix used when referring to *Vitis vinifera*)
PacBio	Pacific BioSciences
SMRT	Single Molecule Real-Time sequencing (developed by PacBio)
MSA	Multiple sequence alignment
Chr.	Chromosome

Stellenbosch University https://scholar.sun.ac.za

Chapter 1

General Introduction, Literature Overview, and Aims and Objectives

Stellenbosch University https://scholar.sun.ac.za

General Introduction, Literature Overview, and Aims and Objectives

1.1 Introduction

Terpenes are ubiquitous to plants, with the more than 50,000 described examples in nature representing the most diverse class of naturally occurring compounds (Buckingham et al., 2015; Christianson, 2017; Osbourn and Lanzotti, 2009). Terpene biosynthesis is catalysed by terpene synthases (TPS) with most plants having a medium to large sized family of 40-152 *TPS* genes (Chen et al., 2011). Terpenes are regarded as specialised secondary metabolites due to the fact that they allow for specific abiotic and biotic responses and/or interactions. Terpenes are therefore extensively studied for their function in plant defence, which can be a direct interaction with, for example an insect or herbivore, or indirect, by attracting natural predators to serve as bodyguards (Bennett and Wallsgrove, 1994; Heil, 2009; Moore et al., 2014).

Terpenes involved in direct defence are often toxic, bitter or repulsive. The oleoresin secreted by certain tree species, for example, is rich in antimicrobial and insecticidal terpenoids (Bohlmann et al., 1998a; Miller et al., 2005). Terpenes can also act as kairomones to various pests with the sesquiterpene (*E*)-β-farnesene, for example being a known alarm pheromone to aphids (Francis et al., 2004; Schnee et al., 2002; Vandermoten et al., 2012).

Terpenes display extensive chemical diversity. Adaptation of plants to specific ecological niches have contributed to this diversity which is achieved through enzymatic modifications of terpenes, allowing for specialised adaption (Nagegowda, 2010; Pichersky et al., 2006). Three key gene families involved in terpene modification are the cytochrome P450's (CYP450's), glycosyl transferases and their inverse counterparts, glycosyl hydrolases (Bönisch et al., 2014; Ilc et al., 2018; Rivas et al., 2013; Schwab et al., 2015). The CYP450's often hydroxylate or oxygenate terpenes to direct a pool of diverse structurers to a targeted function, in essence tailoring their chemical properties (Höfer et al., 2013; Ilc et al., 2016a, 2016b; Sohrabi et al., 2015). Glycosyl transferases allow for increased solubility for transport or storage of terpenes, with the hydrolases removing the glycosyl moiety (Bönisch et al., 2014). The term terpenoid is used in reference to such modified terpene structures.

Grapevine (*Vitis vinifera* L.) has a large *TPS* gene family (hereafter referred to as *VviTPS*) that consists of 152 *VviTPS*-like loci (Martin et al., 2010). Our current understanding of this gene family is based on a few key studies. Firstly, the *VviTPS* gene family has been annotated for the grapevine reference genome, which showed that the gene family is highly duplicated and expanded in this species, compared to other annotated plant genomes (Jaillon et al., 2007; Martin et al., 2010). Secondly, numerous *VviTPS* genes have been functionally characterised with observed terpenes largely associated with important organoleptic properties of grapes and wines (Martin et al., 2010; Siebert et al., 2008; Wood et al., 2008). Thirdly, the role of terpenes in modulating flavour and aroma

Stellenbosch University https://scholar.sun.ac.za

profiles have been extensively studied with numerous metabolomic and transcriptomic analyses conducted, reviewed in Schwab and Wüst (2015).

The availability of these "omic" resources have significantly contributed to the current understanding of terpene biosynthesis in grapevine (D'Onofrio et al., 2014; Fasoli et al., 2012; Jaillon et al., 2007; Martin et al., 2010; Matarese et al., 2013, 2014). However, the ecophysiological roles of terpenes require further analysis in grapevine. Very few studies are available that have explored grapevine terpenes from the perspective of their natural biological functions in the different plant organs. Grapevine flowers presents an interesting study case from a terpene biosynthesis point of view. There is still significant scope for further analysis of flower terpenes and the synthases involved in flower volatile emissions.

1.2 Rationale, scope and aims of the study

Commercial cultivars of grapevine have hermaphroditic flowers; this feature was a driver of the domestication of grapevine, leading to self-pollination and more consistent fruitset. Limited research has been done in terms grapevine flower terpene biosynthesis, but the available information indicates that there are potential chemotypic differences, driven by terpene profiles. Grapevine flowers show highly upregulated *VviTPS* gene expression in the grapevine expression atlas (Fasoli et al., 2012) with concordant emission of terpenes observed (Martin et al., 2009; Matarese et al., 2014). All that we know about grapevine flower terpenes come from five studies, representing fifteen cultivars, with only two studies investigating the genetic factors involved in flower terpene biosynthesis (Buchbauer et al., 1994a, 1994b, 1995; Martin et al., 2009; Matarese et al., 2014). Extensive paralogy for the *VviTPS* family and an increased mutation rate due to vegetative propagation of grapevines is hypothesised to have a direct impact on *VviTPS* functionality. A disjoint seems to exist between observed terpenes and the genetic potential (i.e. complement of coding genes) of grapevine.

The aims of this study were as follows:

(i) To evaluate *VViTPS* genes from flowers to establish to what extent terpene biosynthesis determines chemotypic differences in grapevine flowers emissions; and

(ii) To evaluate genotype-specific *VviTPS* genes from flowers using available genomic resources.

To contextualise the study, a condensed review of literature will be presented before elaborating on the objectives for each of the aims and general presentation of the thesis.

1.3 Literature review

In this review the general characteristics of a terpene compound and the enzymes that form it will be discussed, followed by a brief overview of grapevine genomics, given the importance of genomic tools in our ability to study grapevine gene families. The grapevine terpene synthase (*VviTPS*) family

Stellenbosch University https://scholar.sun.ac.za

and the key focus areas involving grapevine terpene research will be discussed in brief, including the impact grapevine domestication had on this complex gene family, particularly in the bisexual flowers.

1.3.1 Terpene classification

Terpenes and the genes that encode for them are classified according to the structure (skeleton) of the terpene metabolite and the polygenetic conservation of the encoding genes. Extensive efforts have resulted in the elucidation of important protein futures that result in specific catalytic functions. In the following sub-section these aspects will discussed to provide a fundamental understanding of *TPS* genes, proteins and their resulting metabolites.

1.3.1.1 Hydrocarbon structures and gene subfamilies

Terpenes are grouped into distinct subclasses based on the number of hydrocarbons in the terpene structure. The number of carbons is in turn, a function of the head-to-tail coupling of the C_5 building blocks, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), to generate substrates for the TPSs. Coupling of these building blocks is facilitated by prenyl transferases resulting in C_{10} geranyl diphosphate (GPP), C_{15}, farnesyl diphosphate (FPP), or C_{20}, geranyl geranyl diphosphate (GGPP) (Christianson, 2006; Lichtenthaler, 1999). TPS have conserved structural features allowing for utilisation of specific substrates with monoterpenes formed using GPP, sesquiterpenes from FPP and diterpenes from GGPP, illustrated in Figure 1.1a.

Biosynthesis of precursors and substrates are compartmentalised to two independent pathways, namely the cytosolic 2-C-methyl-D-erythritol 4-phosphate (MEP) and plastidial mevalonate (MVA) pathways, both resulting in IPP and DMAPP biosynthesis (Bloch et al., 1959; Eisenreich et al., 2001; Rohmer, 1999). Prenyl diphosphate substrates, are also used in the biosynthesis of hormones, photosynthetic pigments, sterols and electron carriers to fulfil important metabolic and physiological functions (McGarvey and Croteau, 1995), as shown in Figure 1.1b.

TPS genes are subdivided based on enzyme function and sequence identity in seven subfamilies, designated TPS-a to -g (Bohlmann et al., 1998b; Dudareva et al., 2003). The TPS-c and -e subfamilies are involved in primary metabolism, encoding for diterpene synthases involved in *ent*-copalyl diphosphate biosynthesis required for biosynthesis of the gibberellin plant hormones (Sun and Kamiya, 1994; Yamaguchi et al., 1998). The remaining subfamilies encode for TPS genes involved in specialised metabolism, as illustrated in Figure 1.1b.

1.3.1.2 The conserved features of TPS proteins

Terpene synthases are grouped into two enzymatic classes, distinguished by the amino acid motifs involved in generating the initial carbocation (reactive hydrocarbon) intermediate, illustrated in Figure 1.2. Class I enzymes contain the DDxxD and NSE/DTE motifs that bind divalent metal ions which in turn triggers catalysis through ionization of prenyl substrates. Class II synthases contain a DxDD

Stellenbosch University https://scholar.sun.ac.za

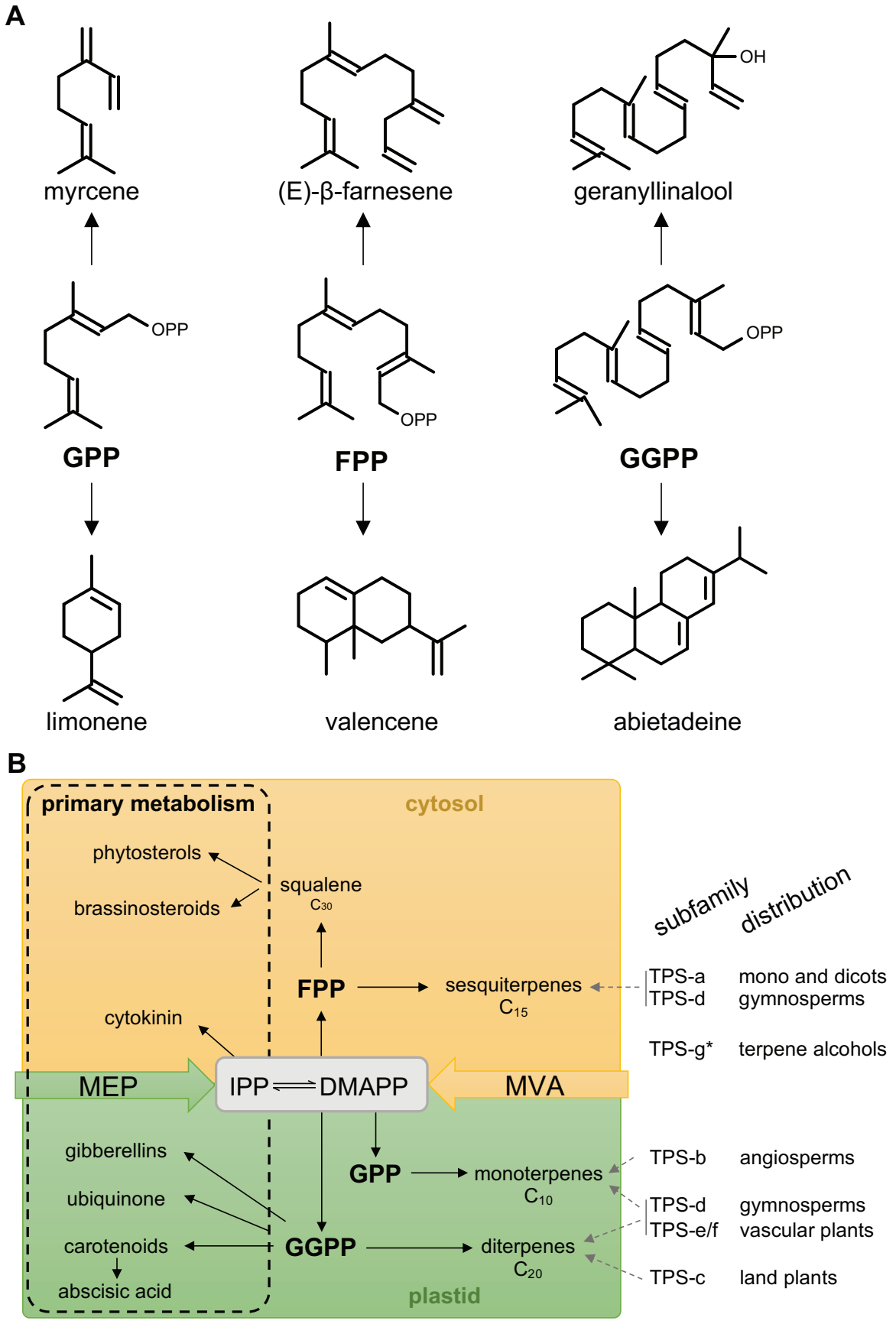

Figure 1.1 (A) Examples of cyclic and acyclic terpene skeletons that are formed from the substrates GPP, FPP or GGPP. **(B)** Compartmentalisation of the plastidial MEP and cytosolic MVA pathways that result in the universal precursors IPP and DMAPP. Examples of primary and specialised metabolites that use the different terpene substrates are shown. The subfamily distribution through the plant kingdoms is also shown, excluding the TPS-g subfamily (indicated by the asterisk).

Stellenbosch University https://scholar.sun.ac.za

motif which facilitates protonation of the substrate double bond to create the initial carbocation. After the initial carbocation has been generated, a multi-step cyclisation cascade ensues until the reaction is terminated by deprotonation or nucleophilic capture of the final carbocation (Cane, 1999; Cane and Iyengar, 1979; Köksal et al., 2011a). The majority of plant terpene synthases are classified as Class I enzymes, with Class II enzymes usually involved in primary metabolism.

The function and taxonomic distribution of TPS subfamilies (Figure 1.1b), point to a shared evolution for TPS genes, thought to have risen from the TPS-c clade of bifunctional diterpene synthases (Chen et al., 2011; Trapp and Croteau, 2001). Diterpene synthases contain three α-helical domains, designated as the α-, β- and γ-domain (Figure 1.2). The α-domain is on the C-terminal side of Class I enzymes and contains the catalytic active site, defined by the DDxxD and NSE/DTE motifs. The β-domain is on the N-terminal region and does not contain a catalytic region, however some enzymes have a second DDxxD motif in this domain. The γ-domain is inserted on the N-terminal side, in between the α- and β-domains with the γ-β interface forming an active site cavity with the β-domain DxDD motif, resulting in Class II activity (Christianson, 2006, 2008; Köksal et al., 2011a). The bifunctional abietadiene synthase has both catalytic mechanisms, resulting in the biosynthesis of diterpene acids that form part of the terpenoid rich oleoresin secreted by grand fir (*Abies grandis*) (Peters et al., 2000; Peters and Croteau, 2002b, 2002a; Ravn et al., 2002). The class I enzyme *ent*-copalyl diphosphate synthase is involved in the biosynthesis of gibberellins, but lacks a functional α-domain catalytic site, while taxadiene synthase is a class II diterpene synthase, lacking a functional γ-β active site (Köksal et al., 2011b, 2011a). The loss of the γ-domain is thought to have resulted in class I synthases that lead to mono- and sesquiterpene biosynthesis. It is not yet clear if the loss of the transit peptide, resulting in biosynthetic compartmentalisation, preceded the loss of the γ-domain (Pazouki and Niinemets, 2016).

The active site dictates how the enzyme will interact with the substrate with amino acid changes (nonsynonymous mutations) in the active site altering the product profile. In general a shallow, and therefore more restrictive active site shows fewer terpene products, while a deeper active site allows for greater movement of the substrate leading to a greater diversity of products (Degenhardt et al., 2009; Lesburg, 1997; Miller et al., 2008). Substrate accessibility within the active site therefore determines which reaction mechanisms a TPS can facilitate, commonly referred to as a carbocation cascade due to the reactive hydrocarbon intermediates through which catalysis sequentially proceeds (Christianson, 2008; Gonzalez et al., 2014; Greenhagen et al., 2006; Köllner et al., 2004; Piechulla et al., 2016; Schwab et al., 2001; Wymore et al., 2011).

Stellenbosch University https://scholar.sun.ac.za

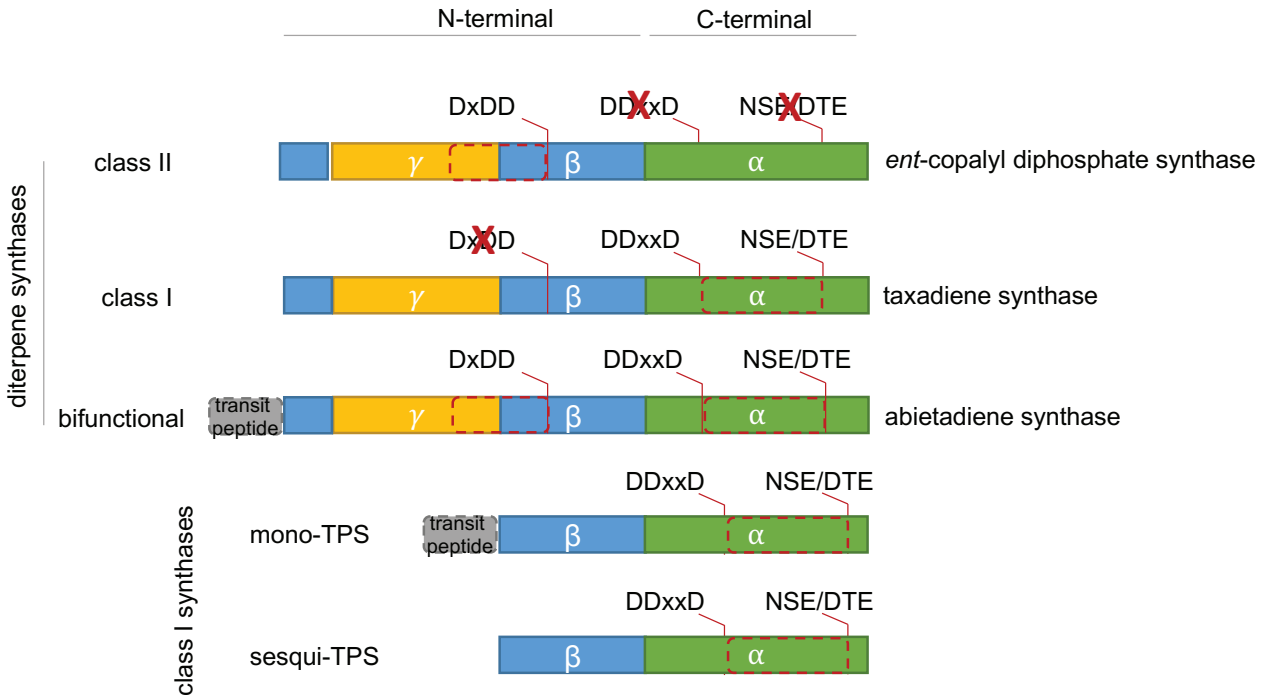

Figure 1.2 The αβγ domain structure of Class I and II terpene synthases. The DxDD, DDxxD and NSE/DTE motifs that define the active site are indicated by the dashed red squares. Red crosses indicate that the functional motif is not present, allowing for the specified enzyme class, with examples of diterpene synthases associated with these enzyme classes listed.

These cascades involve various cyclisations, rearrangements, protonation, deprotonation and nucleophilic attacks which result in the extensive diversity of chemical structures (Gao et al., 2012; Gonzalez et al., 2014; Isegawa et al., 2014). Details on the active site cavity and the carbocation cascades it can accommodate has been extensively reviewed by Christianson (2017). The aforementioned phylogenomic conservations within subfamilies have been combined with the conservation of enzyme mechanisms (enforced by functional characterisation) to identify conserved amino acid residues and motifs. The correlation between specific amino acid residues and carbocation cascades required to form specific terpenes suggests that the amino acid residues could be used to predict the TPS function (Degenhardt et al., 2009; Durairaj et al., 2019; Wymore et al., 2011).

1.3.2 The importance of terpenes in grapevine

Grape berry metabolites are extensively studied for their organoleptic properties with terpenes, phenolics and norisoprenoids imparting important flavours to both table grapes and wine, usually becoming odour active after hydrolysis of their glyosidic moiety (Maicas and Mateo, 2005; Yauk et al., 2014). Grapevine terpene studies have been largely focussed on grape berries, focussing on the molecular mechanism involved in biosynthesis and the aromatic properties of terpenes in wine (Boss, 2011). This has been extensively reviewed by Schwab and Wüst (2015) and will therefore only be discussed in brief. Monoterpenes are major contributors to grape and berry aroma with floral and Muscat aromas imparted by geraniol, citronellol, linalool and α-terpineol (Gholami et al., 1995;

Stellenbosch University https://scholar.sun.ac.za

King and Dickinson, 2000; Martin et al., 2012; Pardo et al., 2015). Aromatic cultivars like Gewürztraminer, Muscat and Riesling that are high in these terpenoids have been shown to have an increased biosynthetic flux through the MEP pathway that result in monoterpene biosynthesis (Dalla Costa et al., 2018). This flux is largely driven by functional differences inferred by point mutations of the *VviDXS* gene which determinises how much substrate is available for monoterpenes biosynthesis (Battilana et al., 2011; Dalla Costa et al., 2018; Emanuelli et al., 2010; Sun et al., 2015).

Sesquiterpenes are less involved in grape and wine aroma, with the exception of rotundone. Rotundone is the key metabolite for the sought after pepper aroma is red wine, mainly Shiraz (Siebert et al., 2008; Wood et al., 2008). Its biosynthesis is a two-step process where α-guaiene synthase, an allelic variants of *VviTPS24*, catalyses the formation of α-guaiene which is then oxygenated by a CYP450 to rotundone (Drew et al., 2015; Huang et al., 2014; Takase et al., 2015). Sesquiterpene biosynthesis, mainly volatile emission, has been largely studied in grapevine flowers where they impart cultivar-specific floral bouquets (Buchbauer et al., 1994a, 1994b, 1995; Martin et al., 2009).

In the subsequent sections the status quo of grapevine genomics will firstly be discussed as it provides the necessary context for discussions on the *VviTPS* gene family. Particular emphasis is placed on how genomic and transcriptomic recourses were generated to facilitate the molecular characterisation and annotation of the *VviTPS* family.

1.3.2.1 The grapevine genome: a brief overview of the last two decades

Domesticated grapevine (*Vitis vinifera* L.) is one of the most important horticultural crops in the world with 7.4 million hectares planted producing 77.8 million tons of grapes. Humans are thought to have interacted with grapevine over the past 20,000 years, with the active cultivation of vines hypothesised to have started 8,000 years ago (Zhou et al., 2019). The first plant genome, *Arabidopsis thaliana*, was sequenced in 2000 (The Arabidopsis Genome Initiative, 2000) with grapevine being the first woody crop to be sequenced. Two genomes were concurrently released for a heterozygous and homozygous clone of the cultivar Pinot Noir, respectively (Jaillon et al., 2007; Velasco et al., 2007). A timeline representing these events and when some of the most important genomic resources became available is shown in Figure 1.3.

Plant genomes have complex structures due to extensive repeat regions and heterozygosity. Inbreeding of the PN40024 clone of Pinot Noir reduced these complexities to some extent resulting in a more contiguous and complete genome assembly (8X coverage) than that of the ENTAV115 heterozygous genome (4-6X coverage) (Jaillon et al., 2007; Velasco et al., 2007). Due to the quality of the assembly, PN40024 was chosen to be the reference genome. Both genomes followed a shotgun approach with Sanger sequencing. In late 2007 massively parallel sequencing using the Illumina platform superseded Sanger shotgun sequencing signalling the dawn of second generation sequencing technology, reviewed in Hutchison (2007) and Metzker (2005), illustrated in Figure 1.3.

Stellenbosch University https://scholar.sun.ac.za

Illumina short-reads, an increased number of ESTs and simple sequence repeat (SSR) markers were later incorporated into the reference genome, improving the read depth to 12X as well as the ordering and orientation of contigs to chromosomes (chr.), constituting a 27% assembly improvement (Adam-Blondon, 2014; Canaguier et al., 2014; Doligez et al., 2006). This version is commonly referred to as 12x.v0. A second major assembly update (12x.v2) was made in 2017, improving the chromosome assembly even further. Despite these extensive efforts, a large portion of the contigs remain unassembled (2.65 Mb), represented as the chr. 00 pseudo-molecule. Concomitant to this assembly was an updated genome annotation (VCost.v3), incorporating various specialised gene family annotations that were not part of previous annotations (Canaguier et al., 2017).

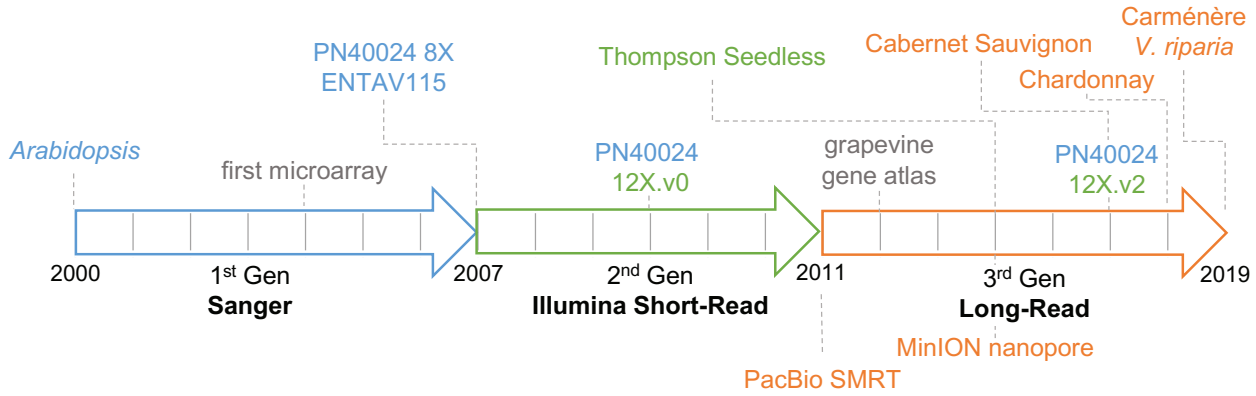

Figure 1.3 Timeline for when the most significant grapevine genomic and transcriptomic resources became available. *Arabidopsis* was the first plant to be sequenced, as indicated. Arrows indicate the type and generation of sequence technology used, with the genomic resources coloured accordingly. Major transcriptomic resources are shown in grey

Pre-genome efforts utilised expressed sequence tags (ESTs) and amplified fragment length polymorphisms (AFLP) to determine the transcript composition of different organs and during grape development (Burger and Botha, 2004; Venter et al., 2001; Vuylsteke et al., 2007). These efforts were however costly and low throughput. Microarrays to study grapevine gene expression presented the first opportunity for high-throughput analyses with the first commercial grapevine specific array becoming available in 2004 (Waters et al., 2005). Since the release of the genome, more comprehensive arrays were developed culminating in the NimbleGen 12X 135K array designed on the 12X assembly and V1 annotation of PN40024, representing 29 549 genes (Fasoli et al., 2012). The most notable application of this array was in the establishment of the grapevine gene expression atlas which provided the first comprehensive transcriptome of grapevine. The atlas considered 54 different samples, representing various tissue types, organs and developmental stages (Fasoli et al., 2012). This provided valuable insight into the transcriptional responses that drive the plant into maturation.

Transcriptome studies utilizing RNA sequencing (RNAseq) has surpassed array-based experiments with an exponential increase over the past five years (26 in 2013 to 2632 in 2018) (Matus et al., 2019). The ability to perform *de novo* assembly of RNAseq transcripts has resulted in

Stellenbosch University https://scholar.sun.ac.za

the genome-independent discovery of novel genes (also referred to as private genes) that are not present in PN40024. For example, the transcriptome of Corvina showed 180 novel genes that were not present in PN40024, while Tannat had 1873 private genes (Da Silva et al., 2013; Venturini et al., 2013). RNAseq approaches are usually narrow in focus due to the cost and time involved in preparing cDNA libraries as well as the magnitude of sequence data generated from a single sequencing run. *De novo* assemblies of transcripts furthermore require specialised bioinformatic approaches that often necessitate extensive computational power (Han et al., 2015). The rapid development of user-friendly computational tools to perform these tasks more efficiently has partly contributed to the increase in grapevine RNAseq data.

It has become increasingly evident that PN40024 does not capture the extent of genetic variability within *V. vinifera*. Genetic variability was already evident when the ENTAV115 and PN40024 genomes were compared with a later attempt to sequence the table grape cultivar Thompson Seedless, showing extensive structural differences to PN40024 (Da Silva et al., 2013; Di Genova et al., 2014). Despite the read-depth (>300), short-read sequencing of this cultivar was still hampered by extensive repeat regions and heterozygosity, resulting in a highly fragmented genome (Di Genova et al., 2014).

Third generation sequencing technology, see Figure 1.3, became accessible in 2011 through the Pacific Biosciences Single Molecule Real Time Sequencing (PacBio-SMRT; hereafter referred to as “PacBio”) platform. This platform allows for long-read sequencing (>30 kb), resulting in highly contiguous reads that are easier to assemble, but with a greater error rate (7-15%) than short-read sequencing (Rhoads and Au, 2015). However, increased read-depth (>115X) overcomes the latter limitation with a hybrid, albeit costly, approach consisting of PacBio sequencing and short-read Illumina sequencing, allowing for tremendous sequencing and assembly accuracy (Chin et al., 2016; Girollet et al., 2019; Minio et al., 2019). Since 2016, four new *V. vinifera* (Chin et al., 2016; Minio et al., 2019; Roach et al., 2018; Vondras et al., 2019) and a *V. riparia* (Girollet et al., 2019) *de novo* genome became available using PacBio sequencing.

The first application of PacBio sequencing of plant genomes was coupled with the development of specialised assembly algorithms, namely FALCON-UNZIP, that allowed for the resolution of haplotypes without relying on an existing reference assembly (Chin et al., 2016). This, in essence, results in two assemblies where the primary assembly consists of highly contiguous pseudo-molecules that contain both haplotypes. The second assembly, referred to as haplotig sequences/contigs, consists of shorter phased reads and therefore represent alternate alleles. The differences between diploid genome sequencing and short-read sequencing (i.e. reference genome) is contrasted, in Figure 1.4.

Phased diploid sequencing has resulted in less fragmented and more complete genomes (reflected by the N50-value) compared to the haploid genomes of PN40024 and ENTAV115, as summarised in Table 1.1. The diploid genomes sizes are, however, inflated (more than double the

Stellenbosch University https://scholar.sun.ac.za

haploid genomes) due to haplotype regions being missed in regions of high heterozygosity, resulting in the haplotypes being incorrectly assigned to the primary assembly (Figueroa-Balderas et al., 2019; Minio et al., 2017).

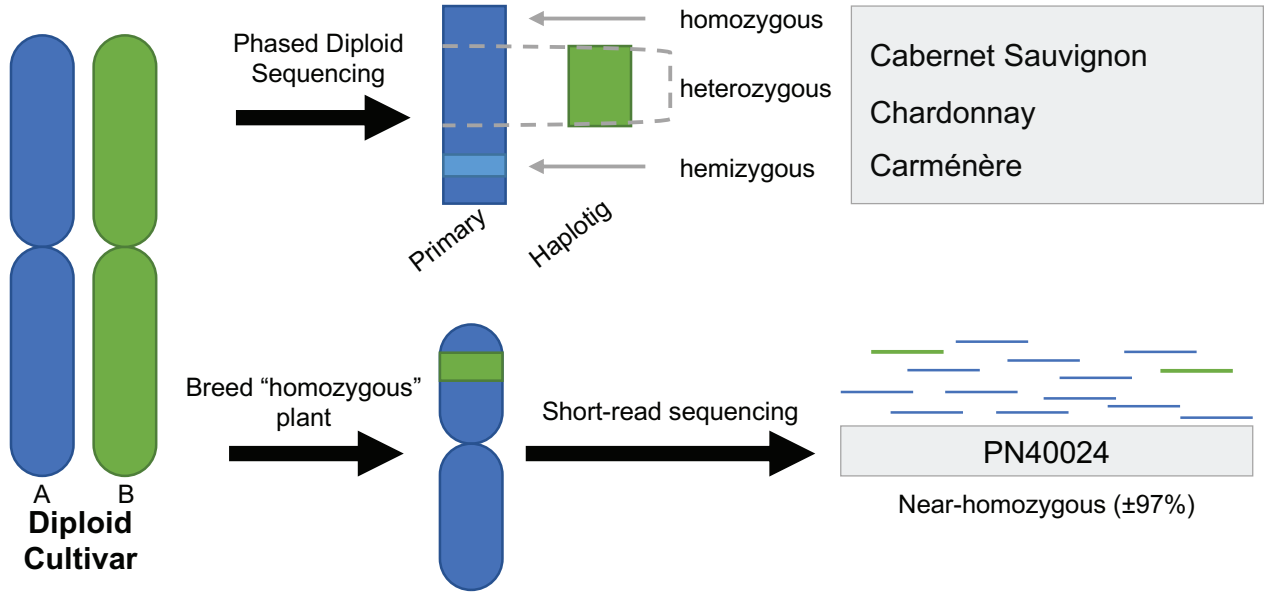

Figure 1.4 Different sequencing approaches used to generate the phased-diploid and reference genomes of grapevine.

Table 1.1 Comparative data highlighting how sequencing technology has improved the quality of available grapevine genome sequences.

Cultivar	N50	Coverage	Genome Size	Repeat Regions
PN40024	0.103 Mbp	12X	457 Mbp*	41%
ENTAV115	0.018 Mbp	4.2X	505 Mbp*	27%
Cabernet Sauvignon	2.17 Mbp	140X	959 Mbp	51%
Chardonnay	0.936 Mbp	115X	868 Mbp	39%
Carménère	1.04 Mbp	115X	1.04 Gbp	48%

* Haploid genome size

1.3.2.2 The grapevine terpene synthase (*VviTPS*) family

Terpenoids received attention in grapevine prior to the release of the genome due to their aforementioned role in modulating wine flavour and aroma. Screening of EST libraries resulted in the characterisation of the first *VviTPS* genes: a (-)-α-terpineol monoterpene synthase and two sesqui-TPSs, respectively encoding for a (+)-valencene and Germacrene D synthases (Lücker et al., 2004; Martin and Bohlmann, 2004). The *VviTPS* and stilbene synthase (*VviSTS*) families were highlighted by Jaillon et al. (2007) as being expanded in grapevine, relative to other plant genomes of the time. The initial 8x genome assembly showed 89 putative full-length and 27 pseudo *VviTPS* genes with curation of this family using the 12x.v0 assembly resulting in the identification of 152 *VviTPS*-like loci, 69 predicted as putatively functional and 83 as pseudogenes (Martin et al., 2010). This manual curation of 12x.v0 automated gene annotations resulted in extensive gene structure corrections with 54 new loci discovered. Mono- (TPS-b) and sesqui-TPS (TPS-a) genes represented

Stellenbosch University https://scholar.sun.ac.za

45 and 65 of members of the *VviTPS* family, with the TPS-g subfamily involved in terpene alcohol biosynthesis comprising of 34 genes. Thirty of the 69 putative genes were subsequently characterised through bacterial *in vitro* or yeast *in vivo* heterologous expression systems (Martin et al., 2010). This seminal work has been fundamental to our understanding of grapevine terpene biosynthesis. The expansion and complexity of this gene family suggests that domestication focussed on the flavour and aroma aspects of terpenes, evident in the number of mono- and sesquiterpene synthases.

1.3.3 Crop domestication and the impact thereof on specialised metabolism

It has become evident that grapevine genotypes display specific capacities to produce terpenes with cultivars specifically being selected for their aromatic properties, or lack thereof. Expansion of the *VviTPS* family and the presence of large *VviTPS* gene clusters (Martin et al., 2010) suggested that evolution and subsequent domestication had a significant impact on this family. It is therefore prudent to understand the forces involved in crop domestication, with the subsequent sub-sections providing a TPS-centric view of genome changes that potentially impact the *VviTPS* family.

1.3.3.1 Traits associated with domestication

A growing body of evidence suggests that crop domestication has been a protracted rather than a rapid process (Purugganan, 2019). Archaeobotanical studies on domesticated cereal crops provided the first insights to this by studying the rachis phenotype, a key domestication trait associated with loss of natural seed dispersal. It was estimated that this trait took 2 000-3 000 years to be fixed (Tanno and Wilcox, 2006). A similar observation was made for seed size enlargement (Fuller et al., 2012, 2014; Purugganan and Fuller, 2011). More recent observations using computational modelling methods and whole genome data of crops species have looked at coalescence events involved in the gradual decline of effective population size for domesticated crops. *V. vinifera*, African rice and maize all showed a similar pattern where the population size decrease precedes archaeological evidence of domestication (Cubry et al., 2018; Gaut et al., 2018; Meyer et al., 2016; Zhou et al., 2017a). This suggests that interactions with the crops before obvious domestication had an unconscious impact on the wild population.

The traits associated with a crop as a result of divergence from the wild ancestor are described as the “domestication syndrome” (Allaby, 2014). Selection for a particular trait can be either conscious (purposeful selection of a trait) or unconscious (unintended consequence of changing the conditions under which the plant is grown) (Purugganan and Fuller, 2011). In either case, genetic diversity is reduced, favouring a particular beneficial allele (this is also referred to as a selection sweep) (Zhou et al., 2019). An example of unconscious selection is flowering time (time to maturation) of maize (*Zea maize*). Changes in the genetic architecture occurred as the plant adapted to new geographic/environmental conditions and as a consequence loci were altered, resulting in divergent haplotypes (long and short maturation times) (Buckler et al., 2009).

Stellenbosch University https://scholar.sun.ac.za

Domestication syndrome implies a conscious selection process which explains why toxicity, pigments and flavour are commonly associated with this syndrome (Meyer et al., 2012). These domestication traits also include important specialised metabolites involved in plant defence and stress responses. *TPS* gene clusters of rice were shown to be divergent between species, indicative of genomic structural differences. Differences in enzyme functions were furthermore observed due to active site mutations resulting in a phenotypic change that correlates with expression and emission pattern differences in response to insect damage (Chen et al., 2019). Various terpene phytoalexins, for example capsidiol, momilactone and oryzalexins have antimicrobial properties while other terpenes, for example limonene and valencene, impart characteristic “citrus” flavours to certain fruits (Akatsuka et al., 1985; Jaillon et al., 2007; Li et al., 2015; Naim et al., 1999; Otomo et al., 2004; Sharon-Asa et al., 2003).

Phytoalexins are however, most often toxic to humans and were likely selected against during domestication. For example, a reduction in toxic (and often bitter) compounds is observed in domesticated potatoes and beans when compared to wild relatives (Johns, 1986; Lindig-Cisneros et al., 1997, 2002; Wink, 1988). In grapevine, selection sweeps are evident for berry colour, size and aroma, as well the flower sex determination (Migicovsky et al., 2017).

1.3.3.2 Domestication of *V. vinifera*

The *Vitis* genus consists of 79 species: 28 across North America, 30 across Asia, two in Europe with the remaining species spread across Mediterranean climates. The vast majority of commercial grapevines are varieties of the European species *V. vinifera* spp. *sativa* (Walker et al., 2019; Wan et al., 2013). Zhou et al. (2019) defined the events shaping domesticated grapevine into four stages, with the first representing the aforementioned protracted period of slow population decline, estimated to have started ~22 000 years ago, illustrated in Figure 1.5. Stage two of domestication is hypothesised to have started ~8 000 years ago with the purposeful cultivation of *V. vinifera* spp. *sativa* (McGovern et al., 2017). This period is the subject of extensive debate, centring around where the origin of domestication is. Archaeological evidence currently points to modern day Georgia as the origin of domesticated grapes, however, an argument can be made for a separate domestication area in Western Europe (McGovern et al., 2017; Riaz et al., 2018). Regardless of where the origin was, grapevine lacks a dramatic bottleneck that is usually associated with domestication, indicating that genetic diversity is still largely maintained exhibiting fewer phenotypic changes than what is usually observed in annual crops (Meyer et al., 2012).

Stellenbosch University https://scholar.sun.ac.za

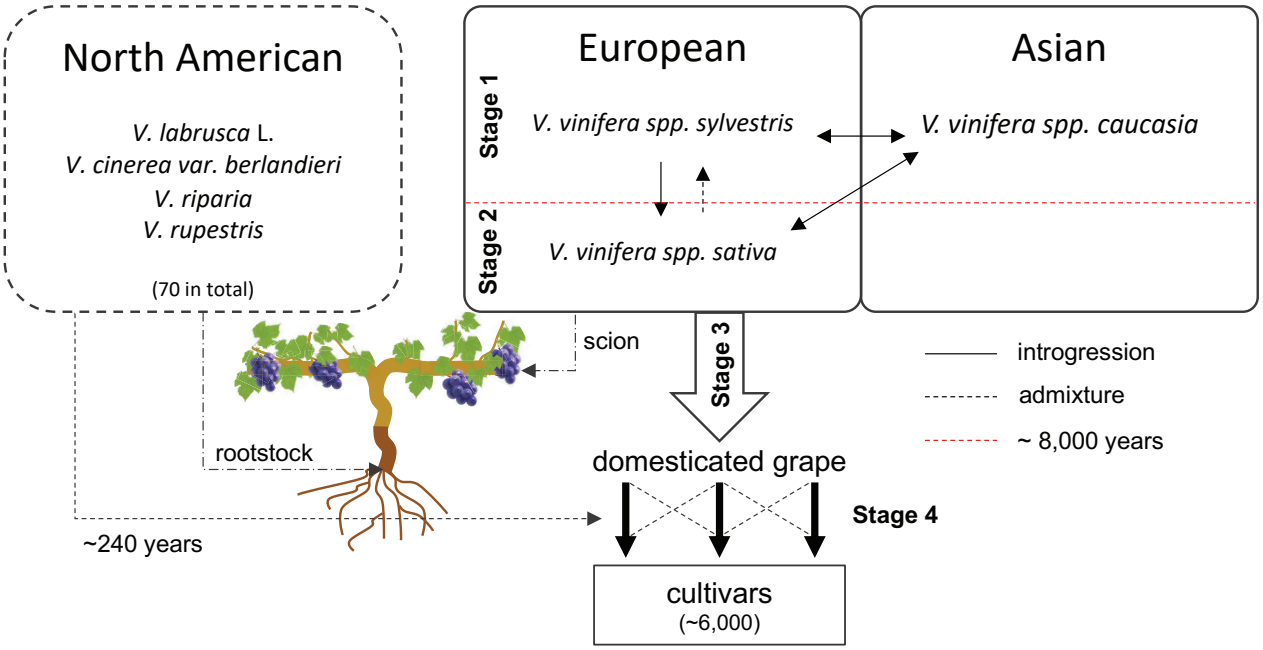

Figure 1.5 Overview of four stages of grapevine domestication and geographic separation of major species (Zhou et al., 2017b, 2019). The dashed red line indicates stage 2 of domestication. Introgression with North American species started ~240 years ago (Walker et al., 2019; Zhou et al., 2019).

Of particular interest to this study is phenotypic changes associated with sex determination. Transitioning from dioecy to hermaphroditism was one of the first traits selected for during domestication due to all individuals being fruit bearing rather than just half, as is the case with a dioecious population (Meyer et al., 2012). This is thought to have happened somewhere during the early part of stage 2, through introgression from the second European species *V. vinifera* spp. *sylvestris*, which exhibits hermaphroditism in the wild population (Zhou et al., 2019). The greater size of the fruit bearing population, along with other traits (for example, increased sweetness), resulted in increased cultivation, allowing for domestication to progress into stage 3, defined by geographic expansion and adaptation to new environments. Introgression between *V. sativa* and *V. sylvestris* species (stages 1 and 2) and the prevalence of subsequent admixture between varieties of *V. sativa* in stages 3 and 4 have resulted in tremendous genetic diversity for domesticated grapevine, evident in the ~6 000 cultivars available for commercial production (This et al., 2006). It must be noted that this increase happened over the last 500 years as European nations started traveling to the East and West. Continued introgression with wild *V. vinifera* spp. from Asia and North America have contributed to the introduction of new alleles, with the purposeful breeding (stage 4) for new/sought after traits increasing in demand (Terral et al., 2010; This et al., 2006; Zhou et al., 2017b).

An early domestication focus on berry aroma has likely shaped the *VviTPS* landscape of cultivars resulting in the expansion of TPS-a and -b subfamilies associated with sesqui- and monoterpene biosynthesis, respectively (Martin et al., 2010). By consciously selecting for aroma it is likely that the terpenes associated with defence/stress responses were unconsciously affected.

Stellenbosch University https://scholar.sun.ac.za

When aromatic traits are coupled with selection for hermaphroditic flowers, one can assume that terpene biosynthesis was altered.

1.3.4 The hermaphroditic grapevine flower: genome complexity and terpene biosynthesis

Irrespective of how or when hermaphroditism took place during domestication, parts of sex-specific organelles remain to form the new inflorescence. In this structure the calyptra (flower cap) is the equivalent of flower petals and enclose the stamens (male) and pistil (female) (Keller, 2010). It is therefore reasonable to assume that the flower contains two sub-organelle specific sources for terpene biosynthesis, illustrated in Figure 1.6. Some insight into this has been gained from Cabernet Sauvignon flowers which are dominated by sesquiterpene volatiles (Martin et al., 2009). The major sesquiterpene emitted is (+)-valencene with its catalytic enzyme VvVal localising to the pollen grains of the flower. It is believed that VvVal transcripts accumulate in this sub-organelle until the calyptra falls off, resulting in the rapid biosynthesis/release of terpene volatiles. The temporal pattern observed in volatiles emissions would suggest that flower sesquiterpene biosynthesis is tightly regulated. It must be emphasised that these insights are from a single study on greenhouse grown, inflorescences (Martin et al., 2009).

The grapevine gene atlas supports the observation of *VviTPS* transcript accumulation in the flowers, with transcript abundance being the greatest in pollen tissue (Fasoli et al., 2012). The limited number of grapevine flowers analysed have shown a sesquiterpene dominated emission profile, however, Muscat Blanc, has been the exception with monoterpene emissions exceeding that of the sesquiterpenes (Buchbauer et al., 1994a, 1994b, 1995; Martin et al., 2009; Matarese et al., 2014). At the time, limited genomic resources were not amenable for exploring structural variations that could impact *VviTPS* function. The genetic structure of domesticated grapevine is therefore an important factor that needs further consideration when looking at the molecular mechanisms involved in terpene biosynthesis.

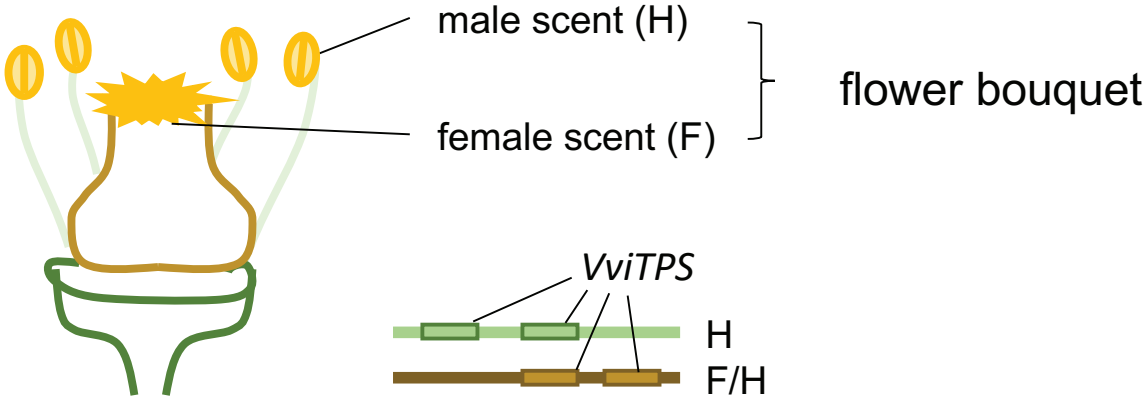

Figure 1.6 Model for grapevine flower haplotypes and terpene biosynthesis.

The PN40024 genome revealed that the *VviTPS* family has undergone significant expansion (Martin et al., 2010). Gene expansion results in a complex genome structure that can be difficult to

Stellenbosch University https://scholar.sun.ac.za

analyse due to paralogy (Morrell et al., 2012; Wang et al., 2013). This is reflected in the PN40024 genome where ~94% of the genome is paralogous (Jaillon et al., 2007). Within the *VviTPS* family, a large cluster of 45 genes, situated on chr. 18, all encode for sesquiterpene synthases, illustrating the extent of paralogy. However, nearly a third of the *VviTPS* family is mapped to chr. 00, adding great difficulties when attempting to study the genome structure of this family (Canaguier et al., 2017; Jaillon et al., 2007; Martin et al., 2010). Even if the reference genome was completely resolved onto chromosomes, it will still effectively only represent one haplotype due to it being near-homozygous. Diploid grapevine genomes therefore hold great promise for studying the structure of the *VviTPS* family.

Nearly all commercial *V. vinifera* cultivars are hermaphroditic with bifurcation of the traditional male and female sub-organelles in the hermaphroditic flower structure presenting two haplotype scenarios. In Figure 1.6, the H and F alleles refer to the parental origin of *VviTPS* alleles. If the cultivar is homozygous for hermaphroditism (H) then both haplotypes will have the H allele. If heterozygous then the haplotypes will be presented by a female (F) and H allele. The haplotype genetic structure is, however, not as straight forward as represented in the figure due the different domestication routes that could have resulted in the haplotypes.

Genome-wide association studies (GWAS) have been applied to *Vitis* germplasms in an attempt to ascertain how cultivars are related. Myles et al. (2011) provide a good representation of first order relationships of common grape cultivars. The cultivar Traminer, based on the SNP analysis of 1009 germplasm accessions, is thought to be a major progenitor of commercial cultivars which includes Pinot Noir and Sauvignon Blanc. The latter cultivar and Cabernet Franc are the parents to Cabernet Sauvignon. The diploid genome for Cabernet Sauvignon and the reference genome is therefore only separated by two first-degree relationships (Chin et al., 2016; Jaillon et al., 2007; Myles et al., 2011). Unless the parents are sequenced, the haplotype origins cannot be definitively ascertained. Nevertheless, sequencing of diploid genomes will provide insights into the fundamental structural differences of haplotypes that have been lacking in the reference genome. In isolation, this is not majorly beneficial to understanding the *VviTPS* family, however, when these structural differences are compared between genomes, one should be able to predict which *VviTPS* genes contribute to a specific phenotype.

1.3.5 Conclusion

Early introgression between the European species *V. vinifera* spp. *sativa* and *sylvestris* provided the initial genetic background that shaped the *VviTPS* family, with the selection for hermaphroditic flowers having a significant, albeit unconscious impact. Subsequent cultivation practices, further introgressions, geographic expansion and the independent adaptation of varieties likely resulted in the generation of new alleles associated with aroma. The expansion of the *VviTPS-a, -b* and *-g* subfamilies suggests that mono and sesquiterpene metabolites were traits continuously selected for due to the pleasant aromas they impart. Compounding to this genetic variability is vegetative

Stellenbosch University https://scholar.sun.ac.za

propagation which increases the frequency of mutations. The domestication history and modern-day cultivation practices are therefore continually shaping the *VviTPS* family of any given genotype under cultivation. The ecophysiological and commercial importance of terpenes will benefit greatly from the various genomic resources that are becoming available, specifically when considering how the *VviTPS* family differs between genotypes and the possible factors that determine functional variation. The hermaphroditic nature of grapevine flowers in particular provides an interesting genetic background for studying both biosynthetic and functional differences between genotypes. Understanding how and why terpene biosynthesis is different between genotypes can provide important fundamental insights into the cultivar-specific genetic potential to synthesise terpenes.

1.4 Project Outline and Objectives

To address the aims of the study, the following objectives were identified:

i. Identify the key volatiles involved in *V. vinifera* floral emissions;
ii. Isolate and characterise candidate *VviTPS* genes potentially involved in flower terpene biosynthesis;
iii. Computationally characterise (i.e. annotate) *VviTPS* genes on the diploid genome of Cabernet Sauvignon, specifically focussing on possible structural differences to the reference genome.

The research results are presented in three chapters (Chapters 2-4), each with more specific literature to complement what is already provided in the literature overview of Chapter 1. Chapter 2 has been peer-reviewed and published in Frontiers in Plant Science. Chapter 3 has been submitted to Frontiers in Plant Science and is currently awaiting review. Chapter 4 will be submitted for publication once Chapter 3 is accepted. We conclude this thesis with an integrated summary and future prospects in Chapter 5.

1.5 References

Adam-Blondon, A. F. (2014). Grapevine genome update and beyond. *Acta Hortic.* 1046, 311–318. doi:10.17660/ActaHortic.2014.1046.42.

Akatsuka, T., Kodama, O., Sekido, H., Kono, Y., and Takeuchi, S. (1985). Novel phytoalexins (oryzalexins A, B and C) isolated from rice blast leaves infected with Pyricularia oryzae. Part I: Isolation, characterization and biological activities of oryzalexins. *Agric. Biol. Chem.* 49, 1689–1694. doi:10.1271/bbb1961.49.1689.

Allaby, R. G. (2014). “Domestication Syndrome in Plants,” in *Encyclopedia of Global Archaeology*, ed. C. Smith (New York, NY: Springer New York), 2182–2184. doi:10.1007/978-1-4419-0465-2_2416.

Battilana, J., Emanuelli, F., Gambino, G., Gribaudo, I., Gasperi, F., Boss, P. K., et al. (2011). Functional effect of grapevine 1-deoxy-D-xylulose 5-phosphate synthase substitution K284N on Muscat flavour formation. *J. Exp. Bot.* 62, 5497–5508. doi:10.1093/jxb/err231.

Bennett, R. N., and Wallsgrove, R. M. (1994). Secondary metabolites in plant defence mechanisms. *New Phytol.* 127, 617–633.

Bloch, K., Chaykin, S., Phillips, A. ., and De Waard, A. (1959). Mevalonic acid pyrophospahte and isopentenylpyrphosphate. *J. Biol. Chemsitry* 234, 2595–2604.

Bohlmann, J., Crock, J., Jetter, R., and Croteau, R. (1998a). Terpenoid-based defenses in conifers: cDNA cloning, characterization, and functional expression of wound-inducible (E)-alpha-bisabolene synthase from grand fir (Abies grandis). *Proc. Natl. Acad. Sci. U. S. A.* 95, 6756–61. doi:10.1073/pnas.95.12.6756.

Stellenbosch University https://scholar.sun.ac.za

Bohlmann, J., Meyer-Gauen, G., and Croteau, R. (1998b). Plant terpenoid synthases: Molecular biology and phylogenetic analysis. *Proc. Natl. Acad. Sci. U. S. A.* 95, 4126–33. doi:10.1073/pnas.95.8.4126.

Bönisch, F., Frotscher, J., Stanitzek, S., Rühl, E., Wüst, M., Bitz, O., et al. (2014). A UDP-glucose:monoterpenol glucosyltransferase adds to the chemical diversity of the grapevine metabolome. *Plant Physiol.* 165, 561–581. doi:10.1104/pp.113.232470.

Boss, P. (2011). Vines to wine – linking fruit quality to wine flavour and aroma. Glen Osmond.

Buchbauer, G., Jirovetz, L., Wasicky, M., Herlitschka, A., and Nikiforov, A. (1994a). Aroma von WeiBweinblüten: Korrelation sensorischer Daten mit Headspace-Inhaltsstoffen. *Zeitschrift für Leb. Und-forsch.* 199, 1–4. doi:http://sci-hub.tw/10.1007/BF01192941.

Buchbauer, G., Jirovetz, L., Wasicky, M., and Nikiforov, A. (1994b). Headspace Analysis of Vitis vinifera (Vitaceae) Flowers. *J. Essent. Oil Res.* 6, 311–314. doi:10.1080/10412905.1994.9698383.

Buchbauer, G., Jirovetzl, L., Wasieky, M., and Nikiforov, A. (1995). Aroma von Rotweinblüten: Korrelation sensorischer Daten mit Headspace-Inhaltsstoffen. *Zeitschrift für Leb. Und-forsch.* 200, 443–446.

Buckingham, J., Cooper, C. M., and Purchase, R. (2015). *Natural Products Desk Reference*. Boca Raton: CRC Press.

Buckler, E. S., Holland, J. B., Bradbury, P. J., Acharya, C. B., Brown, P. J., Browne, C., et al. (2009). The Genetic Architecture of Maize Flowering Time. *Science (80-.).* 325, 714–718. doi:10.1126/science.1174276.

Burger, A. L., and Botha, F. C. (2004). Ripening-related gene expression during fruit ripening in Vitis vinifera L. cv. Cabernet Sauvignon and Clairette blanche. *Vitis - J. Grapevine Res.* 43, 59–64.

Canaguier, A., Grimplet, J., Gaspero, G. Di, Scalabrin, S., Duchêne, E., and Choisne, N. (2017). A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3). *Genomics Data* 14, 56–62. doi:10.1016/J.GDATA.2017.09.002.

Canaguier, A., Le Clainche, I., Bérard, A., Chauveau, A., Vernerey, M. S., Guichard, C., et al. (2014). Toward the Deciphering of Chromosome Structure in Vitis vinifera. *Acta Hortic.*, 319–327. doi:10.17660/ActaHortic.2014.1046.43.

Cane, D. E. (1999). “Isoprenoid Biosynthesis: Overview,” in *Comprehensive Natural Product Chemistry. Isoprenoids Including Carotenoids and Steroids, Vol 2.*, ed. D. E. Cane (Amsterdam: Elsevier), 1–12.

Cane, D. E., and Iyengar, R. (1979). The enzymic conversion of farnesyl to nerolidyl pyrophosphate: Role of the pyrophosphate moiety. *J. Am. Chem. Soc.* 101, 3385–3388. doi:10.1021/ja00506a047.

Chen, F., Tholl, D., Bohlmann, J., and Pichersky, E. (2011). The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* 66, 212–29. doi:10.1111/j.1365-313X.2011.04520.x.

Chen, H., Köllner, T. G., Li, G., Wei, G., Chen, X., Zeng, D., et al. (2019). Combinatorial Evolution of a Terpene Synthase Gene Cluster Explains Terpene Variations in Oryza. *Plant Physiol.*, pp.00948.2019. doi:10.1104/pp.19.00948.

Chin, C., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* 13, 1050–1054. doi:10.1038/nmeth.4035.

Christianson, D. W. (2006). Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev.* 106, 3412–3442. doi:10.1021/cr050286w.

Christianson, D. W. (2008). Unearthing the roots of the terpenome. *Curr. Opin. Chem. Biol.* 12, 141–150. doi:10.1016/j.cbpa.2007.12.008.

Christianson, D. W. (2017). Structural and chemical biology of terpenoid cyclases. *Chem. Rev.* 117. doi:10.1021/acs.chemrev.7b00287.

Cubry, P., Tranchant-Dubreuil, C., Thuillet, A. C., Monat, C., Ndjiondjop, M. N., Labadie, K., et al. (2018). The Rise and Fall of African Rice Cultivation Revealed by Analysis of 246 New Genomes. *Curr. Biol.* 28, 2274-2282.e6. doi:10.1016/j.cub.2018.05.066.

D’Onofrio, C., Matarese, F., Scalabrelli, G., and Boss, P. K. (2014). Functional characterization of terpene synthases of “Aromatic” and “Non-Aromatic” grapevine cultivars. *Acta Hortic.* 1046, 557–564. doi:10.1016/j.anireprosci.2007.05.009.

Da Silva, C., Zamperin, G., Ferrarini, A., Minio, A., Dal Molin, A., Venturini, L., et al. (2013). The High Polyphenol Content of Grapevine Cultivar Tannat Berries Is Conferred Primarily by Genes That Are Not Shared with the Reference Genome. *Plant Cell* 25, 4777–4788. doi:10.1105/tpc.113.118810.

Dalla Costa, L., Emanuelli, F., Trenti, M., Moreno-Sanz, P., Lorenzi, S., Coller, E., et al. (2018). Induction of Terpene Biosynthesis in Berries of Microvine Transformed with VvDXS1 Alleles. *Front. Plant Sci.* 8, 1–14. doi:10.3389/fpls.2017.02244.

Degenhardt, J., Köllner, T. G., and Gershenzon, J. (2009). Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* 70, 1621–1637. doi:10.1016/j.phytochem.2009.07.030.

Di Genova, A., Almeida, A. M., Muñoz-Espinoza, C., Vizoso, P., Travisany, D., Moraga, C., et al. (2014). Whole genome comparison between table and wine grapes reveals a comprehensive catalog of

Stellenbosch University https://scholar.sun.ac.za

structural variants. *BMC Plant Biol.* 14. doi:10.1186/1471-2229-14-7.
Doligez, a., Audiot, E., Baumes, R., and This, P. (2006). QTLs for muscat flavor and monoterpenic odorant content in grapevine (Vitis vinifera L.). *Mol. Breed.* 18, 109–125. doi:10.1007/s11032-006-9016-3.
Drew, D. P., Andersen, T. B., Sweetman, C., Møller, B. L., Ford, C., and Simonsen, H. T. (2015). Two key polymorphisms in a newly discovered allele of the Vitis vinifera TPS24 gene are responsible for the production of the rotundone precursor α-guaiene. *J. Exp. Bot.* 67, 799–808. doi:10.1093/jxb/erv491.
Dudareva, N., Martin, D., Kish, C. M., Kolosova, N., Gorenstein, N., Fäldt, J., et al. (2003). (E)-beta-ocimene and myrcene synthase genes of floral scent biosynthesis in snapdragon: function and expression of three terpene synthase genes of a new terpene synthase subfamily. *Plant Cell* 15, 1227–1241. doi:10.1105/tpc.011015.
Durairaj, J., Di Girolamo, A., Bouwmeester, H. J., de Ridder, D., Beekwilder, J., and van Dijk, A. D. (2019). An analysis of characterized plant sesquiterpene synthases. *Phytochemistry* 158, 157–165. doi:10.1016/j.phytochem.2018.10.020.
Eisenreich, W., Rohdich, F., and Bacher, A. (2001). Deoxyxylulose phosphate pathway to terpenoids. *Trends Plant Sci.* 6, 78–84.
Emanuelli, F., Battilana, J., Costantini, L., Le Cunff, L., Boursiquot, J.-M., This, P., et al. (2010). A candidate gene association study on muscat flavor in grapevine (Vitis vinifera L.). *BMC Plant Biol.* 10, 241. doi:10.1186/1471-2229-10-241.
Fasoli, M., Dal Santo, S., Zenoni, S., Tornielli, G. B., Farina, L., Zamboni, A., et al. (2012). The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. *Plant Cell* 24, 3489–505. doi:10.1105/tpc.112.100230.
Figueroa-Balderas, R., Minio, A., Morales-Cruz, A., Vondras, A. M., and Cantu, D. (2019). “Strategies for Sequencing and Assembling Grapevine Genomes,” in *The Grape Genome*, eds. D. Cantu and M. A. Walker (Cham: Springer International Publishing), 77–88. doi:10.1007/978-3-030-18601-2_5.
Francis, F., Lognay, G., and Haubruge, E. (2004). Olfactory responses to aphid and host plant volatile releases: (E)-β-farnesene an effective kairomone for the predator Adalia bipunctata. *J. Chem. Ecol.* 30, 741–755. doi:10.1023/B:JOEC.0000028429.13413.a2.
Fuller, D. Q., Asouti, E., and Purugganan, M. D. (2012). Cultivation as slow evolutionary entanglement: Comparative data on rate and sequence of domestication. *Veg. Hist. Archaeobot.* 21, 131–145. doi:10.1007/s00334-011-0329-8.
Fuller, D. Q., Denham, T., Arroyo-Kalin, M., Lucas, L., Stevens, C. J., Qin, L., et al. (2014). Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. *Proc. Natl. Acad. Sci. U. S. A.* 111, 6147–6152. doi:10.1073/pnas.1308937110.
Gao, Y., Honzatko, R. B., and Peters, R. J. (2012). Terpenoid synthase structures: a so far incomplete view of complex catalysis. *Nat. Prod. Rep.* 29, 1153–75. doi:10.1039/c2np20059g.
Gaut, B. S., Seymour, D. K., Liu, Q., and Zhou, Y. (2018). Demography and its effects on genomic variation in crop domestication. *Nat. Plants* 4, 512–520. doi:10.1038/s41477-018-0210-1.
Gholami, M., Hayasaka, Y., Coombe, B. G., Jackson, J. F., Robinson, S. P., and Williams, P. J. (1995). Biosynthesis of flavour compounds in Muscat Gordo Blanco grape berries. *Aust. J. Grape Wine Res.* 1, 19–24. doi:10.1111/j.1755-0238.1995.tb00073.x.
Girollet, N., Rubio, B., and Bert, P.-F. (2019). De novo phased assembly of the Vitis riparia grape genome. *Sci. Data* 6, 1–8. doi:10.1038/s41597-019-0133-3.
Gonzalez, V., Touchet, S., Grundy, D. J., Faraldos, J. A., and Allemann, R. K. (2014). Evolutionary and mechanistic insights from the reconstruction of α-humulene synthases from a modern (+)-germacrene A synthase. *J. Am. Chem. Soc.* 136, 14505–14512. doi:10.1021/ja5066366.
Greenhagen, B. T., O’Maille, P. E., Noel, J. P., and Chappell, J. (2006). Identifying and manipulating structural determinates linking catalytic specificities in terpene synthases. *Proc. Natl. Acad. Sci. U. S. A.* 103, 9826–9831. doi:10.1073/pnas.0601605103.
Han, Y., Gao, S., Muegge, K., Zhang, W., and Zhou, B. (2015). Advanced applications of RNA sequencing and challenges. *Bioinform. Biol. Insights* 9, 29–46. doi:10.4137/BBI.S28991.
Heil, M. (2009). Damaged-self recognition in plant herbivore defence. *Trends Plant Sci.* 14, 356–63. doi:10.1016/j.tplants.2009.04.002.
Höfer, R., Dong, L., André, F., Ginglinger, J. F., Lugan, R., Gavira, C., et al. (2013). Geraniol hydroxylase and hydroxygeraniol oxidase activities of the CYP76 family of cytochrome P450 enzymes and potential for engineering the early steps of the (seco)iridoid pathway. *Metab. Eng.* 20, 221–232. doi:10.1016/j.ymben.2013.08.001.
Huang, A.-C., Burrett, S., Sefton, M. A., and Taylor, D. K. (2014). Production of the pepper aroma compound, (-)-rotundone, by aerial oxidation of α-guaiene. *J. Agric. Food Chem* 62, 10809–10815. doi:10.1021/jf504693e.
Hutchison, C. A. (2007). DNA sequencing: Bench to bedside and beyond. *Nucleic Acids Res.* 35, 6227–6237. doi:10.1093/nar/gkm688.
Ilc, T., Arista, G., Tavares, R., Navrot, N., Duchêne, E., Velt, A., et al. (2018). Annotation, classification,

Stellenbosch University https://scholar.sun.ac.za

genomic organization and expression of the Vitis vinifera CYPome. *PLoS One* 13, 1–22. doi:10.1371/journal.pone.0199902.
Ilc, T., Parage, C., Boachon, B., Navrot, N., and Werck-Reichhart, D. (2016a). Monoterpenol Oxidative Metabolism: Role in Plant Adaptation and Potential Applications. *Front. Plant Sci.* 7, 1–16. doi:10.3389/fpls.2016.00509.
Ilc, T., Werck-Reichhart, D., and Navrot, N. (2016b). Meta-analysis of the core aroma components of grape and wine aroma. *Front. Plant Sci.* 7, 1–15. doi:10.3389/fpls.2016.01472.
Isegawa, M., Maeda, S., Tantillo, D. J., and Morokuma, K. (2014). Predicting pathways for terpene formation from first principles – routes to known and new sesquiterpenes. *Chem. Sci.* 5, 1555. doi:10.1039/c3sc53293c.
Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449, 463–7. doi:10.1038/nature06148.
Johns, T. (1986). Detoxification function of geophagy and domestication of the potato. *J. Chem. Ecol.* 12, 635–646. doi:10.1007/BF01012098.
Keller, M. (2010). “Botany and Anatomy,” in *The Science of Grapevines: Anatomy and Physiology* (Elsevier), 1–47. doi:10.1016/B978-0-12-374881-2.00001-5.
King, A., and Dickinson, J. R. (2000). Biotransformation of monoterpene alcohols by Saccharomyces cerevisiae, Torulaspora delbrueckii and Kluyveromyces lactis. *Yeast* 16, 499–506. doi:10.1002/(SICI)1097-0061(200004)16:6<499::AID-YEA548>3.0.CO;2-E.
Köksal, M., Hu, H., Coates, R. M., Peters, R. J., and Christianson, D. W. (2011a). Structure and mechanism of the diterpene cyclase ent-copalyl diphosphate synthase. *Nat. Chem. Biol.* 7, 431–433. doi:10.1038/nchembio.578.
Köksal, M., Jin, Y., Coates, R. M., Croteau, R., and Christianson, D. W. (2011b). Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature* 469, 116–122. doi:10.1038/nature09628.
Köllner, T. G., Schnee, C., Gershenzon, J., and Degenhardt, J. (2004). The variability of sesquiterpenes emitted from two Zea mays cultivars is controlled by allelic variation of two terpene synthase genes encoding stereoselective multiple product enzymes. *Plant Cell* 16, 1115–1131. doi:10.1105/tpc.019877.tive.
Lesburg, C. A. (1997). Crystal Structure of Pentalenene Synthase: Mechanistic Insights on Terpenoid Cyclization Reactions in Biology. *Science (80-.).* 277, 1820–1824. doi:10.1126/science.277.5333.1820.
Li, R., Tee, C.-S., Jiang, Y.-L., Jiang, X.-Y., Venkatesh, P. N., Sarojam, R., et al. (2015). A terpenoid phytoalexin plays a role in basal defense of Nicotiana benthamiana against Potato virus X. *Sci. Rep.* 5, 9682. doi:10.1038/srep09682.
Lichtenthaler, H. K. (1999). The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50, 47–65. doi:10.1146/annurev.arplant.50.1.47.
Lindig-Cisneros, R., Benrey, B., and Espinosa-García, F. J. (1997). Phytoalexins, resistance traits, and domestication status in Phaseolus coccineus and Phaseolus lunatus. *J. Chem. Ecol.* 23, 1997–2011. doi:10.1023/B:JOEC.0000006485.38713.8c.
Lindig-Cisneros, R., Dirzo, R., and Espinosa-García, F. J. (2002). Effects of domestication and agronomic selection on phytoalexin antifungal defense in Phaseolus beans. *Ecol. Res.* 17, 315–321. doi:10.1046/j.1440-1703.2002.00491.x.
Lücker, J., Bowen, P., and Bohlmann, J. (2004). Vitis vinifera terpenoid cyclases: functional identification of two sesquiterpene synthase cDNAs encoding (+)-valencene synthase and (−)-germacrene D synthase and expression of mono- and sesquiterpene synthases in grapevine flowers and berries. *Phytochemistry* 65, 2649–2659. doi:10.1016/j.phytochem.2004.08.017.
Maicas, S., and Mateo, J. J. (2005). Hydrolysis of terpenyl glycosides in grape juice and other fruit juices: A review. *Appl. Microbiol. Biotechnol.* 67, 322–335. doi:10.1007/s00253-004-1806-0.
Martin, D., Aubourg, S., Schouwey, M., Daviet, L., Schalk, M., Toub, O., et al. (2010). Functional Annotation, Genome Organization and Phylogeny of the Grapevine (Vitis vinifera) Terpene Synthase Gene Family Based on Genome Assembly, FLcDNA Cloning, and Enzyme Assays. *BMC Plant Biol.* 10, 226. doi:10.1186/1471-2229-10-226.
Martin, D. M., and Bohlmann, J. (2004). Identification of Vitis vinifera (-)-α-terpineol synthase by in silico screening of full-length cDNA ESTs and functional characterization of recombinant terpene synthase. *Phytochemistry* 65, 1223–1229. doi:10.1016/j.phytochem.2004.03.018.
Martin, D. M., Chiang, A., Lund, S. T., and Bohlmann, J. (2012). Biosynthesis of wine aroma: transcript profiles of hydroxymethylbutenyl diphosphate reductase, geranyl diphosphate synthase, and linalool/nerolidol synthase parallel monoterpenol glycoside accumulation in Gewürztraminer grapes. *Planta* 236, 919–29. doi:10.1007/s00425-012-1704-0.
Martin, D. M., Toub, O., Chiang, A., Lo, B. C., Ohse, S., Lund, S. T., et al. (2009). The bouquet of grapevine (Vitis vinifera L. cv. Cabernet Sauvignon) flowers arises from the biosynthesis of sesquiterpene

Stellenbosch University https://scholar.sun.ac.za

volatiles in pollen grains. *Proc. Natl. Acad. Sci. U. S. A.* 106, 7245–7250. doi:10.1073/pnas.0901387106.
Matarese, F., Cuzzola, A., Scalabrelli, G., and D'Onofrio, C. (2014). Expression of terpene synthase genes associated with the formation of volatiles in different organs of Vitis vinifera. *Phytochemistry*. doi:10.1016/j.phytochem.2014.06.007.
Matarese, F., Scalabrelli, G., and D'Onofrio, C. (2013). Analysis of the expression of terpene synthase genes in relation to aroma content in two aromatic Vitis vinifera varieties. *Funct. Plant Biol.* 40, 552. doi:10.1071/FP12326.
Matus, J. T., Ruggieri, V., Romero, F. J., Moretto, M., and Wong, D. C. J. (2019). "Status and Prospects of Systems Biology in Grapevine Research," in *The Grape Genome*, eds. D. Cantu and M. A. Walker (Cham: Springer International Publishing), 137–166. doi:10.1007/978-3-030-18601-2_8.
McGarvey, D. J., and Croteau, R. (1995). Terpenoid metabolism. *Plant Cell* 7, 1015–1026. doi:10.1105/tpc.7.7.1015.
McGovern, P., Jalabadze, M., Batiuk, S., Callahan, M. P., Smith, K. E., Hall, G. R., et al. (2017). Early Neolithic wine of Georgia in the South Caucasus. *Proc. Natl. Acad. Sci. U. S. A.* 114, E10309–E10318. doi:10.1073/pnas.1714728114.
Metzker, M. L. (2005). Emerging technologies in DNA sequencing. *Genome Res.* 15, 1767–1776. doi:10.1101/gr.3770505.
Meyer, R. S., Choi, J. Y., Sanches, M., Plessis, A., Flowers, J. M., Amas, J., et al. (2016). Domestication history and geographical adaptation inferred from a SNP map of African rice. *Nat. Genet.* 48, 1083–1088. doi:10.1038/ng.3633.
Meyer, R. S., DuVal, A. E., and Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196, 29–48. doi:10.1111/j.1469-8137.2012.04253.x.
Migicovsky, Z., Sawler, J., Gardner, K. M., Aradhya, M. K., Prins, B. H., Schwaninger, H. R., et al. (2017). Patterns of genomic and phenomic diversity in wine and table grapes. *Hortic. Res.* 4, 1–11. doi:10.1038/hortres.2017.35.
Miller, B., Madilao, L. L., Ralph, S., and Bohlmann, J. (2005). Insect-induced conifer defense. White pine weevil and methyl jasmonate induce traumatic resinosis, de novo formed volatile emissions, and accumulation of terpenoid synthase and putative octadecanoid pathway transcripts in Sitka spruce. *Plant Physiol.* 137, 369–382. doi:10.1104/pp.104.050187.
Miller, D. J., Gao, J., Truhlar, D. G., Young, N. J., Gonzalez, V., and Allemann, R. K. (2008). Stereochemistry of eudesmane cation formation during catalysis by aristolochene synthase from Penicillium roqueforti. *Org. Biomol. Chem.* 6, 2346–2354. doi:10.1039/b804198a.
Minio, A., Lin, J., Gaut, B. S., and Cantu, D. (2017). How Single Molecule Real-Time Sequencing and Haplotype Phasing Have Enabled Reference-Grade Diploid Genome Assembly of Wine Grapes. *Front. Plant Sci.* 8, 1–6. doi:10.3389/fpls.2017.00826.
Minio, A., Massonnet, M., Figueroa-Balderas, R., Castro, A., and Cantu, D. (2019). Diploid Genome Assembly of the Wine Grape Carménère. *G3: Genes|Genomes|Genetics* 9, g3.400030.2019. doi:10.1534/g3.119.400030.
Moore, B. D., Andrew, R. L., Kulheim, C., and Foley, W. J. (2014). Explaining intraspecific diversity in plant secondary metabolites in an ecological context. *New Phytol.* 201, 733–750.
Morrell, P. L., Buckler, E. S., and Ross-Ibarra, J. (2012). Crop genomics: Advances and applications. *Nat. Rev. Genet.* 13, 85–96. doi:10.1038/nrg3097.
Myles, S., Boyko, A. R., Owens, C. L., Brown, P. J., Grassi, F., Aradhya, M. K., et al. (2011). Genetic structure and domestication history of the grape. *Proc. Natl. Acad. Sci.* 108, 3530–3535. doi:doi.org/10.1073/pnas.1009363108.
Nagegowda, D. A. (2010). Plant volatile terpenoid metabolism: Biosynthetic genes, transcriptional regulation and subcellular compartmentation. *FEBS Lett.* 584, 2965–2973. doi:10.1016/j.febslet.2010.05.045.
Naim, M., Zehavi, U., and Rouseff, R. L. (1999). Formation of ␣ -terpineol in Citrus Juices , Model and Buffer Solutions. 64, 838–841.
Osbourn, A. E., and Lanzotti, V. (2009). *Plant-derived Natural Products*. , eds. A. E. Osbourn and V. Lanzotti New York, NY: Springer US doi:10.1007/978-0-387-85498-4.
Otomo, K., Kanno, Y., Motegi, A., Kenmoku, H., Yamane, H., Mitsuhashi, W., et al. (2004). Diterpene cyclases responsible for the biosynthesis of phytoalexins, momilactones A, B, and oryzalexins A-F in rice. *Biosci. Biotechnol. Biochem.* 68, 2001–6. doi:10.1271/bbb.68.2001.
Pardo, E., Rico, J., Gil, J. V., and Orejas, M. (2015). De novo production of six key grape aroma monoterpenes by a geraniol synthase-engineered S. cerevisiae wine strain. *Microb. Cell Fact.* 14, 136. doi:10.1186/s12934-015-0306-5.
Pazouki, L., and Niinemets, Ü. (2016). Multi-Substrate Terpene Synthases: Their Occurrence and Physiological Significance. *Front. Plant Sci.* 7, 1–16. doi:10.3389/fpls.2016.01019.
Peters, R. J., and Croteau, R. B. (2002a). Abietadiene synthase catalysis: Conserved residues involved in

Stellenbosch University https://scholar.sun.ac.za

protonation-initiated cyclization of geranylgeranyl diphosphate to (+)-copalyl diphosphate. *Biochemistry* 41, 1836–1842. doi:10.1021/bi011879d.

Peters, R. J., and Croteau, R. B. (2002b). Abietadiene synthase catalysis: Mutational analysis of a prenyl diphosphate ionization-initiated cyclization and rearrangement. *Proc. Natl. Acad. Sci. U. S. A.* 99, 580–584. doi:10.1073/pnas.022627099.

Peters, R. J., Flory, J. E., Jetter, R., Ravn, M. M., Lee, H. J., Coates, R. M., et al. (2000). Abietadiene synthase from grand fir (Abies grandis): Characterization and mechanism of action of the "pseudomature" recombinant enzyme. *Biochemistry* 39, 15592–15602. doi:10.1021/bi001997l.

Pichersky, E., Noel, J. P., and Dudareva, N. (2006). Biosynthesis of plant volatiles: nature's diversity and ingenuity. *Science* 311, 808–811. doi:10.1126/science.1118510.

Piechulla, B., Bartelt, R., Brosemann, A., Effmert, U., Bouwmeester, H., Hippauf, F., et al. (2016). The α-Terpineol to 1,8-Cineole Cyclization Reaction of Tobacco Terpene Synthases. *Plant Physiol.* 172, 2120–2131. doi:10.1104/pp.16.01378.

Purugganan, M. D. (2019). Evolutionary Insights into the Nature of Plant Domestication. *Curr. Biol.* 29, R705–R714. doi:10.1016/j.cub.2019.05.053.

Purugganan, M. D., and Fuller, D. Q. (2011). Archaeological data reveal slow rates of evolution during plant domestication. *Evolution (N. Y).* 65, 171–183. doi:10.1111/j.1558-5646.2010.01093.x.

Ravn, M. M., Peters, R. J., Coates, R. M., and Croteau, R. (2002). Mechanism of abietadiene synthase catalysis: Stereochemistry and stabilization of the cryptic pimarenyl carbocation intermediates. *J. Am. Chem. Soc.* 124, 6998–7006. doi:10.1021/ja017734b.

Rhoads, A., and Au, K. F. (2015). PacBio Sequencing and Its Applications. *Genomics, Proteomics Bioinforma.* 13, 278–289. doi:10.1016/j.gpb.2015.08.002.

Riaz, S., De Lorenzis, G., Velasco, D., Koehmstedt, A., Maghradze, D., Bobokashvili, Z., et al. (2018). Genetic diversity analysis of cultivated and wild grapevine (Vitis vinifera L.) accessions around the Mediterranean basin and Central Asia. *BMC Plant Biol.* 18, 1–14. doi:10.1186/s12870-018-1351-0.

Rivas, F., Parra, A., Martinez, A., and Garcia-Granados, A. (2013). Enzymatic glycosylation of terpenoids. *Phytochem. Rev.* 12, 327–339. doi:10.1007/s11101-013-9301-9.

Roach, M. J., Johnson, D. L., Bohlmann, J., van Vuuren, H. J. J., Jones, S. J. M., Pretorius, I. S., et al. (2018). Population sequencing reveals clonal diversity and ancestral inbreeding in the grapevine cultivar Chardonnay. *PLoS Genet.*, 1–24. doi:https://doi.org/10.1371/journal.pgen.1007807.

Rohmer, M. (1999). The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat. Prod. Rep.* 16, 565–574. doi:10.1039/a709175c.

Schnee, C., Kollner, T. G., Gershenzon, J., Degenhardt, J., Ko, T. G., Gershenzon, J., et al. (2002). The maize gene terpene synthase 1 encodes a sesquiterpene synthase catalyzing the formation of (E)-á-farnesene, (E)-nerolidol, and (E,E)-farnesol after herbivore damage. *Plant Physiol.* 130, 2049–2060. doi:10.1104/pp.008326.the.

Schwab, W., Fischer, T. C., Giri, A., and Wüst, M. (2015). Potential applications of glucosyltransferases in terpene glucoside production: impacts on the use of aroma and fragrance. *Appl. Microbiol. Biotechnol.* 99, 165–174. doi:10.1007/s00253-014-6229-y.

Schwab, W., Williams, D. C., Davis, E. M., and Croteau, R. (2001). Mechanism of monoterpene cyclization: Stereochemical aspects of the transformation of noncyclizable substrate analogs by recombinant (-)-limonene synthase, (+)-bornyl diphosphate synthase, and (-)-pinene synthase. *Arch. Biochem. Biophys.* 392, 123–136. doi:10.1006/abbi.2001.2442.

Schwab, W., and Wüst, M. (2015). Understanding the constitutive and induced biosynthesis of mono- and sesquiterpenes in grapes (Vitis vinifera) – A key to unlocking the biochemical secrets of unique grape aroma profiles. *J. Agric. Food Chem.* 63, 10591–10603. doi:10.1021/acs.jafc.5b04398.

Sharon-Asa, L., Shalit, M., Frydman, A., Bar, E., Holland, D., Or, E., et al. (2003). Citrus fruit flavor and aroma biosynthesis: Isolation, functional characterization, and developmental regulation of Cstps1, a key gene in the production of the sesquiterpene aroma compound valencene. *Plant J.* 36, 664–674. doi:10.1046/j.1365-313X.2003.01910.x.

Siebert, T. E., Wood, C., Elsey, G. M., and Pollnitz, A. P. (2008). Determination of rotundone, the pepper aroma impact compound, in grapes and wine. *J. Agric. Food Chem.* 56, 3745–8. doi:10.1021/jf800184t.

Sohrabi, R., Huh, J.-H., Badieyan, S., Rakotondraibe, L. H., Kliebenstein, D. J., Sobrado, P., et al. (2015). In planta variation of volatile biosynthesis: an alternative biosynthetic route to the formation of the pathogen-induced volatile homoterpene DMNT via triterpene degradation in Arabidopsis roots. *Plant Cell Online* 27, 874–890. doi:10.1105/tpc.114.132209.

Sun, L., Zhu, B. Q., Sun, X. R., Xu, X. Q., Zhang, G. J., Yan, A. L., et al. (2015). Monoterpene accumulation and its biosynthesis: Gene transcript profiles of two grape cultivars during berry development. *Acta Hortic.* 1082, 37–42. doi:10.17660/ActaHortic.2015.1082.2.

Sun, T., and Kamiya, Y. (1994). The Arabidopsis GA1 locus encodes the cyclase ent-kaurene synthetase A of gibberellin biosynthesis. *Plant Cell* 6, 1509–1518. doi:10.1105/tpc.6.10.1509.

Takase, H., Sasaki, K., Shinmori, H., Shinohara, A., Mochizuki, C., Kobayashi, H., et al. (2015). Cytochrome

Stellenbosch University https://scholar.sun.ac.za

P450 CYP71BE5 in grapevine (Vitis vinifera) catalyzes the formation of the spicy aroma compound. *J. Exp. Bot.* 67, 787–798. doi:10.1093/jxb/erv496.

Tanno, K., and Wilcox, G. (2006). How Fast Was Wild Wheat Domesticated? *Science (80-.).* 311, 1886–1886. doi:10.1126/science.1124635.

Terral, J. F., Tabard, E., Bouby, L., Ivorra, S., Pastor, T., Figueiral, I., et al. (2010). Evolution and history of grapevine (Vitis vinifera) under domestication: new morphometric perspectives to understand seed domestication syndrome and reveal origins of ancient European cultivars. *Ann. Bot.* 105, 443–455. doi:10.1093/aob/mcp298.

The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* 408, 796–815. doi:10.1038/35048692.

This, P., Lacombe, T., and Thomas, M. R. (2006). Historical origins and genetic diversity of wine grapes. *Trends Genet.* 22, 511–519. doi:10.1016/j.tig.2006.07.008.

Trapp, S. C., and Croteau, R. B. (2001). Genomic organization of plant terpene synthases and molecular evolutionary implications. *Genetics* 158, 811–832.

Vandermoten, S., Mescher, M. C., Francis, F., Haubruge, E., and Verheggen, F. J. (2012). Aphid alarm pheromone: An overview of current knowledge on biosynthesis and functions. *Insect Biochem. Mol. Biol.* 42, 155–163. doi:10.1016/j.ibmb.2011.11.008.

Velasco, R., Zharkikh, A., Troggio, M., Cartwright, D. a, Cestaro, A., Pruss, D., et al. (2007). A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS One*, e1326. doi:10.1371/journal.pone.0001326.

Venter, M., Burger, A. L., and Botha, F. C. (2001). Molecular analysis of fruit ripening: The identification of differentially expressed sequences in Vitis vinifera using cDNA-AFLP technology. *Vitis* 40, 191–196.

Venturini, L., Ferrarini, A., Zenoni, S., Tornielli, G. B., Fasoli, M., Dal Santo, S., et al. (2013). De novo transcriptome characterization of Vitis vinifera cv. Corvina unveils varietal diversity. *BMC Genomics* 14, 41. doi:10.1186/1471-2164-14-41.

Vondras, A. M., Minio, A., Blanco-Ulate, B., Figueroa-Balderas, R., Penn, M. A., Zhou, Y., et al. (2019). The genomic diversification of grapevine clones. *BMC Genomics* 20, 972. doi:10.1186/s12864-019-6211-2.

Vuylsteke, M., Peleman, J. D., and van Eijk, M. J. T. (2007). AFLP-based transcript profiling (cDNA-AFLP) for genome-wide expression analysis. *Nat. Protoc.* 2, 1399–1413. doi:10.1038/nprot.2007.174.

Walker, M. A., Heinitz, C., Riaz, S., and Uretsky, J. (2019). “Grape Taxonomy and Germplasm,” in *The Grape Genome*, eds. D. Cantu and M. A. Walker (Cham: Springer International Publishing), 25–38. doi:10.1007/978-3-030-18601-2_2.

Wan, Y., Schwaninger, H. R., Baldo, A. M., Labate, J. A., Zhong, G. Y., and Simon, C. J. (2013). A phylogenetic analysis of the grape genus (Vitis L.) reveals broad reticulation and concurrent diversification during neogene and quaternary climate change. *BMC Evol. Biol.* 13, 1. doi:10.1186/1471-2148-13-141.

Wang, N., Xiang, Y., Fang, L., Wang, Y., Xin, H., and Li, S. (2013). Patterns of Gene Duplication and Their Contribution to Expansion of Gene Families in Grapevine. *Plant Mol. Biol. Report.* 31, 852–861. doi:10.1007/s11105-013-0556-5.

Waters, D. L. E., Holton, T. A., Ablett, E. M., Lee, L. S., and Henry, R. J. (2005). cDNA microarray analysis of developing grape (Vitis vinifera cv. Shiraz) berry skin. *Funct. Integr. Genomics* 5, 40–58. doi:10.1007/s10142-004-0124-z.

Wink, M. (1988). Plant breeding: importance of plant secondary metabolites for protection against pathogens and herbivores. *Theor. Appl. Genet.* 75, 225–233. doi:10.1007/BF00303957.

Wood, C., Siebert, T. E., Parker, M., Capone, D. L., Elsey, G. M., Pollnitz, A. P., et al. (2008). From wine to pepper: Rotundone, an obscure sesquiterpene, is a potent spicy aroma compound. *J. Agric. Food Chem.* 56, 3738–3744. doi:10.1021/jf800183k.

Wymore, T., Chen, B. Y., Nicholas, H. B., Ropelewski, A. J., and Brooks, C. L. (2011). A mechanism for evolving novel plant sesquiterpene synthase function. *Mol. Inform.* 30, 896–906. doi:10.1002/minf.201100087.

Yamaguchi, S., Smith, M. W., Brown, R. G. S., Kamiya, Y., and Sun Tai-ping (1998). Phytochrome regulation and differential expression of gibberellin 3β-hydroxylase genes in germinating Arabidopsis seeds. *Plant Cell* 10, 2115–2126. doi:10.1105/tpc.10.12.2115.

Yauk, Y.-K., Ged, C., Wang, M. Y., Matich, A. J., Tessarotto, L., Cooney, J. M., et al. (2014). Manipulation of flavour and aroma compound sequestration and release using a glycosyltransferase with specificity for terpene alcohols. *Plant J.* 80, 317–330. doi:10.1111/tpj.12634.

Zhou, Y., Massonnet, M., Sanjak, J. S., Cantu, D., and Gaut, B. S. (2017a). Evolutionary genomics of grape (Vitis vinifera ssp. vinifera) domestication. *Proc. Natl. Acad. Sci. U. S. A.* 114, 11715–11720. doi:10.1073/pnas.1709257114.

Zhou, Y., Massonnet, M., Sanjak, J. S., Cantu, D., and Gaut, B. S. (2017b). Evolutionary genomics of grape (Vitis vinifera ssp. vinifera) domestication. *Proc. Natl. Acad. Sci.* 114, 201709257. doi:10.1073/pnas.1709257114.

Stellenbosch University https://scholar.sun.ac.za

Zhou, Y., Muyle, A., and Gaut, B. S. (2019). “Evolutionary Genomics and the Domestication of Grapes,” in *The Grape Genome*, eds. D. Cantu and M. A. Walker (Cham: Springer International Publishing), 39–55. doi:10.1007/978-3-030-18601-2_3.

Stellenbosch University https://scholar.sun.ac.za

Chapter 2

Linking Terpene Synthases to Sesquiterpene Metabolism in Grapevine Flowers

This chapter was published in Frontiers in Plant Science. Associated supplementary material is available online, as listed in the publication.

Stellenbosch University https://scholar.sun.ac.za

frontiers in Plant Science

ORIGINAL RESEARCH
published: 21 February 2019
doi: 10.3389/fpls.2019.00177

Linking Terpene Synthases to Sesquiterpene Metabolism in Grapevine Flowers

*Samuel Jacobus Smit, Melané Alethea Vivier and Philip Richard Young**

Institute for Wine Biotechnology, Department of Viticulture and Oenology, Stellenbosch University, Stellenbosch, South Africa

Grapevine (*Vitis vinifera* L.) terpene synthases (VviTPS) are responsible for the biosynthesis of terpenic volatiles. Volatile profiling of nine commercial wine cultivars showed unique cultivar-specific variation in volatile terpenes emitted from grapevine flowers. The flower chemotypes of three divergent cultivars, Muscat of Alexandria, Sauvignon Blanc and Shiraz were subsequently investigated at two flower developmental stages (EL-18 and -26). The cultivars displayed unique flower sesquiterpene compositions that changed during flower organogenesis and the profiles were dominated by either (*E*)-β-farnesene, (*E,E*)-α-farnesene or (+)-valencene. *In silico* remapping of microarray probes to *VviTPS* gene models allowed for a meta-analysis of *VviTPS* expression patterns in the grape gene atlas to identify genes that could regulate terpene biosynthesis in flowers. Selected sesquiterpene synthase genes were isolated and functionally characterized in three cultivars. Genotypic differences that could be linked to the function of a targeted gene model resulted in the isolation of a novel and cultivar-specific single product sesquiterpene synthase from Muscat of Alexandria flowers (VvivMATPS10), synthesizing (*E*)-β-farnesene as its major volatile. Furthermore, we identified structural variations (SNPs, InDels and splice variations) in the characterized *VviTPS* genes that potentially impact enzyme function and/or volatile sesquiterpene production in a cultivar-specific manner.

Keywords: TPS, grapevine, chemotype, flower, sesquiterpene

OPEN ACCESS

Edited by:
Deyu Xie,
North Carolina State University, United States

Reviewed by:
Guodong Wang,
Institute of Genetics and Developmental Biology (CAS), China
Massuo Jorge Kato,
University of São Paulo, Brazil

***Correspondence:**
Philip Richard Young
pryoung@sun.ac.za

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

Received: *07 November 2018*
Accepted: *05 February 2019*
Published: *21 February 2019*

Citation:
Smit SJ, Vivier MA and Young PR (2019) Linking Terpene Synthases to Sesquiterpene Metabolism in Grapevine Flowers. Front. Plant Sci. 10:177. doi: 10.3389/fpls.2019.00177

INTRODUCTION

Evolution has resulted in tremendous chemical diversity of floral scent within and across species. Terpene synthases (TPS) are responsible for the biosynthesis of terpenoids, a class of natural products consisting of more than 50,000 compounds in plants (Christianson, 2008; Osbourn and Lanzotti, 2009; Buckingham et al., 2015), of which ∼556 are known to contribute to floral scent (Knudsen and Gershenzon, 2006). A TPS typically catalyzes the final step in terpene biosynthesis with enzymes having the capacity to synthesize either a single terpene or multiple compounds (Christianson, 2017). This is mainly due to the complex mechanism of the enzyme; one of the most significant aspects being how the enzyme's active site interacts with its substrate (Degenhardt et al., 2009). TPS substrate biosynthesis results from the head-to-tail coupling of C_5 prenylated precursors, namely isopentenyl phosphate (IPP) and dimethylallyl phosphate (DMAPP), that are synthesized by the plastidial 2*C*-methyl-D-erythritol-4-phosphate (MEP) or cytosolic mevalonate (MVA) pathway (Bloch et al., 1959; Lichtenthaler, 1999; Rohmer, 1999). Although these pathways

Stellenbosch University https://scholar.sun.ac.za

are compartmentalized, metabolic "crosstalk" has been shown to result in these precursors being transported between the plastids and cytosol (Piel et al., 1998; Adam et al., 1999; Jux et al., 2001; Bick and Lange, 2003; Hemmerlin et al., 2003; Schuhr et al., 2003). Regulation of these pathways has been shown to be spatial, temporal and/or diurnal, depending on the species and organ involved in biosynthesis (Dudareva and Pichersky, 2006). The C_{10} geranyl diphosphate (GPP) and C_{15} farnesyl diphosphate (FPP) substrates result in the biosynthesis of the majority of flower terpenes, namely mono- and sesquiterpenes, respectively (Davis and Croteau, 2000).

A TPS facilitates a complex biochemical cascade involving cyclizations and/or rearrangement of the substrate to form acyclic and cyclic terpenes. These cascades proceed through reactive intermediates, referred to as carbocations, that serve as branchpoints for specific trajectories in the chemical cascade. It is thus possible to group terpenes based on the similarity of carbocations/cascade required for biosynthesis (Allemann et al., 2007; Hare and Tantillo, 2016; Christianson, 2017). Although the crystal structures of mono- and sesquiterpene synthases have been elucidated (Lesburg, 1997; Starks, 1997; Caruthers et al., 2000; Rynkiewicz et al., 2001; Shishova et al., 2007; Gennadios et al., 2009), the exact path from substrate to terpene is not always known, or conclusively determined. Computational chemistry has proven useful in predicting these structures and the reaction mechanism that will result in terpenes under biologically relevant conditions (Allemann et al., 2007; Miller et al., 2008; Hess et al., 2011; Tantillo, 2011; Wedler et al., 2015; O'Brien et al., 2016).

Plant terpenes are typically studied for their ecological/biological roles which include pollinator attraction (Pichersky and Gershenzon, 2002), direct and indirect pest/pathogen/cellular defense (Köllner et al., 2008; Sabater-Jara et al., 2010; Zulak and Bohlmann, 2010; Lawo et al., 2011) and chemical signaling (Shen et al., 2000; Van Poecke et al., 2001; Köllner et al., 2008; Copolovici et al., 2012). Grapevine (*Vitis vinifera* L.) is a commercially important crop with an expanded *VviTPS* family consisting of 152 loci, of which 69 encode for putatively functional proteins (Martin et al., 2010). Grapevine terpenes have been mainly studied for their roles in modulating flavor and aroma profiles of grape berries and wine, with a particular focus on VviTPSs that synthesize terpenes imparting floral (e.g., linalool and limonene) and pepper (e.g., rotundone) aromas (Siebert et al., 2008; Skinkis et al., 2008; Wood et al., 2008; Matarese et al., 2013). The biological/ecological role of grapevine terpenes is, however, not well established, although a limited number of studies hold promise for identifying such roles. For example, the terpenes (*E*)-β-farnesene, (*E*)-β-caryophyllene and (*E*)-4,8-dimethyl-1,3,7-nonatriene were shown to act as semiochemicals for the phytophagous moth *Lobesia botrana*, a major pest in European vineyards (Tasin et al., 2005; Anfora et al., 2009; von Arx et al., 2011; Salvagnin et al., 2018). Also, cultivar-specific resistance toward phylloxera (*Daktulosphaira vitifoliae*) has been linked to root terpene biosynthesis (Lawo et al., 2011). A potential role in antioxidant protection in response to ultraviolet light has also been proposed for grapevine leaf terpenes (Gil et al., 2012).

Grapevine flowers show the most significant expression of *VviTPS* genes, compared to other organs in the grapevine gene atlas (Fasoli et al., 2012). A concordant emission of terpenes has been observed in a limited number of cultivars profiled for their flower volatile emissions (Buchbauer et al., 1994a,b, 1995; Martin et al., 2009; Barbagallo et al., 2014; Matarese et al., 2014). These results clearly showed that grapevine flowers have a unique transcriptional and biosynthetic capacity to produce and emit terpenes, with the majority of cultivars emitting mainly sesquiterpenes, even though the biological/ecological role(s) for domesticated grapevine flower terpenes remain to be established. Furthermore, the reported volatile profiles suggested that there are differences between cultivars, but it is difficult to directly compare the results from the different studies, given the variety of analytical techniques used to profile the grapevine flowers (Buchbauer et al., 1994a,b, 1995; Martin et al., 2009; Barbagallo et al., 2014; Matarese et al., 2014).

One of the aims of this study was to link terpenic profiles of the flower terpene emissions of a few selected, globally important commercial cultivars of grapevine, to functionally characterized *VviTPS* genes. Cultivar variations in terpene biosynthesis could be due to a variety of genetic and/or biochemical factors. To date 30 of the 69 putative *VviTPS* gene models (Martin et al., 2010) identified on the PN40024 reference genome (Jaillon et al., 2007) have been functionally characterized, of which 16 encode for sesqui- and 7 for mono-*TPS* genes. These 30 gene models are associated with 42 enzymes producing a broad range of terpenes and were isolated from a multitude of tissue types and cultivars (Lücker et al., 2004; Martin and Bohlmann, 2004; Martin et al., 2010; Drew et al., 2015). The reference genome revealed that the VviTPS family is greatly expanded, likely due to a complicated domestication history where the modern domesticated species shows greater diversity and heterozygosity than the ancient parents (Aradhya et al., 2003; Salmaso et al., 2004; Laucou et al., 2018). Crossing of distantly related parents, coupled with clonal propagation, have resulted in numerous heterozygous genotypes with their genetic diversity not reflected in the highly inbred, near homozygous reference genome (Da Silva et al., 2013; Roach et al., 2018; Minio et al., 2019). For example, a comparison between the reference genome and the Tannat cultivar revealed that 8–10% of genes are unshared, referred to as cultivar specific or "private" genes (Da Silva et al., 2013). Furthermore, these private genes contribute to cultivar specific phenotypes and account for the majority of uniquely expressed genes (Da Silva et al., 2013). More recently, the application of single cell sequencing technology revealed that the genome of Cabernet Sauvignon contains private genes not present in PN40024, Tannat, Nebbiolo or Corvina genomes (Minio et al., 2019) while a similar study in Chardonnay extended genotypic differences even further by showing the extent of structural variations within fifteen clones of this cultivar (Roach et al., 2018). Other approaches to identify structural variations between genotypes include the analyses of molecular markers, like nuclear microsatellites (nSSRs) or single nucleotide polymorphisms (SNPs), where evidence of extensive genotypic differences is shown (Aradhya et al., 2003; This et al., 2004; Ibáñez et al., 2009; Myles et al., 2011; Emanuelli et al., 2013; Picq et al., 2014; Nicolas et al., 2016; Laucou et al., 2018). A second

Stellenbosch University https://scholar.sun.ac.za

focus of this study was therefore to understand the genetic factors that could determine how cultivar genotypes differ in terms of terpene biosynthesis.

Although the PN40024 reference genome is limiting when viewing genotypic variation, it still allowed for the generation of numerous expression datasets that can be mined to understand the VviTPS family. One of the most useful datasets is that of the grapevine gene atlas (Fasoli et al., 2012) which consists of 54 different organs and tissue types, comprehensively profiling gene expression throughout the plant. Unfortunately it underrepresents the *VviTPS* family due to the microarray probe design being based on computationally identified gene models of the CRIBI.v1 genome annotation (Jaillon et al., 2007; Forcato, 2010; Adam-Blondon et al., 2011; Grimplet et al., 2012; Adam-Blondon, 2014). The 152 *VviTPS*-like loci identified by Martin et al. (2010) and resultant manually corrected *VviTPS* gene models differ greatly from the 70 *VviTPS*-like genes of the CRIBI.v1 genes analyzed on the gene atlas. Furthermore, cross-hybridization of probes on the grapevine microarrays can be extensive leading to a high false discovery rate (Moretto et al., 2016). We addressed these limitations through *in silico* remapping of the microarray probes from the gene atlas to the curated *VviTPS* gene annotations (Martin et al., 2010), allowing for the identification of specific VviTPS expression patterns. Grapevine flowers showed an interesting expression pattern with subsequent volatile profiling of flowers from nine cultivars showing terpene volatile differences. We therefore aimed to explore the extent of genotypic differences in *VviTPS* genes, and their potential impact on terpene metabolism by linking the *in silico* expression analyses with functional characterization of selected *VviTPS* gene models. Gene models were characterized in three different cultivars with gene structure variations (i.e., SNPs and InDels) that could impact enzyme function in a cultivar-specific manner evaluated. The results obtained in this study, and known VviTPS functions mentioned earlier, were used to postulate on the carbocation intermediates commonly utilized in grapevine flower sesquiterpene biosynthesis. This resulted in the generation of a model for metabolic cascades involved in grapevine flower sesquiterpene biosynthesis as dictated by cultivar-specific roles of VviTPSs.

MATERIALS AND METHODS

Sampling and Volatile Analysis of Grapevine Flower Material

Nine *V. vinifera* cultivars, namely Chardonnay (CH), Chenin Blanc (CB), Muscat of Alexandria (MA), Pinot noir (PN), Pinotage (PI), Sauvignon Blanc (SB), Shiraz (SH), Viognier (VG), and Weisser Riesling (WR), were sampled at the pre-anthesis flower stage, corresponding to stage 18 of the modified Eichorn-Lorenz (EL) phenological stage classification system (Coombe, 1995). Six to eight flower clusters per cultivar were obtained from a mother block in the Stellenbosch area (33°57′33.50″S, 18°51′38.09″E), South Africa in a vineyard where the respective cultivars were planted in close geographical proximity. Samples were flash frozen with liquid nitrogen and stored at −80°C. Flower rachises were separated from the samples before flowers were homogenized and stored at −80°C for subsequent analyses.

In a subsequent season, sampling of MA, SB and SH flowers were performed at two distinct developmental stages, the EL-18 and EL-26 (flower bloom) stages. For this sampling, we randomly sampled four biological repeats consisting of six to eight flower clusters per repeat from the same vineyard as described before. All cultivars were sampled between 9 and 10 am on a single day for the respective stages during the 2015 flower season.

A method optimized for grape berry aroma compound analysis (Young et al., 2015) was adapted to analyze flower tissue. 10 mg (±10% SD) frozen tissue was weighed off directly into a 20 mL glass vial containing 2 mL tartrate extraction buffer (5 g/L tartaric acid, 2 g/L ascorbic acid, 8 mg/L sodium azide and 250 g/L NaCl). The deuterated standard Anisole-D_8 (Sigma-Aldrich, United States), prepared in acetonitrile served as internal standard and was added to the buffer at a final concentration of 0.1 mg/L. Vials were sealed using a screw cap. Solid phase micro-extraction (SPME) of the vial head space (HS) was done using a 50/30 μm gray divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) fiber (Supelco, Bellefonte, PA, United States) that underwent pre-conditioning at 270°C for 60 min in the GC injection port according to the manufacturer specifications.

Sample vials were pre-incubated for 5 min at 45°C in the autosampler heating chamber. The heating chamber was maintained at 45°C and agitated at 250 rpm to allow for equilibration of compounds between the sample and headspace. The fiber was inserted through the septa and exposed to the analytes in the headspace for 10 min, while maintaining the agitation speed and temperature at 250 rpm and 45°C, respectively. Desorption of the analytes took place in the GC injection port for 5 min, where after, the fiber was maintained for 20 min in order to prevent any carryovers.

An Agilent 6890N gas chromatograph (Agilent, Santa Clara, CA, United States) system coupled to a CTC CombiPal Analytics auto-sampler and an Agilent 5975B inert XL EI/CI MSD mass spectrometer detector through a transfer line was used for the analyses. A Zebron 7HG-G009-11 ZB-FFAP capillary 55 column, 30 m × 250 ID μm, 0.25 μm film thickness, (Phenomenex, United States) was used. The desorption temperature for the analytes was 250°C for 5 min with a 10:1 split. Helium served as the carrier gas having an initial flow rate of 1 mL/min. Initial oven temperature was maintained for 2 min at 40°C, followed by a linear increase of 10°C/min to a final temperature of 240°C which was held for an additional 2 min.

Authentic standards for identification and quantification of volatiles were purchased from Sigma-Aldrich, United States for (+)-valencene (≥70%), (*E*)-β-farnesene (≥90%), β-caryophyllene (≥80%), and α-humulene (≥96%). Stock solutions of the standards were prepared in methanol. A calibration curve was prepared in 2 mL tartrate buffer as described above containing 0.1 mg/L Anisole-D_8 as internal standard.

The Qualitative Analysis package of MassHunter Workstation software (Agilent, Santa Clara, CA, United States) was used to visualize extracted ion chromatograms (IEC) using the

Stellenbosch University https://scholar.sun.ac.za

cumulative response of the following masses: 41 and 55 for (*E*)-2-hexenal; 70 and 116 for the internal standard; 93, 161 and 189 for sesquiterpenes. IEC chromatogram peak areas were integrated using default parameters and normalized to the area of the internal standard. Compounds were identified using authentic standards, when available, and the Wiley 275 and NIST14 mass spectral libraries. Concentrations were determined according to the calibration curve of the respective authentic standards. Where an authentic standard was not available, we determined compound concentrations semi-quantitatively using the (+)-valencene standard curve.

In silico Expression Pattern and Phylogenetic Analysis for the *VviTPS*-a Gene Family

Manual curations for the *VviTPS* gene family (Martin et al., 2010) were incorporated in the recently released 12X.v2 genome assembly and accompanying VCost.v3 (V3) annotation (Canaguier et al., 2017) and are referred to accordingly. We aimed to supplement the existing compendium of expression data (Moretto et al., 2016), generated using the CRIBI.V1 annotation, as described below.

Putatively functional *VviTPS* genes (Martin et al., 2010) were evaluated for their expression patterns in the grapevine gene atlas [GEO Accession GSE36128 (Fasoli et al., 2012)]. Probe sequences for the NimbleGen 090918 *Vitis vinifera* exp HX12 array (NCBI GEO Acc. GPL13936) were retrieved from the GEO database (Edgar et al., 2002) followed by analysis of probe binding ambiguity using BLAST homology with cut-off parameters that allowed for two sequence miss matches of the full-length probe sequence as aligned to the *VviTPS* gene models. RMA normalized expression values of the re-mapped probes were used to analyze the expression patterns with the clustermap function of the Seaborn package in Python (version 3.5.3).

VviTPS-a members (Lücker et al., 2004; Martin et al., 2009, 2010) were compared through multiple sequence alignments (MSAs) of derived protein sequences. CLC Main Workbench 7 (CLC Bio-Qiagen, Denmark) was used to perform MUSCLE alignments followed by phylogenetic tree construction using Maximum Likelihood Phylogeny with UPGMA as construction method, Jukes Cantor as substitution model and 100 bootstrap replicates.

Isolation and Characterization of *VviTPS* Genes

Total nucleic acids were extracted from the MA, SB, and SH cultivars using the method described in Reid et al. (2006). RNA was purified and gDNA removed by on-column DNase I treatment using the Bioline Isolate II Plant RNA kit (Celtic Molecular Diagnostics, South Africa). RNA integrity was assessed on an agarose gel followed by cDNA synthesis using the ImPromII Reverse Transcription System (Promega, United States). Primers were designed with restriction digestion sites to facilitate directional cloning (**Supplementary Table 1**) using predicted cDNA sequences for *VviTPS* gene models described by Martin et al. (2010), as available on FLAGdb++ (Dèrozier et al., 2011). PCR reactions were performed using Phusion High Fidelity DNA polymerase (Thermo Fisher Scientific, United States). PCR products of expected sizes were purified from an agarose gel using the Qiagen Gel Extraction kit (Qiagen, United States) and A-tailed by incubation with the TaKaRa ExTaq proof-reading polymerase. A-tailed PCR products were ligated into a pGEM-T Easy vector (Promega, United States), transformed into chemically competent *Escherichia coli* and verified through bi-directional sequencing (Central Analytical Facility, Stellenbosch University, South Africa).

Isolated genes were sequenced and named according to the grapevine nomenclature standard for *V. vinifera* L. (Vviv) (Grimplet et al., 2014) with the gene model numbers used in the VCost.v3 annotation (Canaguier et al., 2017) preceded by the cultivar abbreviations for Muscat of Alexandria (MA), Sauvignon Blanc (SB), and Shiraz (SH): *VvivMATPS01* (MK100068), *VvivSBTPS01* (MK100069), *VvivSBTPS02* (MK100070), *VvivMATPS10* (MK100071), *VvivMATPS27* (MK100072), *VvivSHTPS27* (MK100073), *VvivMATPS28* (MK100074), *VvivSHTPS01* (MK100075), *VvivMATPS02* (MK100076), *VvivSHTPS02* (MK100077), *VvivSBTPS10* (MK100078), *VvivSHTPS10* (MK100079), *VvivSBTPS27* (MK100080), *VvivSBTPS28a* (MK100081), *VvivSBTPS28b* (MK100082), *VvivSHTPS28* (MK100083). Details regarding the specific cultivar clones are included in the above GenBank accessions.

Sequence analysis of gene isolates was performed using the CLC Main Workbench 7 (CLC Bio-Qiagen, Denmark) by searching for the presence of an open reading frame (ORF). Gene structures were predicted using Splign (Kapustin et al., 2008) with genomic sequences of target gene models (retrieved from FLAGdb++) used as the reference. Gene structures were visualized using the Gene Structure Display Server (Hu et al., 2015). Derived protein sequences were used to identify the N-terminal RRx_8W and C-terminal DDxxD and NSE/DTE motifs described to be characteristic of *TPS* genes (Bohlmann et al., 1998; Aubourg et al., 2002) using the FIMO tool of the MEME suite (Bailey et al., 2009; Grant et al., 2011). The CLC Main Workbench 7 (CLC Bio-Qiagen, Denmark) was used to generate all MSAs.

In vivo Heterologous Expression of *VviTPS* Cultivar Variants in Yeast and Volatile Profiling

Sub-cloning of putatively functional *VviTPS* genes from pGem-T Easy (Promega, United States) vectors were performed through restriction enzyme excision and ligation with T4 ligase (Promega, United States) into an inducible yeast expression vector harboring a *GAL1* promoter and the *URA3* auxotrophic marker. Expression vectors were transformed into *E. coli*, followed by PCR screening for positives and subsequent plasmid isolations. Positive expression vectors were linearized with *Apa*I (Thermo Fisher Scientific, United States) and transformed into *Saccharomyces cerevisiae* strain GT051 using the TRAFO method (Gietz and Woods, 2002). The GT051 strain was modified from the Thomas and Rothstein (1989) W303a strain to

Stellenbosch University https://scholar.sun.ac.za

increase the metabolic flux for the FPP terpene precursor by over-expression of a truncated *HMG1* (SGD:S000004540) and an *IDI1* (SGD:S000006038) gene. Yeast transformants were plated on modified TRAFO synthetic drop-out (SD) plates (Gietz and Woods, 2002) containing 2% (w/v) glucose and the amino acids adenine, leucine and uracil omitted for auxotrophic selection. Putative yeast transformants were verified by colony PCR.

Synthetic complete (SC) media (Gietz and Woods, 2002) was supplemented with $MgSO_4$ to a final Mg^{2+} concentration of 5 mM and buffered to pH 6 using citrate-phosphate buffer. Pre-cultures of the respective yeast transformants were prepared in SC media with glucose (2% w/v) as a carbon source. Cells were harvested through centrifugation and washed with sterile water. *TPS*-expression was induced in sealed 20 mL GC-vials containing 5 mL SC media with galactose (2% w/v) as carbon source. Assays were performed in triplicate (three cultures per positive transformant). The starting optical density (OD) was 0.7 at 600 nm. After 16 h of induction at 30°C with shaking, vials were placed at 4°C for 1 h before analysis. A 1 mL mixture of natamycin (Delvocid at 2 mg/mL in 0.1 M NaOH) and the internal standard Anisole-D_8, prepared in acetonitrile at 50 μg/L final concentration, was added to each vial by piercing the vial septa using a sterile syringe. Delvocid was added to arrest biomass production, allowing for normalization to the internal standard.

HS-SPME-GC-MS was conducted using the same fiber, column, chromatograph and mass spectrometer detector as described before. The fiber was inserted through the septa and exposed to the analytes in the headspace for 20 min, while maintaining the agitation speed and temperature at 250 rpm and 35°C, respectively. Desorption of the analytes took place in the GC injection port where after the fiber was maintained for 20 min in order to prevent any carryovers. Desorption temperature for the analytes was 250°C for 5 min with a 10:1 split. Helium served as carrier gas with an initial flow rate of 1 mL/min. Initial oven temperature was maintained for 2 min at 40°C, followed by a linear increase of 10°C/min to a final temperature of 240°C which was held for an additional 2 min. The total run time was 24 min and the transfer line temperature 250°C. Calibration curves prepared in SC media, using the standards described earlier, were used for quantification and compound identification in combination with the Wiley 275 and NIST14 mass spectral libraries. Chromatograms were analyzed as described earlier.

Transient Expression in *Nicotiana benthamiana*

Putative *VviTPS* genes were cloned into pDONR-Zeocin, using the 2-step PCR protocol to add attB sites, followed by an overnight BP reaction as described in the product manual (Thermo Fisher Scientific, United States). Entry clones were transformed into electrocompetent *E. coli* and colonies confirmed to be positive through sequencing. Expression clones were created using the pEAQ-HT-DEST1 vector (Sainsbury et al., 2009; Peyret and Lomonossoff, 2013) by performing an overnight LR reaction, followed by transformation into *E. coli* as above and restriction enzyme digestion of plasmids to confirm positive colonies. Clonases for Gateway cloning and the pDONR-Zeocin vector were purchased from Thermo Fisher Scientific, United States.

Destination vectors were transformed into electrocompetent *Agrobacterium tumefaciens* GV3101 and plated on LB plates with 30 μg/mL gentamycin, 50 μg/mL kanamycin, and 50 μg/mL rifampicin. Transient expression and volatile analysis in *N. benthamiana* was performed according to the method described by Bach et al. (2014) with minor adaptations: Overnight cultures were washed thrice with 0.9% (w/v) saline solution and resuspended to a final OD_{600} of 0.6 using MMA buffer [10 mM 2-[*N*-morpholino]ethanesulfonic acid (MES) pH 5.6, 10 mM $MgCl_2$, 200 μM acetosyringone] instead of water. Resuspended cultures were incubated for 1 h at room temperature before infiltration. Two fully expanded leaves per plant were infiltrated in triplicate. Mock infiltrations with MMA buffer and non-infiltrated wild type plants served as controls. Qualitative analysis and compound identification were performed with the GC-MS instrument, software and (*E*)-β-farnesene analytical standard described earlier.

Southern and Northern Blot Analysis of *VviTPS10*

A DIG probe targeting *VviTPS10* was obtained through PCR amplification of an 862 bp internal region of the coding sequence followed by DIG labeled as described in the DIG Application Manual for Filter Hybridization (Roche, Germany) and diluted to 8.2 ng/mL in DIG Easy Hyb. The same probe solution was used for both Southern and Northern blotting at the appropriate temperatures described in the DIG Application Manual for Filter Hybridization (Roche, Germany).

For Southern blot analysis genomic DNA was isolated from MA, SB, and SH using the method described by Lovato et al. (2012), followed by single digests of 10 μg gDNA using *Bam*HI, *Eco*RI and *Xba*I restriction enzymes (Thermo Fisher Scientific, United States). Southern blotting was performed as described in the DIG Application Manual for Filter Hybridization (Roche, Germany).

Total RNA was isolated from ±100 mg tissue for EL-18 and EL-26 stages from MA, SB, and SH using the method described by Reid et al. (2006). RNA was selectively purified using the RNeasy Mini kit (Qiagen) according to the RNA clean-up protocol described in the product manual. RNA samples were separated on a 1.2% formaldehyde agarose (FA) gel followed by Northern blot analysis according to the DIG Application Manual for Filter Hybridization (Roche, Germany).

Biosynthetic Network of VviTPS and Heterologous Sesquiterpenes

VviTPS enzymes that have reported heterologous function (Lücker et al., 2004; Martin et al., 2009, 2010; Drew et al., 2015) along with enzymes isolated in this study were used to construct a virtual interaction network using Cytoscape (Version 3.4) (Shannon et al., 2003), available from http://www.cytoscape.org/. VviTPS enzymes were used as source nodes with their associated volatiles serving as target nodes, connected by an edge.

Stellenbosch University https://scholar.sun.ac.za

Edges were weighted as a major volatile when their percentage contribution was greater than 10% with all volatiles contributing less then 10% deemed a minor volatile. Source nodes were colored according to the likely carbocation intermediate used in the majority of volatiles from the respective enzymes. We referred to Bülow and König (2000); Davis and Croteau (2000); Tantillo (2011); Miller and Allemann (2012); Wedler et al. (2015), and Durairaj et al. (2019) to predict the likely carbocation intermediate involved.

RESULTS

In silico Expression Patterns of *VviTPS* Genes

Available *VviTPS* gene models (Martin et al., 2010) were re-assessed by re-mapping of probes-to-genes, as annotated on FLAGdb++ (Dèrozier et al., 2011), followed by expression pattern identification. The 69 putatively functional *VviTPS* gene models (predicted pseudo- and partial genes not considered for probe-to-gene remapping) (Martin et al., 2010) were re-analyzed to generate a network model (**Supplementary Data Sheet 1**). It was observed that probes often cross-hybridize with multiple *VviTPS* gene models, highlighting the close relatedness within the gene family. *In silico* remapping revealed a total number of 306 probes binding to the 69 putative *VviTPS* genes, with only 133 of these probes binding uniquely to a single gene model (**Supplementary Figures 1A,B**). Of these probes, only eight gene models showed the expected four probes per gene. The remaining probes had probe-to-gene binding ambiguity ratios between 1:2 and 1:6.

Using the remapped probe sets, *in silico* expression analysis was performed using the grapevine gene atlas. The *VviTPS* mapping provided for the NimbleGen 090918 *Vitis vinifera* exp HX12 platform can, however, also be used to view *VviTPS* expression for all experiments available in the GPL13936 platform.

Global *VviTPS* expression was assessed by looking at all the probes individually. Two expression hotspots were identified, shown in the red and blue squares of **Supplementary Figure 2**. The blue square represented organs undergoing initial differentiation from budburst (EL-14) up to inflorescence establishment (EL-17), and include probes associated with mainly the VviTPS-a and -b subfamilies. The second hotspot (red square) showed high *VviTPS* expression in flower tissues from early bloom (EL-20) to full-bloom (EL-25), with the majority of probes also associated with VviTPS-a and -b subfamilies. Gene specific patterns were subsequently calculated by averaging all probes that bind uniquely to *VviTPS*-a and -b transcripts, illustrated in **Figure 1**. Only 35 of the 49 putatively functional VviTPS-a and -b members could be considered for **Figure 1**, with the remaining members represented only by ambiguously binding probes. A differential expression pattern for male and female flower organs was observed with *VviTPS-a* members (VviTPS07, -08, -10, -12, -14, and -16) showing greater expression in male parts while significantly lower expression in female parts. VviTPS27 and -28 showed the inverse with higher relative expression in female parts. The hotspot associated with inflorescence development (EL-14 to -17) was much less pronounced when probes are averaged, while the high relative expression at flower anthesis was still evident. In combination, the two approaches (the per probe and gene-averaged expression clustermaps) showed that VviTPS-a and -b subfamilies were highly expressed in floral organs with a differential pattern between pre- and full-bloom stages, suggesting that mono- and sesquiterpene biosynthesis could be upregulated during flower organogenesis.

Profiling of Grapevine Flower Chemotypes

A selection of nine cultivars formed part of an initial screen to evaluate the formation of mono- and sesquiterpenes at flowering. Volatile analysis of flower samples at EL-18 stage of these cultivars (presumed to be the *VviTPS* transcriptional transition point from pre-bloom to bloom and including genes from hotspots identified in **Figure 1** and **Supplementary Figure 2**) revealed that the cultivars differed significantly in terms of volatile content and composition, and that the majority of compounds present were sesquiterpenes (**Supplementary Figure 3**). (*E*)-2-hexenal was present at high concentrations for all cultivars along with heptadecene, tridecanone, eicosene, and 2-pentadecanone alkanes, at low abundance. Hierarchical clustering of the sesquiterpene volatiles identified cultivar differences in the chemotypes (**Supplementary Figure 3**) and identified the volatiles driving the differentiation. Two main clusters were identified with (*E,E*)-α-farnesene, (+)-valencene and its rearrangement 7-epi-α-selinene consistently present in all cultivars, except for CH and PI which lacked the latter two and produced (*E,E*)-α-farnesene as the major volatile. SB and SH were therefore selected as white and red varieties to represent this common chemotype with MA selected due to its unique chemotype, dominated by (*E*)-β-farnesene.

In-depth profiling of these three cultivars at two phenological stages were performed to expand on the different compositional ratios observed in the initial nine cultivar screen (**Supplementary Figure 3**). We identified a total of 12 flower sesquiterpenes with seven, namely β-caryophyllene, α-humulene, (*E*)-β-farnesene, (+)-valencene, α-selinene, 7-epi-α-selinene and (*E,E*)-α-farnesene consistently present in all three cultivars, regardless of flower stage (**Supplementary Table 2**). Chromatograms illustrating the volatile differences for these three cultivars can be viewed in **Supplementary Figure 4**. (+)-Aromadendrene, β-selinene, (*E*)-β-caryophyllene and (*Z,E*)-α-farnesene were emitted at low levels in a cultivar and/or stage specific manner, as shown in **Supplementary Table 2**. Multivariate data analysis tools were applied to identify variables that explain the variation observed between cultivars. Firstly, we used unsupervised principal component analysis (PCA) of the sesquiterpene volatiles, shown in **Figures 2A,B**. Cultivar sesquiterpene composition was shown to be the main driver for differences, contributing to 74.2% in the first component while stage differences explained 18.3% of the variation as the second component. The loadings plots (**Figure 2B**) was subsequently used to identify the volatiles that impart the

Stellenbosch University https://scholar.sun.ac.za

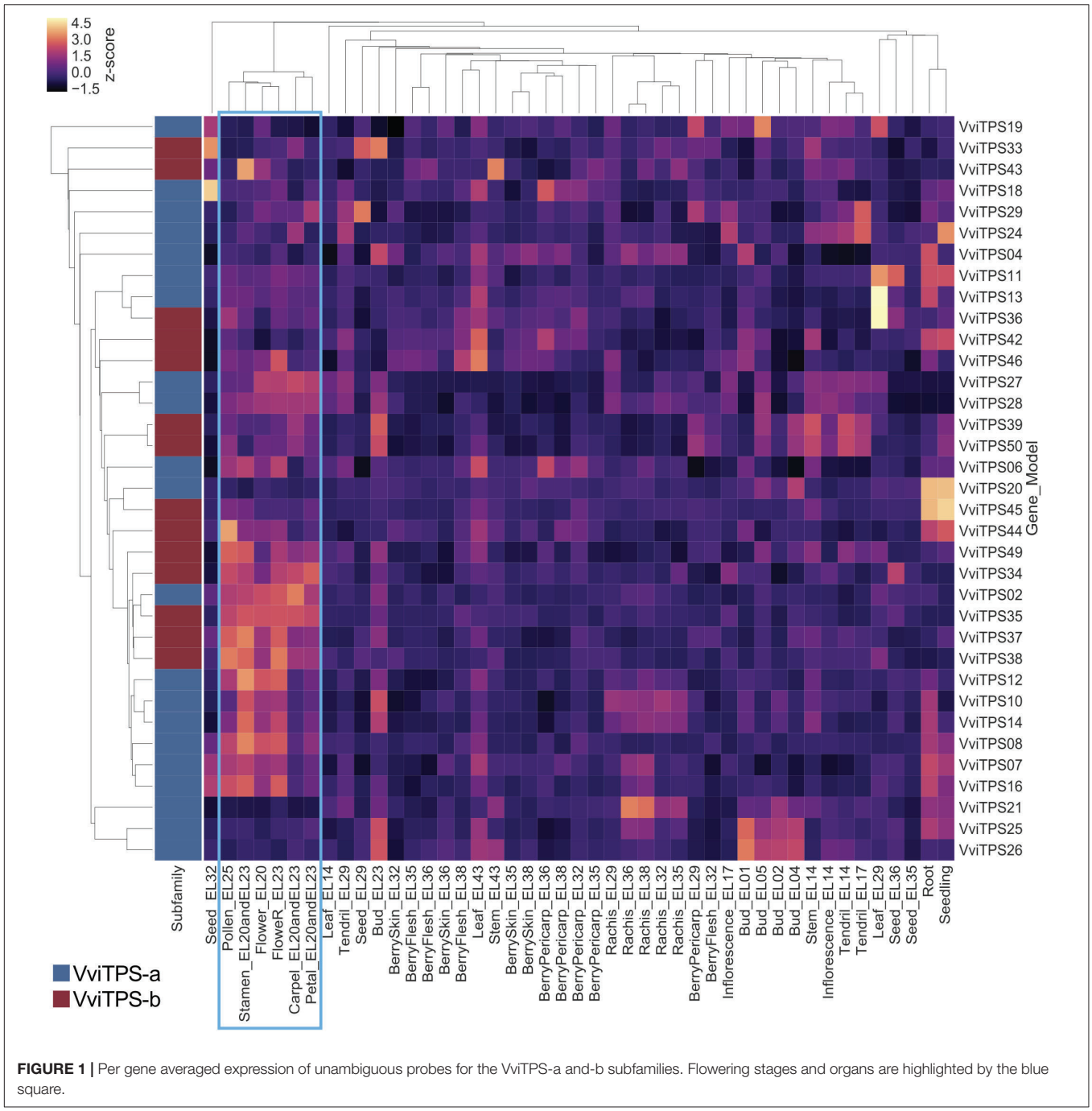

FIGURE 1 | Per gene averaged expression of unambiguous probes for the VviTPS-a and-b subfamilies. Flowering stages and organs are highlighted by the blue square.

most variation to the dataset with a multivariate analysis of variance (MANOVA) of these volatiles showing the extent of statistically significant differences between cultivars and/or stages (**Figures 2C–F**).

(*E*)-β-farnesene (**Figure 2C**) was significantly different for MA, compared to SB and SH. Furthermore, a significant difference for MA was observed between stages with 74% higher (*E*)-β-farnesene emission at EL-18 relative to EL-26. SB and SH produced (*E*)-β-farnesene at similar levels, regardless of phenological stage but at a concentration at least three times lower than MA. SB emitted (*E,E*)-α-farnesene (**Figure 2D**) as major volatile at near identical levels in both phenological stages. (*E,E*)-α-farnesene levels were significantly lower in MA and SH (ranging between 31 and 59%) relative to SB for both stages. However, SH showed an 85% relative increase for (*E,E*)-α-farnesene from EL-18 to EL-26. (+)-valencene (**Figure 2E**) and its rearrangement, 7-epi-α-selinene, (**Figure 2F**) had near identical emission levels that were statistically different between the cultivars, but not between stages within a cultivar. (+)-Valencene was the major volatile of SH. In summary, the results

Stellenbosch University https://scholar.sun.ac.za

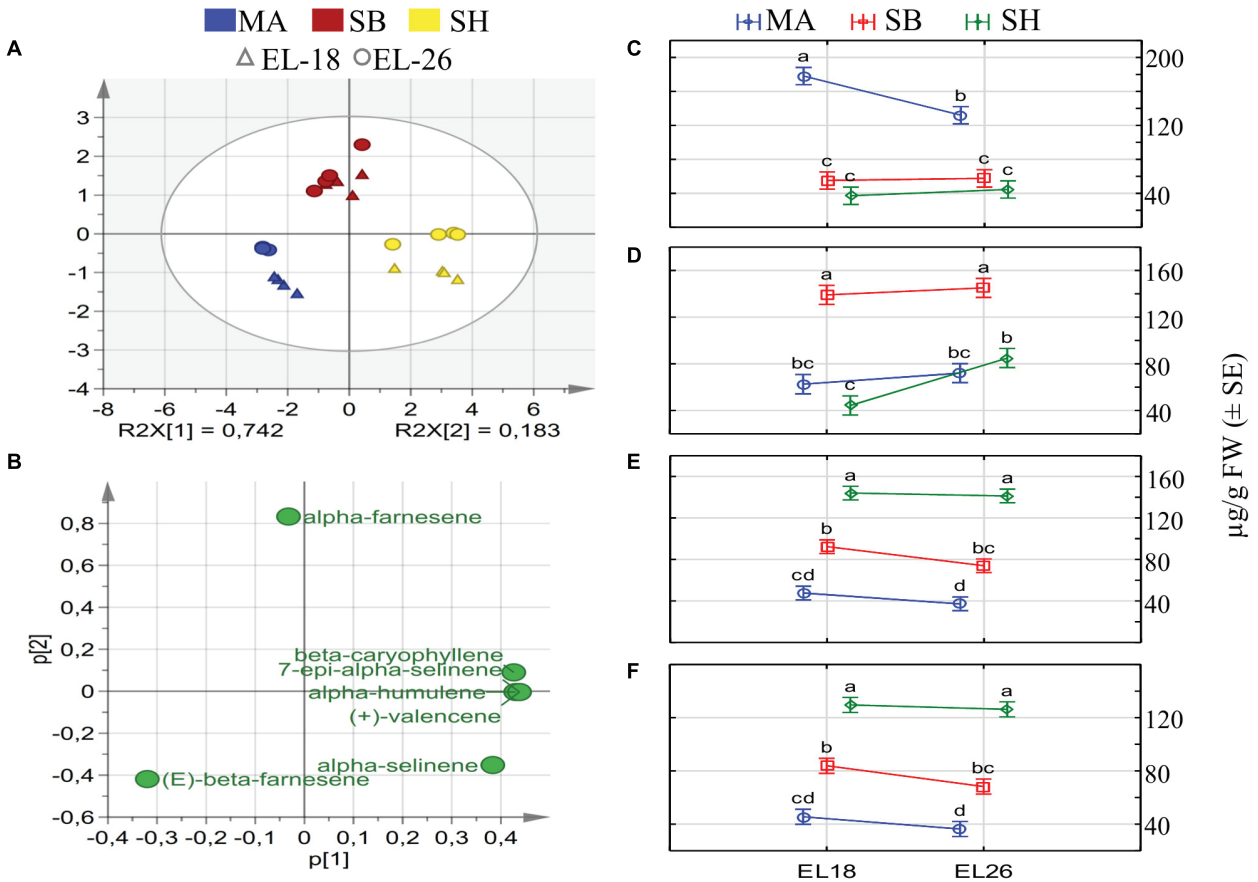

FIGURE 2 | PCA scores **(A)** and loadings **(B)** of sesquiterpenes that drive differentiation between the cultivars Muscat of Alexandria (MA), Sauvignon Blanc (SB), and Shiraz (SH) at two phenological stages (EL-18 and EL-26). MANOVA of (*E*)-β-farnesene **(C)**, (*E,E*)-α-farnesene **(D)**, (+)-valencene **(E),** and 7-epi-α-selinene **(F)** shows statistically significant differences between cultivars and phenological stages of the most abundant sesquiterpenes emitted by grapevine flowers.

showed that the three cultivars each produced a specific major sesquiterpene and that their emission compositions changed between the EL-18 and EL-26 stages. The compositional changes were minor between cultivars, and within a cultivar, as flower organogenesis progressed. However, the presence/absence for the minor sesquiterpenes (+)-aromadendrene, β-selinene, (*E*)-caryophyllene and (*Z,E*)-α-farnesene contributed significantly to the cultivar- and stage-specific chemotypes.

Selection of VviTPS-a Genes for Comparative Functional Characterization

Protein sequences derived from the predicted gene models showed subtle differences to the protein sequences of isolated and functionally characterized, illustrated in **Figure 3**. For example, five (*E*)-β-caryophyllene synthases, from two different cultivars, are associated with four different gene models (VviTPS02, -02, -13, and -27) (Martin et al., 2010), and although of similar function form distinctly different clades on the phylogenetic tree.

To investigate the extent of the genotypic variations and their potential impact on cultivar specific chemotypes we selected five candidate gene models (highlighted in blue in **Figure 3**). The *VviTPS01* gene model has been associated with two different functional sesquiterpene synthases, namely VvGwECar1 and VvGwGerA, producing (*E*)-β-caryophyllene and Germacrene A, respectively (Martin et al., 2010). This gene model also had a high number of ambiguously binding probes (**Supplementary Data Sheet 1**), suggesting that multiple variants or closely related genes exist. *In silico* expression patterns of the probes associated with *VviTPS01* furthermore show high relative expression in flowering tissue. *VviTPS02* and *VviTPS27* were dissimilar to *VviTPS01* on a sequence level but both were associated with functional enzymes, VvGwECar3/VvPnECar1 and VvGwECar2, respectively, producing (*E*)-β-caryophyllene as major product (Martin et al., 2010). *VviTPS10* was chosen due to its associated functional enzyme, VvGwaBer, producing (*E*)-β-farnesene as a minor secondary product. Twelve probes bound to this gene model, with only one binding unambiguously. Expression patterns for *VviTPS10* probes showed high relative expression in flowers. *VviTPS28* is associated with VvGerD (Lücker et al., 2004), which was characterized before the design of the microarray, resulting in four unique probes for the gene model (**Supplementary Data Sheet 1**). Furthermore, *VviTPS28*, along with *VviTPS27* showed high expression in both inflorescence and flower bloom stages (**Supplementary Figure 2**).

Stellenbosch University https://scholar.sun.ac.za

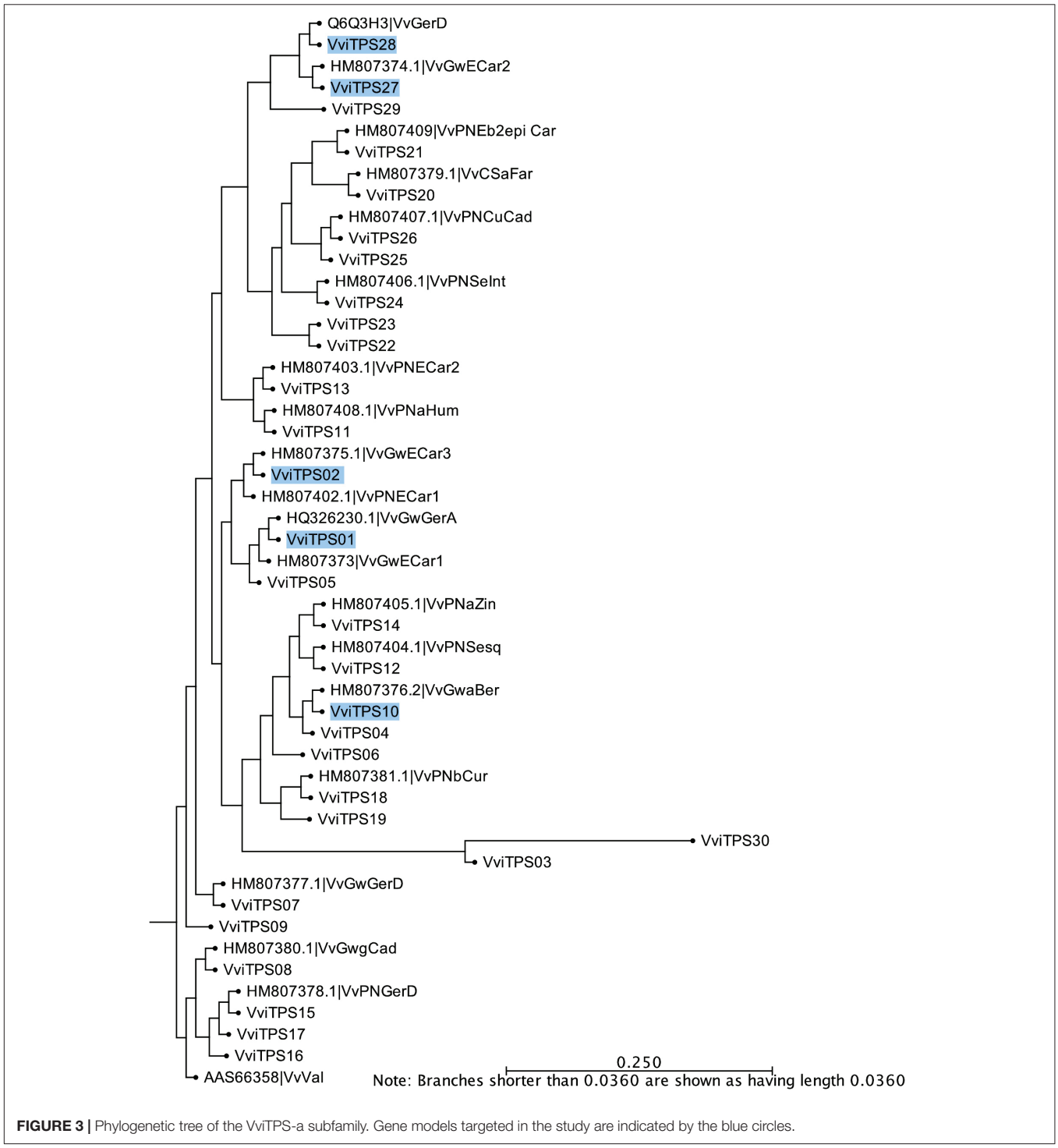

FIGURE 3 | Phylogenetic tree of the VviTPS-a subfamily. Gene models targeted in the study are indicated by the blue circles.

Analysis of Isolated VviTPS-a Gene Sequences

Sequenced isolates were compared to the predicted gene model and existing characterized genes mentioned earlier. This comparison revealed sequence and structural variations that potentially impact gene function, illustrated in **Figure 4**. *VvivTPS01 -02* isolates differed in gene structure to the gene model but contained a full length ORF and were therefore deemed putatively functional. The most prevalent cause for loss of function was due to SNPs that result in a premature stop codon. In addition to a premature stop we observed intron retention for *VvivSHTPS10*. Curiously PCR amplification with VviTPS28 primers resulted in two amplicons for SB with the second amplicon, *VvivSBTPS28b*, not being of the expected size.

Stellenbosch University https://scholar.sun.ac.za

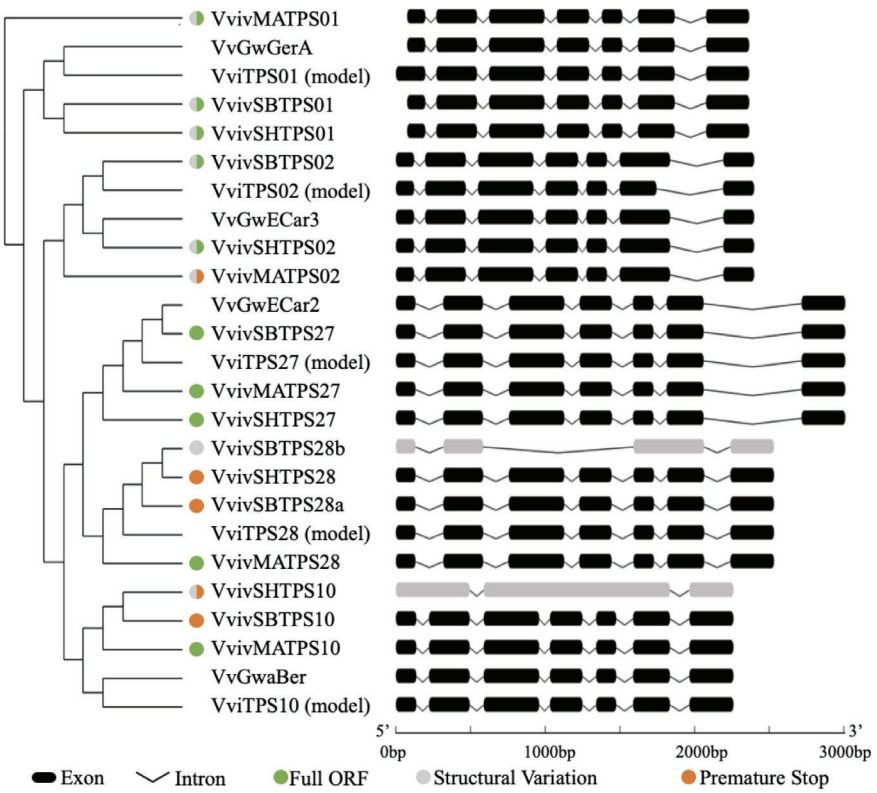

FIGURE 4 | Structural organization of isolated *VvivTPS* cultivar variants compared to the reference gene model and associated functional genes. The dendrogram reflects maximum likelihood phylogeny using MUSCLE aligned coding sequences. Structural variations to the reference sequence, sequences with premature stop codons and those with a full-length open reading frame are indicated by the colored circles.

Gene sequencing results suggest that it is a partial duplicate of *VvivSBTPS28a*. *VvivSBTPS28b* maintained exons one and two, compared to the full-length sequence of *VvivSBTPS28b*, with a 596 nucleotide deletion resulting in the loss of exons three, four and a short part of exon five which shifted the start position for exon five. The intron between exons five and six was also retained. This isolate, however, has higher sequence homology to the SH variant than the SB variant.

Protein sequences were derived for the genes with a predicted full-length ORF and compared to that of the gene model (i.e., reference sequence) and its associated functional proteins (Lücker et al., 2004; Martin et al., 2010). These results are available in **Supplementary Table 3** and the MSAs in **Supplementary Data Sheet 2**. VvivMATPS10 showed extensive sequence differences to both the reference sequence and VvGwaBer, with 37 of the 50 missense mutations located in the catalytic region of the enzyme. An amino acid deletion in the catalytic site was also observed.

Heterologous Expression and Functional Characterization of VviTPS-a Cultivar Variants

Genes with full length ORFs were expressed *in vivo* using a heterologous yeast system with the percentage contribution of the observed volatiles reported in **Table 1**. Although, putatively functional, VvivSHTPS01 and VvivSBTPS27 produced no detectable volatiles and were therefore considered non-functional *in vivo*.

Agrobacterium mediated transient expression of VvivMATPS10 confirmed functionality *in planta* as a single product enzyme synthesizing (*E*)-β-farnesene (**Supplementary Figure 5**).

Genomic Localization and Flower Expression of VviTPS10

The dominance of (*E*)-β-farnesene in MA and the unique heterologous function of VvivMATPS10 prompted further inspection. *In silico* analysis showed that the *VviTPS10* gene model shared multiple probes with other *VviTPS* genes (**Supplementary Data Sheet 1**), suggesting that genes with high homology to *VviTPS10* are present in the PN40024 genome. Southern blot analysis confirmed the presence of *VviTPS10* homologs in all three cultivars, with three separate restriction enzyme digests of gDNA from MA, SB and SH (**Supplementary Figure 6**). Multiple hybridization patterns in the 4 to 1.9 kB regions, were observed, indicating the presence of numerous homologous genes within a cultivar.

We performed a preliminary analysis of the *VviTPS10* locus using the phased-diploid assembly and annotation of Cabernet Sauvignon (Chin et al., 2016; Minio et al., 2019). Two

Stellenbosch University https://scholar.sun.ac.za

TABLE 1 | Percentage contribution of volatiles produced through *in vivo* expression of VvivTPS cultivar variants.

	VvivSB-TPS01	VvivMA-TPS01	VvivSB-TPS02	VvivMA-TPS10	VvivSH-TPS27	VvivMA-TPS27	VvivMA-TPS28
β-Elemene	6.2%	5.5%	–	–	–	–	–
(*E*)-β-Caryophyllene	–	–	**100.0%**	–	**69.1%**	**63.7%**	–
(*E*)-β-Farnesene	–	–	–	**100.0%**	–	–	–
α-Humulene	5.3%	7.4%	–	–	–	–	–
β-Selinene	–	–	–	–	22.9%	24.2%	–
γ-Selinene	6.3%	–	–	–	–	–	–
Germacrene D	–	–	–	–	2.7%	3.7%	**56.4%**
β-Selinene	16.4%	15.1%	–	–	–	–	–
α-Selinene	**38.1%**	**49.1%**	–	–	–	–	–
Camphene	–	–	–	–	–	–	12.1%
δ-Cadinene	–	–	–	–	1.1%	1.6%	17.6%
α-Amorphene	–	–	–	–	–	–	13.9%
Germacrene A	23.6%	20.0%	–	–	0.9%	1.4%	–

Major volatiles are shown in bold.

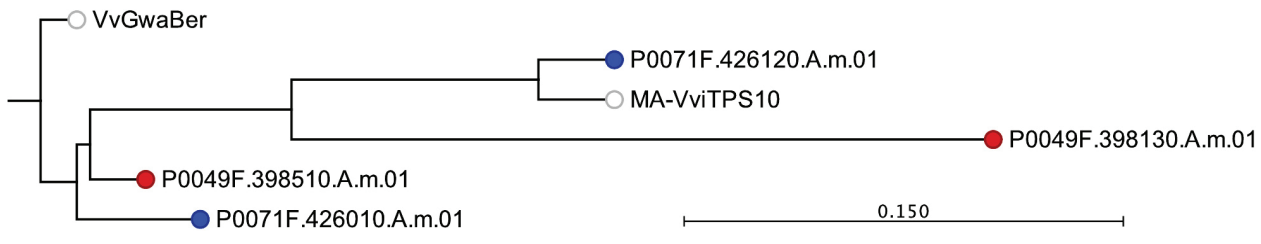

FIGURE 5 | Maximum likelihood phylogeny of proteins homologous to VvivMATPS10 and VvGWaBer identified on two different primary contigs (red and blue nodes) of the phased-diploid Cabernet Sauvignon assembly.

contigs, containing four homologous genes, were found with sequence phylogeny to *VvivMATPS10* and *VvGwaBer* and the four putative *VviTPS10*-like regions shown in **Figure 5**. The Cabernet Sauvignon *VviTPS10*-like genes are located on two different primary contigs with this shared location reflected in their phylogenetic grouping. Determining the expression of VvivMATPS10 was therefore not possible using quantitative PCR. Preliminary Northern blot analysis, however, suggested that *VviTPS10* is expressed in MA at both flower stages (**Supplementary Figure 7**).

Proposed Carbocation Cascades Involved in Flower Chemotypic Differences

By identifying a likely carbocation cascade required to synthesize flower sesquiterpenes, a cultivar-specific prevalence for carbocation intermediates was observed, illustrated in **Figure 6**. Flux through the (*E*)-humulyl cation (gray cascade) toward humulenes and caryophyllenes was consistent between stages for each cultivar. MA directs terpene biosynthesis through the farnesyl cation in both stages due to the prevalence of linear farnesene type sesquiterpenes emitted. MA, however, produced much lower total levels of sesquiterpenes in both flower stages (**Figure 6B**).

A trend of increased farnesene biosynthesis as flower development progressed was seen in all three cultivars (**Figure 6B**) with farnesene levels increasing by more than 10% from EL-18 to EL-26 in SH and SB. In these cultivars, a proportional decrease in cyclized sesquiterpenes, proceeding through the (*E,E*)-germacradienyl cation was observed. Characterized enzymes and their associated products in the context of the carbocation cascades are shown in **Figure 6A**. Based on the carbocation cascades that only utilize FPP, an enzyme-function network was created (**Figure 6C**) to illustrate the biosynthetic potential of characterized grapevine sesquiterpene synthases and how they could contribute to flower chemotypes. Nodes were numbered according to the number of gene models that transcribed an enzyme with identical function. The functional relatedness of these enzymes was represented by edges that connected products synthesized by different enzymes, with major products, <10% of contribution, shown with a thicker edge.

DISCUSSION

Grapevine Flowers Are Hotspots for *VviTPS* Expression and Terpene Production

VviTPS gene were found to be underrepresented in previous annotations. The remapping of probes to curated gene models allowed for analysis of the VviTPS family as presented in the

Stellenbosch University https://scholar.sun.ac.za

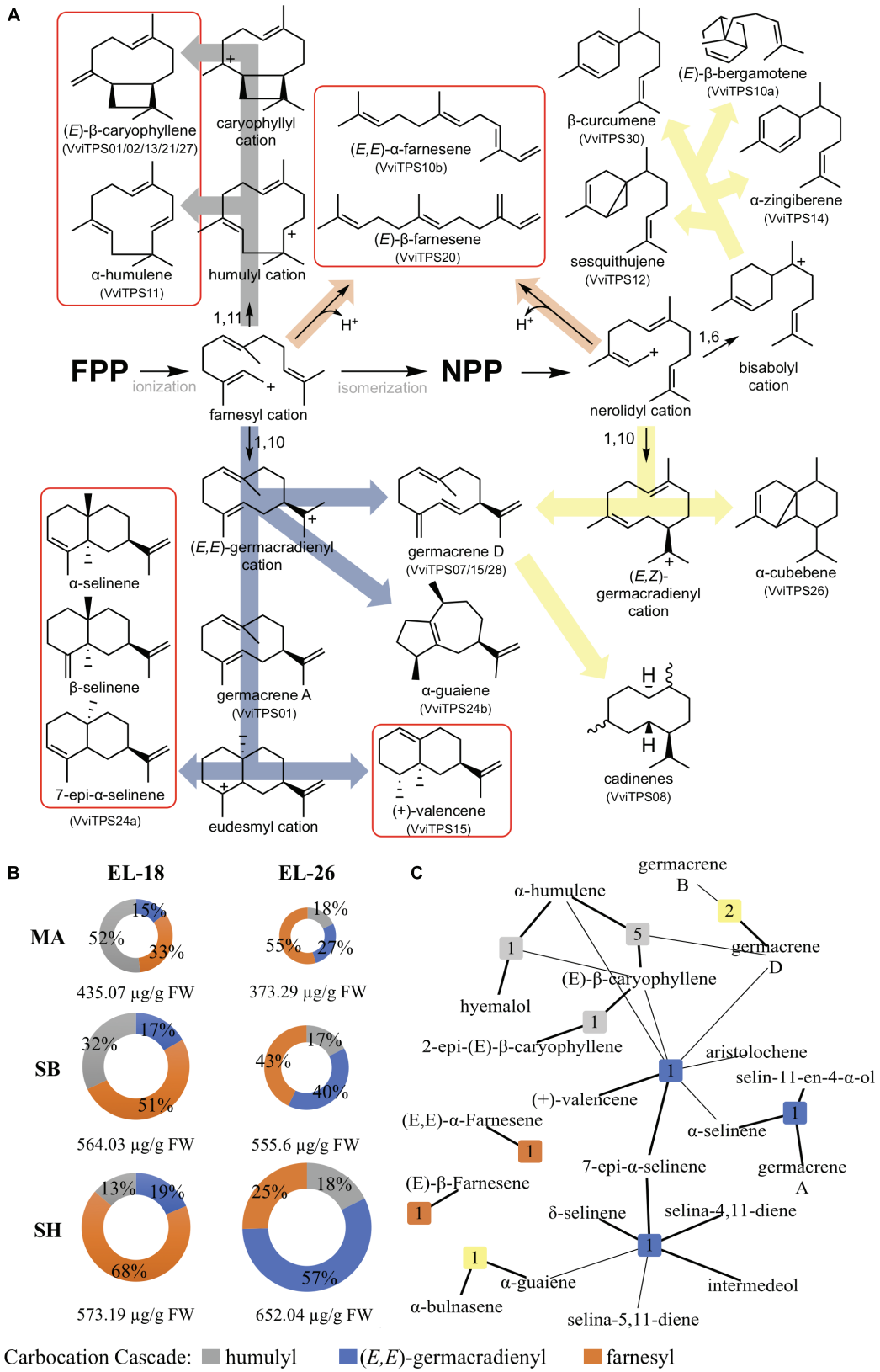

FIGURE 6 | Continued

Stellenbosch University https://scholar.sun.ac.za

FIGURE 6 | (A) Proposed carbocation cascades leading to cultivar-specific flower chemotypes proceed from the farnesyl cation toward the linear farnesenes (orange arrow) or through the humulyl (gray arrow) and (*E,E*)-germacradienyl (blue arrow) cations, respectively, with major end-point sesquiterpenes observed in this study shown in red squares. Yellow arrows indicate cascades for which grapevine has the biosynthetic potential (based on known functional enzymes) but not observed in the flowers studied. *VviTPS* gene models that have been linked to a functional enzyme synthesizing the respective sesquiterpenes are indicated in parentheses. **(B)** Biosynthetic flux as a percentage of the total observed sesquiterpenes for MA, SB and SH flower sesquiterpenes at two flower stages proceed through carbocation cascades where either farnesyl (orange), humulyl (gray) or (*E,E*)-germacradienyl (blue) cations serve as branchpoint intermediates. The total concentration of sesquiterpenes in μg/g FW is shown in the center of the doughnut charts. **(C)** Heterologously characterized sesquiterpene synthases that with products observed in the flowers show that there are numerous enzymes that contribute to specific sesquiterpenes. Products contributing to more than 10% of products synthesized in the heterologous expression assay are indicated by a thicker edge. Nodes are colored according to the dominant cascade that will result in the associated products.

VCost.v3 annotation (Canaguier et al., 2017). Although we present data here only for the grapevine gene atlas, the mapping provided can be applied for analyses on the Nimblegen 090918 Vitus HX12 platform. *In silico* expression patterns (**Figure 1**) showed that flower development and flowering were hotspots for *VviTPS* genes, with probes associated with *VviTPS-a* and *-b* transcripts (sesqui- and monoterpene synthases), showing high relative expression during the progression from inflorescence structure differentiation (EL-14) to flower bloom (EL-23) and specific transcripts localizing to these stages. It was expected to see high *VviTPS* expression in flower tissues as it was previously identified as potential organs for VviTPS biomarkers (Fasoli et al., 2012). These biomarkers were, however, based on computational gene models of the CRIBI.v1 annotation. The *in silico* expression profiles presented in **Figure 1** and **Supplementary Figure 2** therefore shows the expression patterns of corrected VviTPS gene models reported by Martin et al. (2010). Although, we could refine the number genes expressed through an *in silico* re-mapping of probes to the VviTPS gene family we still found a high number of genes could not be accurately analyzed due to the observed number of ambiguous probes. Nevertheless, we clearly showed that mono- and sesquiterpene synthases are upregulated during flowering.

Volatile profiling of flowers, however, only showed high levels of sesquiterpenes with a unique major volatile for the respective cultivars (**Figure 2**). Previous studies showed that (+)-valencene was the major terpene for flowers from red and white cultivars with only two cultivars showing slightly higher levels of β-caryophyllene (Buchbauer et al., 1994a,b, 1995; Martin et al., 2009). The only exception was that of Muscat Bianco where monoterpenes contributed to 20% of the total flower volatiles, with sesquiterpenes contributing less then 1% (Matarese et al., 2014). (*E,E*)-α-farnesene was the second highest sesquiterpene at 22.2% for Cabernet Sauvignon flowers (Martin et al., 2009) with other cultivars showing a total contribution of 2.2% or less (Buchbauer et al., 1994a,b, 1995). The three cultivars profiled in this study emitted a unique major sesquiterpene, with the blend of volatiles emitted consisting of the same compounds, but at different ratios. Furthermore, the initial volatile screen of nine cultivars (**Supplementary Figure 3**) suggests even greater chemotypic differences exist.

The lack of glandular structures in domesticated grapevine (Ma et al., 2016), an accumulation of sesquiterpene transcripts and concordant emissions in flowers (Martin et al., 2009), suggests that expression and emission are linked. Martin et al. (2009) showed sesquiterpene emissions were localized to the anthers. Localization of the VvValCS protein to lipid bodies in microspores of the pollen grain, preceded by an accumulation of *VvValCS* transcripts, suggested that sesquiterpene biosynthesis was confined to male parts of the hermaphroditic flower (Martin et al., 2009), but it is not yet clear if all cultivars synthesize sesquiterpenes in this manner. Our *in silico* analysis of the VviTPS family is to some extent in agreement with the aforementioned observation. However, the lack of substantial monoterpenes volatiles in grapevine flowers, except for Muscat Bianco, suggests that there are aspects of flower terpene metabolism that likely retain monoterpenes in a non-volatile form. This has indeed been shown to be true in grape berries where yeast and/or plant glycosidases release monoterpenes during vinification (Loscos et al., 2007; Martin et al., 2012; Yauk et al., 2014).

Isolation, Characterization, and Functional Analysis of *VvivTPS* Genes Provided Insight Into Genotypic Differences Potentially Impacting Sesquiterpene Production

Vegetative propagation and domestication of grapevine (Myles, 2013) resulted in a SNP ratio that is 2–3 higher than Arabidopsis (Martinez-Zapater et al., 2010). Profiling of these SNP differences in 5,000 germplasm accessions revealed two general domestication paths where aromatic varieties, commonly associated with table grapes, originate from Muscat or Riesling parents and less aromatic varieties used for making wine originating from the Traminer variety (Myles et al., 2011). Recently it was shown that vegetative propagation also allows for the maintenance of aberrant genome scale events where large regions of a genome can be lost due to chromosome breaking which also results widespread recombination events (Carbonell-Bejerano et al., 2017). These genome scale events have been linked to structural events that alter berry color due to deletions of hemizygous genes (Carbonell-Bejerano et al., 2017). Furthermore, evidence of genome wide transposable elements (Carrier et al., 2012), especially around the VviTPS members (Martin et al., 2010), indicate that domestication and propagation of grapevine resulted in cultivar and/or clone specific genetic changes. Linking these genotype specific structural variations with an observable phenotype presents a challenge but can be addressed to some extent when computational chemistry, functional biology, bioinformatics and chemical profiling is utilized in combination to understand enzyme mechanisms.

Stellenbosch University https://scholar.sun.ac.za

Phylogenetic similarity is thought to be an inaccurate predictor for function due to the effect that subtle amino acid changes have on TPS function (Yoshikuni et al., 2006), a fact that is exacerbated by the heterozygosity of grapevine and high level of duplications within the *VviTPS* family (Martin et al., 2010). Previous studies have used sequence phylogeny to establish the evolution of TPS in plants (Bohlmann et al., 1998). Various studies on the active site of sesquiterpene synthases, however, suggests that phylogenetic similarity in this region will allow for a more focused analysis by identifying amino acid residues that correlate with conserved enzyme mechanisms (Degenhardt et al., 2009; Wymore et al., 2011). These insights were recently applied in a sequence-based analysis of 262 experimentally characterized plant sesquiterpene synthases; resulting in the identification of conserved amino acid residues and motifs (Durairaj et al., 2019). Incorporating experimental evidence with the amino acid composition in the active side subsequently allowed for grouping enzymes based on carbocation intermediates utilized to produce the observed end-point sesquiterpenes (Durairaj et al., 2019). By studying the genotypic differences of selected sesquiterpene synthases from three cultivars we identified subtle sequence variations that could impact enzyme function. Although all cultivars produced a transcript for the targeted genes, structural variations resulted in many of these transcripts being non-functional (**Supplementary Table 3**). SNPs resulted in premature stop codons for five of the isolates (**Supplementary Table 3**) in this study with intron retention and partial duplication also shown (**Figure 4**). These structural variations effectively eliminate the targeted genes from contributing to the flower chemotype. Extrapolating these results to the extensive genotypic variation within *V. vinifera* furthermore highlights the limitations of a one-size-fits-all reference genome.

The database of plant sesquiterpene synthases (Durairaj et al., 2019) allowed us to predict the reaction mechanisms for VviTPS involved in flower sesquiterpene biosynthesis. We characterized the sequence space of the five targeted gene models by utilizing the aforementioned database in order understand how the observed genotypic variations influences enzyme function. Furthermore, we extrapolate these findings, in combination with known functional VviTPSs, to the observed flower chemotypes, illustrated in **Figure 6**. VviTPS10 served as a prime example for genotypic differences influencing flower chemotypes. The gene space of the three cultivar variants shows that the SB and SH variants of VviTPS10 contain premature stop codons with the latter also retaining some introns. This gene model was previously characterized as VvGWaBer synthase (VviTPS10a in **Figure 6A**), isolated from Gewürztraminer, producing bergamotene as major product (Martin et al., 2010). The SB and SH non-functional VviTPS10 variants showed high homology with this gene. The MA variant was, however, unique in both sequence and function with *in vivo* and *in planta* characterization resulting in (*E*)-β-farnesene as a single product. No sesquiterpene volatiles that will require isomerisation of FPP to NPP were observed. Nevertheless, a genetic capacity to synthesize NPP derived products is present in grapevine, shown by the yellow cascade of **Figure 6A**.

Observed flower volatiles can be grouped based on the carbocation intermediates required for their production. This allowed us to identify the cultivar specific flux from FPP with known grapevine sesquiterpene synthases producing these volatiles indicated in the cascades (**Figures 6A,B**). Cyclization of FPP was observed to be that first branch point with the majority of known grapevine sesquiterpene synthases proceeding though either 1,11 or 1,10 ring closures. Based on the observed flower sesquiterpenes we showed that 14.7–18.6% of FPP is directed through a 1,11-closure (gray cascade) toward humulenes and caryophyllenes with seven gene models linked with enzymes that perform this as a primary mechanism (**Figure 6C**). Cyclization resulting in the (*E,E*)-germacradienyl cation will be required to account for the majority of sesquiterpenes observed in SB and SH (**Figure 6B**), with VviTPS24 and -15 characterized to produce selinenes and (+)-valencene, respectively (Lücker et al., 2004; Martin et al., 2009, 2010). A Shiraz allelic variant of the VviTPS24 gene model resulted in the characterization of VvGuaS (indicated as VviTPS24b in **Figure 6A**), producing α-guaiene. This sesquiterpene serves a precursor for rotundone, which is linked to the peppery aroma profile of Shiraz wine (Siebert et al., 2008; Huang et al., 2014, 2015). Although this metabolite is not observed in flowers it serves as an example of genotypic variation impacting on terpene metabolism in a cultivar specific manner. A single amino acid difference between these allelic variants was identified as the mechanistic switch leading to either selinenes or α-guaiene (Drew et al., 2015).

The production of linear farnesenes are facilitated by enzymes that have an active site cavity where cyclization of this cation is prevented by early deprotonation of the substrate (Deligeorgopoulou and Allemann, 2003). Deprotonation of the farnesyl cation will result in stereoisomers of farnesene with VvivMATPS10 and VvCSaFar (VviTPS20) producing those in the *E* orientation. The presence of (*E,Z*)-α-farnesene in SB and SH at the flower bloom stage indicates the presence of a yet to be characterized enzyme that utilizes NPP as substrate with the nerolidyl cation being deprotonated. This novel variant of VviTPS10 presented an interesting scenario due to the extensive amino acid differences between VvGWaBer (Martin et al., 2010) and VvivMATPS10. The observed sequence differences around the active site and the distinct lineage of MA (Myles et al., 2011) suggested that VvivMATPS10 might be unique to MA, rather than a cultivar variant of VviTPS10. Southern blots targeting VviTPS10 show numerous hybridizations, suggesting the presence of multiple genomic regions homologous to VviTPS10. Probe re-mapping showed that the VviTPS10 gene model shares probes with two other gene models linked to functional enzymes (VviTPS12 and -14) (Martin et al., 2010), supporting this observation. Although each of these enzymes were unique in function it was curious to see that they shared minor products that would require a reaction mechanism proceeding through the bisabolyl carbocation (yellow cascade in **Figure 6A**), suggesting a degree of mechanistic conservation (Hong and Tantillo, 2014). Preliminary insight from the phased diploid Cabernet Sauvignon assembly (Chin et al., 2016; Minio et al., 2019) suggests that four homologous loci exists for VviTPS10 (**Figure 5**). This genome is presently being assembled

Stellenbosch University https://scholar.sun.ac.za

and once chromosome assemblies are accessible, we should be able to elucidate if VviTPS10 represents a gene duplicated on both alleles. Nevertheless, the presence of these four homologs gives credence to our belief that VvivMATPS10 encoded by a different locus to VvGwaBer.

CONCLUSION

The domestication history of grapevine has resulted in a high level of variation for *VviTPS* genes with the inbred near homozygous reference genome masking this complexity. Grapevine sesquiterpene biosynthesis was shown to differ in flowers of commercial grapevine cultivars with functional analyses of the gene space for five sesquiterpene synthases, in three cultivars, highlighting the extent of genotypic variation and the impact on floral chemodiversity. The current sesquiterpene biosynthetic landscape in *V. vinifera* suggests that there are mechanistic switches, dictated by cultivar-specific genes or variants, that allow for chemotypic differences between linear sesquiterpenes and cyclizations of FPP/NPP. The genetic potential of the respective cultivars (i.e., genotypic variation) presents multiple potential cascades toward flower sesquiterpenes with current knowledge applied to model these cascades, notwithstanding metabolic flux toward the substrate or terpene modifying enzymes. The current limitations of the reference genome for studying cultivar- and clone specific phenotypic differences is being addressed by utilizing new sequencing and assembly technologies. The phased-diploid assemblies of Cabernet Sauvignon (Chin et al., 2016; Minio et al., 2019) and 16 individual Chardonnay clones (Roach et al., 2018) will shed light on the extent of structural variations within specialized gene families across cultivars and clones as well as allelic differences within a cultivar. It is likely much more complex than what we see in the reference genome and a pangenomic view will be required in order to annotate this gene family more comprehensively.

AUTHOR CONTRIBUTIONS

SS, MV, and PY conceptualized the study. SS performed all field sampling, molecular biology, chemical analysis, data integration, bioinformatics, and statistical analyses. PY transformed the heterologous yeast used in functional characterization and designed the primers for VviTPS isolation. SS and PY drafted the initial manuscript. All authors contributed to the final manuscript.

FUNDING

The study was financially supported with grants from Wine Industry Network for Expertise and Technology (Winetech; IWBT P14/02), the National Research Foundation (NRF) Thuthuka (TTK13070220277), and the Technology and Human Resources for Industry Programme (THRIP) of the Department of Trade and Industry (DTI).

ACKNOWLEDGMENTS

The authors would like to recognize the following people for their contributions toward this study: Mr. Charles Visser and Vititec for allowing access to their vineyards to conduct field experiments. Mr. Franscisco Campos and Ms. Kalem Garcia for their assistance with sample collection and processing. Mr. Malcolm Taylor for assistance with the implementation of the GC-MS methods.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.00177/full#supplementary-material

REFERENCES

Adam, K. P., Thiel, R., and Zapp, J. (1999). Incorporation of 1- 1-C-13 deoxy-D-xylulose in chamomile sesquiterpenes. *Arch. Biochem. Biophys.* 369, 127–132. doi: 10.1006/abbi.1999.1346

Adam-Blondon, A. F. (2014). Grapevine genome update and beyond. *Acta Hortic.* 1046, 311–318. doi: 10.17660/ActaHortic.2014.1046.42

Adam-Blondon, A.-F., Jaillon, O., Vezzulli, S., Zharkikh, A., Troggio, M., and Velasco, R. (2011). "Genome sequence initiatives," in *Genetics, Genomics and Breeding of Grapes*, eds A.-F. Adam-Blondon, J. M. Martiınez-Zapater, and C. Kole (Boca Raton, FL: CRC Press), 211–234. doi: 10.1201/b10948

Allemann, R. K., Young, N. J., Ma, S., Truhlar, D. G., and Gao, J. (2007). Synthetic efficiency in enzyme mechanisms involving carbocations: aristolochene synthase. *J. Am. Chem. Soc.* 129, 13008–13013. doi: 10.1021/ja0722067

Anfora, G., Tasin, M., de Cristofaro, A., Ioriatti, C., and Lucchi, A. (2009). Synthetic grape volatiles attract mated *Lobesia botrana* females in laboratory and field bioassays. *J. Chem. Ecol.* 35, 1054–1062. doi: 10.1007/s10886-009-9686-5

Aradhya, M. K., Dangl, G. S., Prins, B. H., Boursiquot, J.-M., Walker, M. A., Meredith, C. P., et al. (2003). Genetic structure and differentiation in cultivated grape. *Vitis vinifera* L. *Genet. Res.* 81, 179–192. doi: 10.1017/S0016672303006177

Aubourg, S., Lecharny, A., and Bohlmann, J. (2002). Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. *Mol. Genet. Genom.* 267, 730–745. doi: 10.1007/s00438-002-0709-y

Bach, S. S., Bassard, J. -É, Andersen-Ranberg, J., Møldrup, M. E., Simonsen, H. T., and Hamberger, B. (2014). "High-throughput testing of terpenoid biosynthesis candidate genes using transient expression in *Nicotiana benthamiana*," in *Plant Isoprenoids: Methods and Protocols*, ed. M. Rodríguez-Concepción (New York, NY: Springer Science), 245–255. doi: 10.1007/978-1-4939-0606-2_18

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res.* 37, 202–208. doi: 10.1093/nar/gkp335

Barbagallo, M. G., Pisciotta, A., and Saiano, F. (2014). Identification of aroma compounds of *Vitis vinifera* L. flowers by SPME GC-MS analysis. *Vitis J. Grapevine Res.* 53, 111–113.

Bick, J. A., and Lange, B. M. (2003). Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane. *Arch. Biochem. Biophys.* 415, 146–154. doi: 10.1016/S0003-9861(03)00233-9

Bloch, K., Chaykin, S., Phillips, A., and De Waard, A. (1959). Mevalonic acid pyrophospahte and isopentenylpyrphosphate. *J. Biol. Chemsitry* 234, 2595–2604.

Stellenbosch University https://scholar.sun.ac.za

Bohlmann, J., Meyer-Gauen, G., and Croteau, R. (1998). Plant terpenoid synthases: molecular biology and phylogenetic analysis. *Proc. Natl. Acad. Sci. U.S.A.* 95, 4126–4133. doi: 10.1073/pnas.95.8.4126

Buchbauer, G., Jirovetz, L., Wasicky, M., Herlitschka, A., and Nikiforov, A. (1994a). Aroma von Weiβweinblüten: Korrelation sensorischer Daten mit Headspace-Inhaltsstoffen. *Z. Lebensm. Unters. Forsch.* 199, 1–4. doi: 10.1007/BF01192941

Buchbauer, G., Jirovetz, L., Wasicky, M., and Nikiforov, A. (1994b). Headspace analysis of *Vitis vinifera* (Vitaceae) flowers. *J. Essent. Oil Res.* 6, 311–314. doi: 10.1080/10412905.1994.9698383

Buchbauer, G., Jirovetzl, L., Wasieky, M., and Nikiforov, A. (1995). Aroma von Rotweinblüten: Korrelation sensorischer Daten mit Headspace-Inhaltsstoffen. *Zeitschrift für Leb. Und-forsch.* 200, 443–446. doi: 10.1007/BF01193255

Buckingham, J., Cooper, C. M., and Purchase, R. (2015). *Natural Products Desk Reference*. Boca Raton, FL: CRC Press. doi: 10.1201/b19137

Bülow, N., and König, W. (2000). The role of germacrene D as a precursor in sesquiterpene biosynthesis: investigations of acid catalyzed, photochemically and thermally induced rearrangements. *Phytochemistry* 55, 141–168. doi: 10.1016/S0031-9422(00)00266-1

Canaguier, A., Grimplet, J., Gaspero, G., Di Scalabrin, S., Duchêne, E., and Choisne, N. (2017). A new version of the grapevine reference genome assembly (12X.*v2*) and of its annotation (VCost.v3). *Genom. Data* 14, 56–62. doi: 10.1016/J.GDATA.2017.09.002

Carbonell-Bejerano, P., Royo, C., Torres-Pérez, R., Grimplet, J., Fernandez, L., Franco-Zorrilla, J. M., et al. (2017). Catastrophic unbalanced genome rearrangements cause somatic loss of berry color in grapevine. *Plant Physiol.* 175, 786–801. doi: 10.1104/pp.17.00715

Carrier, G., Le Cunff, L., Dereeper, A., Legrand, D., Sabot, F., Bouchez, O., et al. (2012). Transposable elements are a major cause of somatic polymorphism in *Vitis vinifera* L. *PLoS One* 7:e32973. doi: 10.1371/journal.pone.0032973

Caruthers, J. M., Kang, I., Rynkiewicz, M. J., Cane, D. E., and Christianson, D. W. (2000). Crystal structure determination of aristolochene synthase from the blue cheese mold, *Penicillium Roqueforti*. *J. Biol. Chem.* 275, 25533–25539. doi: 10.1074/jbc.M000433200

Chin, C., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* 13, 1050–1054. doi: 10.1038/nmeth.4035

Christianson, D. W. (2008). Unearthing the roots of the terpenome. *Curr. Opin. Chem. Biol.* 12, 141–150. doi: 10.1016/j.cbpa.2007.12.008

Christianson, D. W. (2017). Structural and chemical biology of terpenoid cyclases. *Chem. Rev.* 117, 11570–11648. doi: 10.1021/acs.chemrev.7b00287

Coombe, B. G. (1995). Adoption of a system for identifying grapevine growth stages. *Aust. J. Grape Wine Res.* 1, 100–110. doi: 10.1111/j.1755-0238.1995.tb00086.x

Copolovici, L., Kännaste, A., Pazouki, L., and Niinemets, Ü (2012). Emissions of green leaf volatiles and terpenoids from *Solanum lycopersicum* are quantitatively related to the severity of cold and heat shock treatments. *J. Plant Physiol.* 169, 664–672. doi: 10.1016/j.jplph.2011.12.019

Da Silva, C., Zamperin, G., Ferrarini, A., Minio, A., Dal Molin, A., Venturini, L., et al. (2013). The high polyphenol content of grapevine cultivar tannat berries is conferred primarily by genes that are not shared with the reference genome. *Plant Cell* 25, 4777–4788. doi: 10.1105/tpc.113.118810

Davis, E. M., and Croteau, R. (2000). Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes. *Top. Curr. Chem.* 209, 53–95. doi: 10.1007/3-540-48146-X_2

Degenhardt, J., Köllner, T. G., and Gershenzon, J. (2009). Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* 70, 1621–1637. doi: 10.1016/j.phytochem.2009.07.030

Deligeorgopoulou, A., and Allemann, R. K. (2003). Evidence for differential folding of farnesyl pyrophosphate in the active site of aristolochene synthase: a single-point mutation converts aristolochene synthase into an (E)-β-farnesene synthase. *Biochemistry* 42, 7741–7747. doi: 10.1021/bi034410m

Dèrozier, S., Samson, F., Tamby, J.-P., Guichard, C., Brunaud, V., Grevet, P., et al. (2011). Exploration of plant genomes in the FLAGdb++ environment. *Plant Methods* 7:8. doi: 10.1186/1746-4811-7-8

Drew, D. P., Andersen, T. B., Sweetman, C., Møller, B. L., Ford, C., and Simonsen, H. T. (2015). Two key polymorphisms in a newly discovered allele of the *Vitis vinifera* TPS24 gene are responsible for the production of the rotundone precursor α-guaiene. *J. Exp. Bot.* 67, 799–808. doi: 10.1093/jxb/erv491

Dudareva, N., and Pichersky, E. (2006). "'Floral scent metabolic pathways: their regulation and evolution,'" in *Biology of Floral Scent*, eds N. Dudareva and E. Pichersky (Boca Raton, FL: CRC Press), 55–78.

Durairaj, J., Di Girolamo, A., Bouwmeester, H. J., de Ridder, D., Beekwilder, J., and van Dijk, A. D. (2019). An analysis of characterized plant sesquiterpene synthases. *Phytochemistry* 158, 157–165. doi: 10.1016/j.phytochem.2018.10.020

Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30, 207–210. doi: 10.1093/nar/30.1.207

Emanuelli, F., Lorenzi, S., Grzeskowiak, L., Catalano, V., Stefanini, M., Troggio, M., et al. (2013). Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. *BMC Plant Biol.* 13:39. doi: 10.1186/1471-2229-13-39

Fasoli, M., Dal Santo, S., Zenoni, S., Tornielli, G. B., Farina, L., Zamboni, A., et al. (2012). The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. *Plant Cell* 24, 3489–3505. doi: 10.1105/tpc.112.100230

Forcato, C. (2010). *Gene prediction and functional annotation in the Vitis vinifera* genome. Doctoral dissertation, Universita' Degli Studi Di Padova, Padova.

Gennadios, H. A., Gonzalez, V., Di Costanzo, L., Li, A., Yu, F., Miller, D. J., et al. (2009). Crystal structure of (+)-δ-cadinene synthase from *Gossypium arboreum* and evolutionary divergence of metal binding motifs for catalysis. *Biochemistry* 48, 6175–6183. doi: 10.1021/bi900483b

Gietz, R. D., and Woods, R. A. (2002). "Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method," in *Methods in Enzymology*, eds C. Guthrie and G. R. Fink (Cambridge, MA: Acadmeic Press), 87–96.

Gil, M., Pontin, M., Berli, F., Bottini, R., and Piccoli, P. (2012). Metabolism of terpenes in the response of grape (*Vitis vinifera* L.) leaf tissues to UV-B radiation. *Phytochemistry* 77, 89–98. doi: 10.1016/j.phytochem.2011.12.011

Grant, C. E., Bailey, T. L., and Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018. doi: 10.1093/bioinformatics/btr064

Grimplet, J., Adam-Blondon, A.-F., Bert, P.-F., Bitz, O., Cantu, D., Davies, C., et al. (2014). The grapevine gene nomenclature system. *BMC Genomics* 15:1077. doi: 10.1186/1471-2164-15-1077

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

Hare, S. R., and Tantillo, D. J. (2016). Dynamic behavior of rearranging carbocations - Implications for terpene biosynthesis. *Beilstein J. Org. Chem.* 12, 377–390. doi: 10.3762/bjoc.12.41

Hemmerlin, A., Hoeffler, J. F., Meyer, O., Tritsch, D., Kagan, I. A., Grosdemange-Billiard, C., et al. (2003). Cross-talk between the cytosolic mevalonate and the plastidial methylerythritol phosphate pathways in tobacco bright yellow-2 cells. *J. Biol. Chem.* 278, 26666–26676. doi: 10.1074/jbc.M302526200

Hess, B. A., Smentek, L., Noel, J. P., and O'Maille, P. E. (2011). Physical constraints on sesquiterpene diversity arising from cyclization of the eudesm-5-yl carbocation. *J. Am. Chem. Soc.* 133, 12632–12641. doi: 10.1021/ja203342p

Hong, Y. J., and Tantillo, D. J. (2014). Branching out from the bisabolyl cation. Unifying mechanistic pathways to barbatene, bazzanene, chamigrene, chamipinene, cumacrene, cuprenene, dunniene, isobazzanene, iso-γ-bisabolene, isochamigrene, laurene, microbiotene, sesquithujene, sesquisabinene, thujopsene, trichodiene, and widdradiene sesquiterpenes. *J. Am. Chem. Soc.* 136, 2450–2463. doi: 10.1021/ja4106489

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817

Huang, A.-C., Burrett, S., Sefton, M. A., and Taylor, D. K. (2014). Production of the pepper aroma compound, (-)-rotundone, by aerial oxidation of α-guaiene. *J. Agric. Food Chem.* 62, 10809–10815. doi: 10.1021/jf504693e

Huang, A. C., Sefton, M. A., Sumby, C. J., Tiekink, E. R. T., and Taylor, D. K. (2015). Mechanistic studies on the autoxidation of α-guaiene: structural diversity of the

Stellenbosch University https://scholar.sun.ac.za

sesquiterpenoid downstream products. *J. Nat. Prod.* 78, 131–145. doi: 10.1021/np500819f

Ibáñez, J., Vélez, M. D., Teresa de Andrés, M., and Borrego, J. (2009). Molecular markers for establishing distinctness in vegetatively propagated crops: a case study in grapevine. *Theor. Appl. Genet.* 119, 1213–1222. doi: 10.1007/s00122-009-1122-2

Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449, 463–467. doi: 10.1038/nature06148

Jux, A., Gleixner, G., and Boland, W. (2001). Classification of terpenoids according to the methylerythritolphosphate or the mevalonate pathway with natural 12C/13C isotope ratios: dynamic allocation of resources in induced plants. *Angew. Chemie Int. Ed.* 40, 2091–2094. doi: 10.1002/1521-3773(20010601)40:11<2091::AID-ANIE2091>3.0.CO;2-5

Kapustin, Y., Souvorov, A., Tatusova, T., and Lipman, D. (2008). Splign: algorithms for computing spliced alignments with identification of paralogs. *Biol. Direct.* 3, 1–13. doi: 10.1186/1745-6150-3-20

Knudsen, J. T., and Gershenzon, J. (2006). "'The chemical diversity of floral scent," in *Biology of Floral Scent*, eds N. Dudareva and E. Pichersky (Boca Raton, FL: CRC Press), 27–54.

Köllner, T. G., Held, M., Lenk, C., Hiltpold, I., Turlings, T. C. J., Gershenzon, J., et al. (2008). A maize (E)-beta-caryophyllene synthase implicated in indirect defense responses against herbivores is not expressed in most american maize varieties. *Plant Cell Online* 20, 482–494. doi: 10.1105/tpc.107.051672

Laucou, V., Launay, A., Bacilieri, R., Lacombe, T., Adam-Blondon, A. F., Bérard, A., et al. (2018). Extended diversity analysis of cultivated grapevine *Vitis vinifera* with 10K genome-wide SNPs. *PLoS One* 13:e0192540. doi: 10.1371/journal.pone.0192540

Lawo, N. C., Weingart, G. J. F., Schuhmacher, R., and Forneck, A. (2011). The volatile metabolome of grapevine roots: first insights into the metabolic response upon phylloxera attack. *Plant Physiol. Biochem.* 49, 1059–1063. doi: 10.1016/j.plaphy.2011.06.008

Lesburg, C. A. (1997). Crystal structure of pentalenene synthase: mechanistic insights on terpenoid cyclization reactions in Biology. *Science* 277, 1820–1824. doi: 10.1126/science.277.5333.1820

Lichtenthaler, H. K. (1999). The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50, 47–65. doi: 10.1146/annurev.arplant.50.1.47

Loscos, N., Hernandez-Orte, P., Cacho, J., and Ferreira, V. (2007). Release and formation of varietal aroma compounds during alcoholic fermentation from nonfloral grape odorless flavor precursors fractions. *J. Agric. Food Chem.* 55, 6674–6684. doi: 10.1021/jf0702343

Lovato, M. B., Federal, U., Gerais, D. M., Horizonte, B., Souza, H. A., Muller, L. A., et al. (2012). Isolation of high quality and polysaccharide-free DNA from leaves of *Dimorphandra mollis* (Leguminosae), a tree from the Brazilian Cerrado. *Genet. Mol. Res.* 11, 756–764. doi: 10.4238/2012.March.22.6

Lücker, J., Bowen, P., and Bohlmann, J. (2004). *Vitis vinifera* terpenoid cyclases: functional identification of two sesquiterpene synthase cDNAs encoding (+)-valencene synthase and (-)-germacrene D synthase and expression of mono- and sesquiterpene synthases in grapevine flowers and berries. *Phytochemistry* 65, 2649–2659. doi: 10.1016/j.phytochem.2004.08.017

Ma, Z.-Y., Wen, J., Ickert-Bond, S. M., Chen, L.-Q., and Liu, X.-Q. (2016). Morphology, structure, and ontogeny of trichomes of the grape genus (*Vitis*, *Vitaceae*). *Front. Plant Sci.* 7:704. doi: 10.3389/fpls.2016.00704

Martin, D., Aubourg, S., Schouwey, M., Daviet, L., Schalk, M., Toub, O., et al. (2010). Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, flcDNA cloning, and enzyme assays. *BMC Plant Biol.* 10:226. doi: 10.1186/1471-2229-10-226

Martin, D. M., and Bohlmann, J. (2004). Identification of *Vitis vinifera* (-)-α-terpineol synthase by *in silico* screening of full-length cDNA ESTs and functional characterization of recombinant terpene synthase. *Phytochemistry* 65, 1223–1229. doi: 10.1016/j.phytochem.2004.03.018

Martin, D. M., Chiang, A., Lund, S. T., and Bohlmann, J. (2012). Biosynthesis of wine aroma: transcript profiles of hydroxymethylbutenyl diphosphate reductase, geranyl diphosphate synthase, and linalool/nerolidol synthase parallel monoterpenol glycoside accumulation in Gewürztraminer grapes. *Planta* 236, 919–929. doi: 10.1007/s00425-012-1704-0

Martin, D. M., Toub, O., Chiang, A., Lo, B. C., Ohse, S., Lund, S. T., et al. (2009). The bouquet of grapevine (*Vitis vinifera* L. cv. *Cabernet Sauvignon*) flowers arises from the biosynthesis of sesquiterpene volatiles in pollen grains. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7245–7250. doi: 10.1073/pnas.0901387106

Martinez-Zapater, J. M., Carmona, M. J., Diaz-Riquelme, J., Fernandez, L., and Lijavetzky, D. (2010). Grapevine genetics after the genome sequence: challenges and limitations. *Aust. J. Grape Wine Res.* 16, 33–46. doi: 10.1111/j.1755-0238.2009.00073.x

Matarese, F., Cuzzola, A., Scalabrelli, G., and D'Onofrio, C. (2014). Expression of terpene synthase genes associated with the formation of volatiles in different organs of *Vitis vinifera*. *Phytochemistry* 105, 12–24. doi: 10.1016/j.phytochem.2014.06.007

Matarese, F., Scalabrelli, G., and D'Onofrio, C. (2013). Analysis of the expression of terpene synthase genes in relation to aroma content in two aromatic *Vitis vinifera* varieties. *Funct. Plant Biol.* 40, 552–565. doi: 10.1071/FP12326

Miller, D. J., and Allemann, R. K. (2012). Sesquiterpene synthases: passive catalysts or active players? *Nat. Prod. Rep.* 29, 60–71. doi: 10.1039/C1NP00060H

Miller, D. J., Gao, J., Truhlar, D. G., Young, N. J., Gonzalez, V., and Allemann, R. K. (2008). Stereochemistry of eudesmane cation formation during catalysis by aristolochene synthase from *Penicillium roqueforti*. *Org. Biomol. Chem.* 6, 2346–2354. doi: 10.1039/b804198a

Minio, A., Massonnet, M., Figueroa-Balderas, R., Vondras, A. M., Blanco-Ulate, B., and Cantu, D. (2019). Iso-seq allows genome-independent transcriptome profiling of grape berry development. *G3* doi: 10.1534/g3.118.201008 [Epub ahead of print].

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

Myles, S. (2013). Improving fruit and wine: What does genomics have to offer? *Trends Genet.* 29, 190–196. doi: 10.1016/j.tig.2013.01.006

Myles, S., Boyko, A. R., Owens, C. L., Brown, P. J., Grassi, F., Aradhya, M. K., et al. (2011). Genetic structure and domestication history of the grape. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3530–3535. doi: 10.1073/pnas.1009363108

Nicolas, S. D., Péros, J. P., Lacombe, T., Launay, A., Le Paslier, M. C., Bérard, A., et al. (2016). Genetic diversity, linkage disequilibrium and power of a large grapevine (*Vitis vinifera* L) diversity panel newly designed for association studies. *BMC Plant Biol.* 16:74. doi: 10.1186/s12870-016-0754-z

O'Brien, T. E., Bertolani, S. J., Tantillo, D. J., and Siegel, J. B. (2016). Mechanistically informed predictions of binding modes for carbocation intermediates of a sesquiterpene synthase reaction. *Chem. Sci.* 7, 4009–4015. doi: 10.1039/C6SC00635C

Osbourn, A. E., and Lanzotti, V. [eds]. (2009). *Plant-derived Natural Products*. New York, NY: Springer, doi: 10.1007/978-0-387-85498-4

Peyret, H., and Lomonossoff, G. P. (2013). The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. *Plant Mol. Biol.* 83, 51–58. doi: 10.1007/s11103-013-0036-1

Pichersky, E., and Gershenzon, J. (2002). The formation and function of plant volatiles: Perfumes for pollinator attraction and defense. *Curr. Opin. Plant Biol.* 5, 237–243. doi: 10.1016/S1369-5266(02)00251-0

Picq, S., Santoni, S., Lacombe, T., Latreille, M., Weber, A., Ardisson, M., et al. (2014). A small XY chromosomal region explains sex determination in wild dioecious *V. vinifera* and the reversal to hermaphroditism in domesticated grapevines. *BMC Plant Biol.* 14:229. doi: 10.1186/s12870-014-0229-z

Piel, J., Donath, J., Bandemer, K., and Boland, W. (1998). Mevalonate-independent biosynthesis of terpenoid volatiles in plants: induced and constitutive emission of volatiles. *Angew. Chemie Int. Ed.* 37, 2478–2481. doi: 10.1002/(SICI)1521-3773(19981002)37:18

Reid, K. E., Olsson, N., Schlosser, J., Peng, F., and Lund, S. T. (2006). An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. *BMC Plant Biol.* 6:27. doi: 10.1186/1471-2229-6-27

Roach, M. J., Johnson, D. L., Bohlmann, J., van Vuuren, H. J. J., Jones, S. J. M., Pretorius, I. S., et al. (2018). Population sequencing reveals clonal diversity and ancestral inbreeding in the grapevine cultivar Chardonnay. *PLoS Genet.* 14:e1007807. doi: 10.1371/journal.pgen.1007807

Rohmer, M. (1999). The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat. Prod. Rep.* 16, 565–574. doi: 10.1039/a709175c

Stellenbosch University https://scholar.sun.ac.za

Rynkiewicz, M. J., Cane, D. E., and Christianson, D. W. (2001). Structure of trichodiene synthase from *Fusarium sporotrichioides* provides mechanistic inferences on the terpene cyclization cascade. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13543–13548. doi: 10.1073/pnas.231313098

Sabater-Jara, A. B., Almagro, L., Belchí-Navarro, S., Ferrer, M. A., Barceló, A. R., and Pedreño, M. A. (2010). Induction of sesquiterpenes, phytoesterols and extracellular pathogenesis-related proteins in elicited cell cultures of *Capsicum annuum*. *J. Plant Physiol.* 167, 1273–1281. doi: 10.1016/j.jplph.2010.04.015

Sainsbury, F., Thuenemann, E. C., and Lomonossoff, G. P. (2009). pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* 7, 682–693. doi: 10.1111/j.1467-7652.2009.00434.x

Salmaso, M., Faes, G., Segala, C., Stefanini, M., Salakhutdinov, I., Zyprian, E., et al. (2004). Genome diversity and gene haplotypes in the grapevine (*Vitis vinifera* L.), as revealed by single nucleotide polymorphisms. *Mol. Breed.* 14, 385–395. doi: 10.1007/s11032-004-0261-z

Salvagnin, U., Malnoy, M., Thöming, G., Tasin, M., Carlin, S., Martens, S., et al. (2018). Adjusting the scent ratio: using genetically modified *Vitis vinifera* plants to manipulate European grapevine moth behaviour. *Plant Biotechnol. J.* 16, 264–271. doi: 10.1111/pbi.12767

Schuhr, C. A., Radykewicz, T., Sagner, S., Latzel, C., Zenk, M. H., Arigoni, D., et al. (2003). Quantitative assessment of crosstalk between the two isoprenoid biosynthesis pathways in plants by NMR spectroscopy. *Phytochem. Rev.* 2, 3–16. doi: 10.1023/B:PHYT.0000004180.25066.62

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Shen, B., Zheng, Z., and Dooner, H. K. (2000). A maize sesquiterpene cyclase gene induced by insect herbivory and volicitin: characterization of wild-type and mutant alleles. *Proc. Natl. Acad. Sci. U.S.A.* 97, 14807–14812. doi: 10.1073/pnas.240284097

Shishova, E. Y., Di Costanzo, L., Cane, D. E., and Christianson, D. W. (2007). X-ray crystal structure of aristolochene synthase from *Aspergillus terreus* and evolution of templates for the cyclization of farnesyl diphosphate. *Biochemistry* 46, 1941–1951. doi: 10.1021/bi0622524

Siebert, T. E., Wood, C., Elsey, G. M., and Pollnitz, A. P. (2008). Determination of rotundone, the pepper aroma impact compound, in grapes and wine. *J. Agric. Food Chem.* 56, 3745–3748. doi: 10.1021/jf800184t

Skinkis, P. A., Bordelon, B. P., and Wood, K. V. (2008). Comparison of monoterpene constituents in Traminette, Gewürztraminer, and Riesling winegrapes. *Am. J. Enol. Vitic.* 59, 440–445.

Starks, C. M. (1997). Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science* 277, 1815–1820. doi: 10.1126/science.277.5333.1815

Tantillo, D. J. (2011). Biosynthesis via carbocations: theoretical studies on terpene formation. *Nat. Prod. Rep.* 28, 1035–1053. doi: 10.1039/c1np00006c

Tasin, M., Anfora, G., Ioriatti, C., Carlin, S., De Cristofaro, A., Schmidt, S., et al. (2005). Antennal and behavioral responses of grapevine moth *Lobesia botrana* females to volatiles from grapevine. *J. Chem. Ecol.* 31, 77–87. doi: 10.1007/s10886-005-0975-3

This, P., Jung, A., Boccacci, P., Borrego, J., Botta, R., Costantini, L., et al. (2004). Development of a standard set of microsatellite reference alleles for identification of grape cultivars. *Theor. Appl. Genet.* 109, 1448–1458. doi: 10.1007/s00122-004-1760-3

Thomas, B. J., and Rothstein, R. (1989). Elevated recombination rates in transcriptionally active DNA. *Cell* 56, 619–630. doi: 10.1016/0092-8674(89)90584-9

Van Poecke, R. M. P., Posthumus, M. A., and Dicke, M. (2001). Herbivore-induced volatile production by *Arabidopsis thaliana* leads to attraction of the parasitoid *Cotesia rubecula*: chemical, behavioral, and gene-expression analysis. *J. Chem. Ecol.* 27, 1911–1928. doi: 10.1023/A:1012213116515

von Arx, M., Schmidt-Büsser, D., and Guerin, P. M. (2011). Host plant volatiles induce oriented flight behaviour in male European grapevine moths, *Lobesia botrana*. *J. Insect Physiol.* 57, 1323–1331. doi: 10.1016/j.jinsphys.2011.06.010

Wedler, H., Pemberton, R., and Tantillo, D. (2015). Carbocations and the complex flavor and bouquet of wine: mechanistic aspects of terpene biosynthesis in wine grapes. *Molecules* 20, 10781–10792. doi: 10.3390/molecules200610781

Wood, C., Siebert, T. E., Parker, M., Capone, D. L., Elsey, G. M., Pollnitz, A. P., et al. (2008). From wine to pepper: rotundone, an obscure sesquiterpene, is a potent spicy aroma compound. *J. Agric. Food Chem.* 56, 3738–3744. doi: 10.1021/jf800183k

Wymore, T., Chen, B. Y., Nicholas, H. B., Ropelewski, A. J., and Brooks, C. L. (2011). A mechanism for evolving novel plant sesquiterpene synthase function. *Mol. Inform.* 30, 896–906. doi: 10.1002/minf.201100087

Yauk, Y.-K., Ged, C., Wang, M. Y., Matich, A. J., Tessarotto, L., Cooney, J. M., et al. (2014). Manipulation of flavour and aroma compound sequestration and release using a glycosyltransferase with specificity for terpene alcohols. *Plant J.* 80, 317–330. doi: 10.1111/tpj.12634

Yoshikuni, Y., Ferrin, T. E., and Keasling, J. D. (2006). Designed divergent evolution of enzyme function. *Nature* 440, 1078–1082. doi: 10.1038/nature04607

Young, P., Eyeghe-Bickong, H. A., du Plessis, K., Alexandersson, E., Jacobson, D. A., Coetzee, Z. A., et al. (2015). Grapevine plasticity in response to an altered microclimate: sauvignon Blanc modulates specific metabolites in response to increased berry exposure. *Plant Physiol.* 170, 1235–1254. doi: 10.1104/pp.15.01775

Zulak, K. G., and Bohlmann, J. (2010). Terpenoid biosynthesis and specialized vascular cells of conifer defense. *J. Integr. Plant Biol.* 52, 86–97. doi: 10.1111/j.1744-7909.2010.00910.x

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 © 2019 Smit, Vivier and Young. 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) and the copyright owner(s) 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.

Stellenbosch University https://scholar.sun.ac.za

Chapter 3

The grapevine terpene synthase (*VviTPS*) compendium: Comparative (within species) genomics to predict catalytic functions for genes involved in specialised metabolism.

This chapter was submitted to Frontiers in Plant Science.

Stellenbosch University https://scholar.sun.ac.za

The grapevine terpene synthase (*VviTPS*) compendium: Comparative (within species) genomics to predict catalytic functions for genes involved in specialised metabolism.

3.1 Abstract

The *Vitis vinifera* L. terpene synthase (*VviTPS*) family was comprehensively annotated on the phased diploid genomes of three closely related cultivars (Cabernet Sauvignon, Carménère and Chardonnay). *VviTPS* gene regions were grouped to chromosomes with the haplotig assemblies used to identify allelic variants. Functional predictions of the *VviTPS* subfamilies were performed using enzyme active site phylogenies resulting in the putative identification of the initial substrate and cyclisation mechanism of VviTPS enzymes. Subsequent groupings into carbocation cascades was coupled with an analysis of cultivar-specific gene duplications resulting in the identification of conserved and unique *VviTPS* clusters. These findings are presented as a compendium of interactive networks where any *VviTPS* of interest can be queried through BLAST, allowing for a rapid identification of VviTPS-subfamily, enzyme mechanism and degree of connectivity (i.e. extent of duplication). The within species comparative genomic analyses presented provides valuable new insights into the complexities of genome structures and how it impacts an important gene family involved in specialised metabolism.

3.2 Introduction

Grapevine has an extensive domestication history that includes various non-*vinifera* hybridisations, resulting in high levels of heterozygosity (Minio et al., 2017). The sequencing of *Vitis vinifera* cultivar Pinot Noir resulted in the first genome of a woody crop species (Jaillon et al., 2007; Velasco et al., 2007). Inbreeding of Pinot Noir simplified the genome to near homozygosity (93%) which facilitated sequencing of PN40024 (Jaillon et al., 2007). Concurrently a heterozygous clone of Pinot Noir, ENTAV115, was sequenced, but difficulties in assembly of the heterozygous and highly repetitive regions resulted in a fragmented genome, limiting its usability (Figueroa-Balderas et al., 2019; Velasco et al., 2007). Continuous improvement over the last decade resulted in numerous assemblies and annotations of the PN40024 reference genome with the latest version (12X.v2 and VCost.v3 annotation) improving the contig coverage and orientation by 14% over the previous assembly (12X.v0) and annotation (V1). However, 2.64 Mbp of contig sequences remain unmapped (chr. 00) while the orientation of numerous mapped contigs remain uncertain (Canaguier et al., 2017).

A combination of crossing (with close relatives as well as non-*vinifera* species) and millennia of propagation have resulted in the expansion of certain *V. vinifera* gene families. Of interest to this study are those linked to volatile organic compounds (VOC) that are often associated with aromatic cultivars. Terpenoids are known to modulate flavour and aroma profiles with monoterpenoids associated with floral and Muscat aromas while a spicy or pepper aroma, in certain wine styles, have

Stellenbosch University https://scholar.sun.ac.za

been attributed to sesquiterpenoids. Flavour and aroma is, however, complex and can be influenced by a multitude of factors that include amongst others the cultivar, vinification style, viticulture practices and extent of compound glycosylation (i.e. bound versus free volatiles) (D'Onofrio et al., 2017; Hjelmeland et al., 2015; Robinson et al., 2014; Swiegers et al., 2005). Terpenoids can furthermore be synthesised *de novo* by certain yeasts during fermentation, while other genera are known to liberate bound terpenoids by cleaving the glycosyl bonds (Carrau et al., 2005). It is therefore difficult to link the vine's genetic potential with terpenoids observed in the wine.

Annotation of the PN40024 terpene synthase (VviTPS) family revealed extensive duplications and neofunctionalization resulting 152 VviTPS-like regions, of which 69 encode for putatively functional proteins (Martin et al., 2010). The extent to which cultivar-specific *VviTPS* gene variation impacts the genetic potential of a vine, with subtle mutations often rendering a gene non-functional has also been demonstrated (Drew et al., 2015; Dueholm et al., 2019; Smit et al., 2019). Extrapolating this to the more than 6000 grapevine accessions planted worldwide (This et al., 2006) suggests extensive *VviTPS* diversity, with the PN40024 genome sequence probably representing only a fraction of the genetic potential.

Genetic variability is only one aspect of *VviTPS* variability with the encoded enzymes known to facilitate complex biochemical cascades. All terpenes consist of the C_5 prenyl diphosphate building blocks isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). These two building blocks arise from the 2-C-methyl-D-erythritol 4-phosphate (MEP) and 1-deoxy-D-xylulose 5-phosphate (DOXP) pathways that are compartmentalised to the cytosol and plastids, respectively, although metabolic crosstalk between these pathways have been shown (Bick and Lange, 2003; Bloch et al., 1959; Hemmerlin et al., 2003; Lichtenthaler, 1999; Rohmer, 1999). Head-to-tail coupling of IPP and DMAPP results in elongated prenylated substrates that are characteristic to the various known terpene classes. Of particular interest in grapevine due to their volatile flavour and aroma properties, are the C_{10} mono- and C_{15} sesquiterpenes.

Monoterpene biosynthesis proceeds through the DOXP pathway with geranyl diphosphate (GPP) as the initial substrate with sesquiterpene biosynthesis proceeding through the (MEP) pathway using farnesyl diphosphate (FPP) and its isomer, nerolidyl diphosphate (NPP) as substrates (Davis and Croteau, 2000). The VviTPS enzymes involved in mono- and sesquiterpene biosynthesis facilitate complex biochemical reactions that include ring-closures, hydride shifts, protonation and deprotonation events and various rearrangements (Cane, 1990; Christianson, 2006; Davis and Croteau, 2000; Wedler et al., 2015). Sesqui-TPS enzymes are more promiscuous in their product profile due to increased number of orientations that can arise from the added double bond of the FPP substrate. Enzyme promiscuity is known to be affected by subtle sequence variations that alter the product specificity or change the enzyme function completely (Drew et al., 2015; Li et al., 2013; Smit et al., 2019).

Stellenbosch University https://scholar.sun.ac.za

The recently available diploid grapevine genomes therefore present an opportunity to explore cultivar-specific *VviTPS* variation to understand the structure-function (i.e. gene-protein-terpene) relationships. Cabernet Sauvignon (CS), Carménère (CR) and Chardonnay (CH) were selected for this purpose as they were sequenced and assembled using the same technology: Pac-Bio SMRT sequencing with FALCON-UNZIP phased assembly (Chin et al., 2016; Minio et al., 2019; Roach et al., 2018). Focussing on the *VviTPS* family, the aim was to evaluate and correct gene models where necessary and then explore the extent of haplotype and genotype variations. An in-depth analysis of the *VviTPS* family ultimately resulted in a compendium of information that includes chromosome groupings, functional prediction (which includes TPS-subfamily, initial substrate and cyclisation mechanisms), cultivar-specific duplication analysis and identification of conserved VviTPS functions. Interactive networks were constructed for gene duplication and genotype/haplotype variations, making the compendium easily accessible. The networks can be queried using BLAST and all relevant VviTPS information interactively accessed in the respective networks.

3.3 Methodology

The computational methodologies employed utilised phased diploid genome sequences of three *V. vinifera* cultivars, available publicly as unassembled contigs. The primary and haplotig assemblies (refer to Chapter 1) were independently annotated using the *VviTPS* family annotations of the PN40024 reference genome. Cultivar-specific VviTPS nucleotide and derived protein sequences were used for gene subfamily and functional predictions as well comparative analyses between the haplotypes and cultivars, respectively.

3.3.1 Genome assemblies and annotations utilised

Genomes for *V. vinifera* cultivars (Chin et al., 2016; Jaillon et al., 2007; Minio et al., 2017, 2019; Roach et al., 2018) listed in Table 3.1 were downloaded from the listed repositories. PN40024 12X.2 assembly and VCost.v3 (V3) annotation was used as the reference genome (Canaguier et al., 2017; Jaillon et al., 2007). The GFF3 annotation for the terpene synthase family (https://urgi.versailles.inra.fr/Species/Vitis/Annotations) was used for *VviTPS* positioning on the reference genome. PN40024 *VviTPS* sequences identified and curated by Martin et al. (2010) were retrieved from FLAGdb++ (Dèrozier et al., 2011).

Table 3.1. Genome assemblies and annotations utilised

Genome	Assembly Type	Annotation Version	Repository
PN40024 12X.v2 (PN)	Haploid	VCost.v3	https://urgi.versailles.inra.fr/Species/Vitis
Cabernet Sauvignon (CS)	Diploid	V1	http://cantulab.github.io/data.html
Carménère (CR)	Diploid	V1	http://cantulab.github.io/data.html
Chardonnay (CH)	Diploid	V1	https://doi.org/10.5281/zenodo.1480037

Stellenbosch University https://scholar.sun.ac.za

The domestication history of these cultivars was inferred by using the Vitis International Variety Catalogue (www.vivc.de, accessed 18 November 2019) and domestication histories described by Myles et al. (2011) and Minio et al. (2019).

3.3.2 Identification of *VviTPS* gene regions on the diploid genomes

The Exonerate tool (Slater and Birney, 2005) was used to identify VviTPS-like regions on the primary and haplotig assemblies of the respective diploid genomes (Table 3.1). PN *VviTPS* gene models served as query sequences with the exonerate parameters set to the est2genome model, percentage of the maximal score set at 90% and intron size limited to 3000 bp. A detailed explanation of the Exonerate analysis can be found in Appendix A under Script 1. Exonerate computations were performed using the Stellenbosch University Central Analytical Facilities' HPC2: http://www.sun.ac.za/hpc. Exonerate-gff outputs were annotated on the respective genome contigs and manually curated with CLC Main Workbench 7 (CLC Bio-Qiagen, Denmark) to identify hit regions with the greatest coverage and highest mapping score.

3.3.3 Annotation of *VviTPS* genes

The identified gene regions were compared with the computational annotations reported for the respective genomes in Table 3.1. Each identified gene region was assigned a unique accession consisting of a two-letter cultivar code (Table 3.1) followed by a sequential TPS number. When automated annotations for the respective genomes (Table 3.1) where congruent with the annotation generated in this study, the annotation specific locus ID was maintained as the parent ID in the annotation file. Annotated coding sequences of these congruent regions were maintained as far as possible, but manual correction of several gene regions was necessary with such corrections noted in the annotation file. Gene regions lacking a parent ID indicates a newly annotated region. Partial genes were noted when there were four or less exons and a lack of start and/or stop codons. Genes were considered to be complete if a start and stop codon was present at the terminal ends and the exons number was greater than four. Complete genes were subsequently evaluated for the presence of a reading frame, with genes lacking a full-length open reading frame (fl-ORF) tagged as a disrupted (d-ORF), with the disruption being either a premature stop, or a frameshift (insertion-deletion) mutation.

Protein sequences were derived for the fl-ORF's and the terpene synthase N-terminal (PF01397) and C-terminal (PF03936) domains predicted using the PFam Domain Search function of CLC Main Workbench 7 (CLC Bio-Qiagen, Denmark). The motif search function CLC Main Workbench 7 was used to identify motifs characteristic to TPS proteins (Durairaj et al., 2019; Gao et al., 2012; Rynkiewicz et al., 2002; Starks, 1997; Williams et al., 1998a).

3.3.4 Putative identification of duplicated gene regions

A BLASTn alignment (Altschul et al., 1997; Camacho et al., 2009) of complete gene regions for each cultivar was performed and duplications identified by calculating the identity (I') using the formula I'

Stellenbosch University https://scholar.sun.ac.za

$= I \text{ X } \text{Min}(n_1/L_1, n_2/L_2)$, described by Li et al. (2001), with I being the number of identities and gaps, n the aligned length and L the total length of the query and subject sequences, respectively. For BLASTn analyses of primary-to-primary and haplotig-to-haplotig complete genes, an *E*-value of 1e-5 was used, with the maximum number of alignments (max-hsps) limited to 5 and number of aligned sequences (max-target-seqs) set to 10. The latter two parameters were set to 1 when haplotigs-to-primary alignments were performed.

3.3.5 Rapid assembly of contigs

Chromosome positions of *VviTPS*-like gene regions were inferred by mapping *VviTPS* containing contigs to the PN reference genome using rapid reference-guided assembly (RaGOO) (Alonge et al., 2019; Canaguier et al., 2017). The RaGOO parameters for chimera breaking (-b), structural variant calling (-s) and a gap padding (-g) of 200 were used with unplaced contigs not assembled to a random chromosome. The random pseudo-molecule (chr. 00) of the reference genome was not included for RaGOO assemblies. The output of this cultivar-specific all-against-all assembly was used to group contigs according to their highest scoring PN40024 chromosome followed by chromosome specific contig assembly. The respective RaGOO outputs were visualised using the contig alignment function of the Alvis tool (Martin and Leggett, 2019).

3.3.6 Functional annotation of *VviTPS* genes

Multiple sequence alignments (MSA) and phylogenetic tree constructions were performed in the CLC Main Workbench 7 (CLC Bio-Qiagen, Denmark). For nucleotide alignments the ClustalO algorithm was used while the MUSCLE algorithm was used for protein sequences. Phylogenetic trees were constructed with UPGMA, Jukes-Cantor as distance measure and 100 bootstrap replicates (Edgar, 2004b, 2004a; Jukes and Cantor, 1969). MSA's were performed at the nucleotide level using the 152 *VviTPS* gDNA and mRNA sequences predicted by Martin et al. (2010) as reference. Phylogenetic position relative to PN40024 gDNA sequences were used to group gene regions into TPS-subfamilies (Bohlmann et al., 1998; Martin et al., 2010) with the eulerr R package (Larsson, 2019; R Core Team, 2013) used to visualise the data.

Protein sequence phylogenies with characterised grapevine TPSs (Appendix B, Table 1) were used to group proteins into TPS-subfamilies. For the TPS-a subfamily, the active site region was identified as described by Durairaj et al. (2019) and aligned as described earlier. This active site phylogeny and the Database of Characterized Plant Sesquiterpene Synthases (Durairaj et al., 2019) was used to divide TPS-a members into groups based on their parent cation and first cyclisation. For the TPS-b subfamily a similar approach to Durairaj et al. (2019) was applied where only the active site region between the C-terminal metal binding motifs, if present, were aligned. The product profiles of TPS-b members (Martin et al., 2010) were used to predict a mono-TPS reaction mechanism (Davis and Croteau, 2000; Hyatt et al., 2007; Schwab et al., 2001; Schwab and Wüst, 2015; Williams et al., 1998a; Xu et al., 2018) and categorise proteins according to their initial carbocation intermediate (terpinyl or linalyl cation). The latter was further subcategorised by

Stellenbosch University https://scholar.sun.ac.za

considering whether or not quenching occurs before deprotonation. The TPS-g subfamily was subcategorised using the full-length protein alignment and phylogenetic position relative to functional proteins (Martin et al., 2010).

3.3.7 Finding homologous proteins between cultivars

The cluster function of MMseqs2 (Steinegger and Söding, 2017) was used for all-against-all clustering of proteins with the following parameters: bidirectional alignment coverage mode with a minimum coverage of 85%, minimum sequence identity of 75%, *E*-value of 1e-5 and greedy clustering (cluster-mode 2). Representative sequences from the clustering were extracted as described in Appendix A, Script 2.

3.3.8 Network construction

Cytoscape v3.7.2 (Shannon et al., 2003) was used to construct all networks presented in this study with the data generated from the aforementioned methodologies used for node and edge metadata.

3.4 Results

3.4.1 Relatedness of the genomes

The domestication history of grapevine (Myles et al., 2011) and available pedigree information (Maul and Töpfer, 2015) shows that CR and CS have a common parent while PN is a parent to CH. All cultivars share Traminer as an ancestor. The relatedness (pedigree) of cultivars used for genomes discussed in this study is shown in Appendix B, Figure 1.

3.4.2 Diploid genome VviTPS-like gene regions

Nearly all of the diploid contigs annotated with a *VviTPS* could be assigned to a reference chromosome, with the exception of 1 CS and 2 CR contigs. The RaGOO grouping scores per chromosome (Appendix B, Table 2) ranged between 53% and 97% with an average of 75%, indicating that the contigs could be placed on a chromosome with an acceptable level of confidence. However, the exact position on a chromosome could not be accurately estimated, as evident by the location scores, reflecting a low level of collinearity to the reference genome (Appendix B, Table 2). Contig alignments to the reference genome using Alvis (Appendix C) illustrates the extent of discontiguity when mapping contigs to the reference genome.

The euler graphs in Figure 3.1 show *VviTPS* subfamily members per chromosome for the diploid assemblies with PN40024 as a reference. Despite the latest assembly improvements for PN40024, a large number *VviTPS* genes are yet to be assembled to a chromosome. The diploid assemblies showed an inverse proportional relationship between unassembled and chr. 10 genes relative to PN40024, indicating that long read sequencing has overcome, to a large extent, the unresolved location of chr. 10 *VviTPS* genes. 28 *VviTPS* genes for CH and 41 for CR and CS, respectively, were placed on chr. 10, compared to a single gene on PN40024 (Canaguier et al., 2017; Martin et al., 2010). The majority of these genes are TPS-g subfamily members based on their

Stellenbosch University https://scholar.sun.ac.za

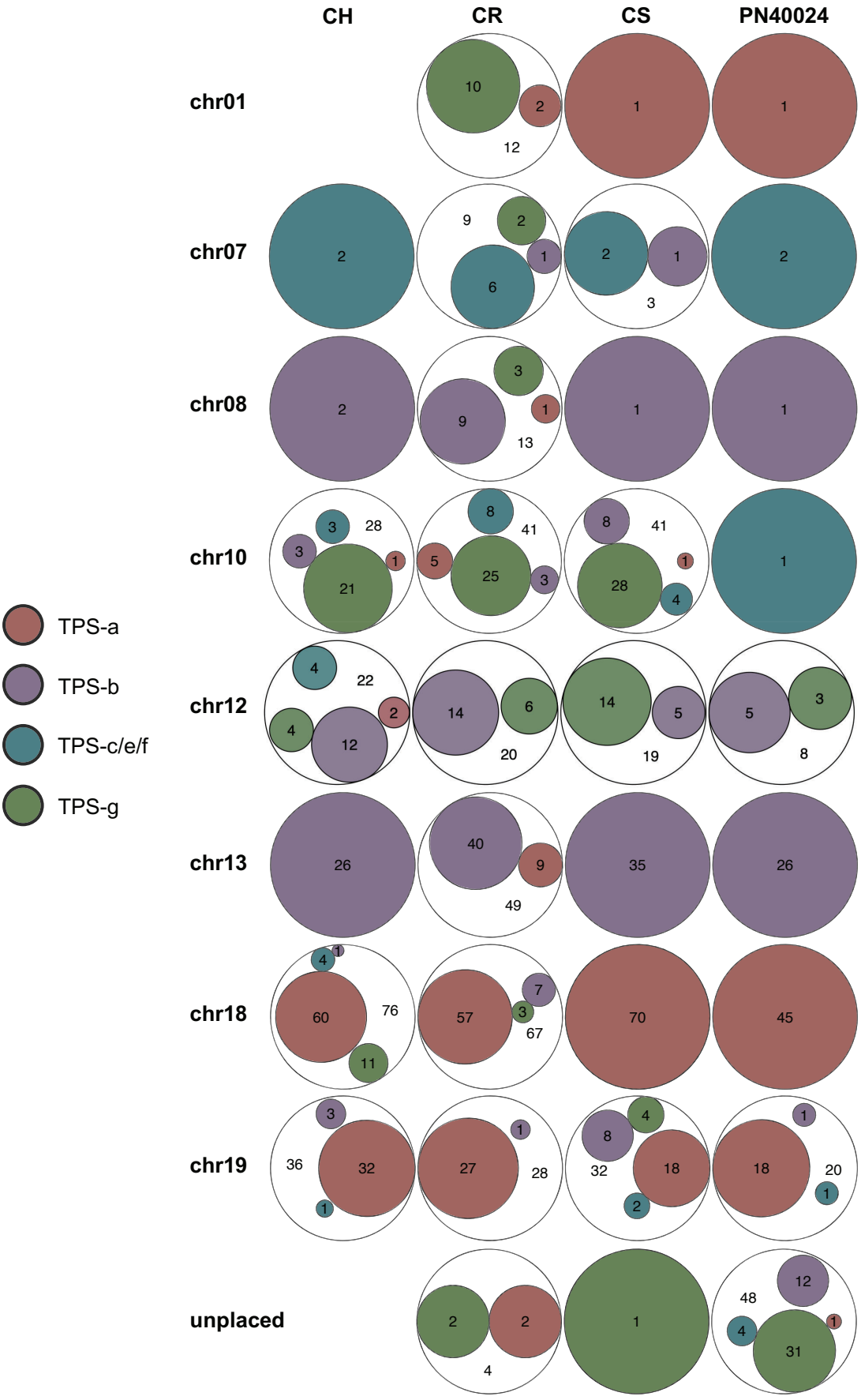

Figure 3.1 Euler diagrams summarising the chromosome specific distribution of *VviTPS*-subfamilies for each of the diploid genomes: Cabernet Sauvignon (CS), Carménère (CR) and Chardonnay (CH)) as well as the Martin et al. (2010) annotation of PN40024 (PN). The legend shows the different *VviTPS* subfamilies, proportionally sized within the euler diagrams to reflect the total number of *VviTPS*-like gene regions per cultivar and chromosome.

Stellenbosch University https://scholar.sun.ac.za

phylogenetic similarity to PN40024 gDNA sequences (Appendix B, Figure 2). Furthermore, CR had more than three times the number of genes on chr. 01, 07 and -08 than CS, CH or PN. The majority of TPS-a genes are located on chr. 18 and -19 with nearly all TPS-b genes on chr. 13.

The distribution of complete and partial gene regions on the primary and haplotig assemblies is shown in Figure 3.2a. Complete gene regions were sub-categorised into fl-ORF or d-ORF, with the latter representing regions that are considered pseudogenes. Although CR had the greatest number of *VviTPS*-like regions (243), only 49% of these regions encode for a putative fl-ORF, shown in Figure 3.2b, with 84% of the complete genes being duplicated (Figure 3.2c). CS and CH had a similar number of *VviTPS*-like regions (203 and 192, respectively), with CH showing the greatest proportion of fl-ORF (77%) of all three cultivars (Figure 3.2b). CS and CH *VviTPS* families are also extensively duplicated, however, ~30% of their complete *VviTPS* genes were hemizygous. Although the diploid genomes are unassembled, the size and contiguity of the diploid contig assemblies allowed for the extent of gene duplications to be investigated, as illustrated by the cultivar specific networks in Figure 3.3. Gene regions with an identity score (I') greater than 80% were considered to be duplicated with those localising to the same contig considered to be tandem duplicates. Tentative duplications show genes that are not on the same contig (i.e. possible genome wide duplications). Appendix B, Figure 3 shows an alternative node colouring for the aforementioned figure, illustrating their chromosomal localisation. The duplication distribution in the edge interactions graph (Figure 3.3d) gives an estimation of the homozygosity for each cultivar, with CS showing the greatest percentage (32%) of haplotype edge connections, i.e. potential allelic variants.

3.4.3 Functional annotation of the *VviTPS* family

Protein sequences derived from fl-ORFs and subsequent phylogenetic similarity to known functional VviTPS enzymes clearly separate the proteins into subfamilies, illustrated in Appendix B, Figure 4. The TPS-a, -b and -g subfamilies represent the majority of putative proteins and were subsequently analysed in a family specific manner to predict their function.

The TPS-a subfamily separates into three major groups based on the initial substrate (FPP and/or NPP) utilized, illustrated in Figure 3.4. Two acyclic subgroups were associated with each of these substrates. With the exception of the acyclic sesquiterpenes, all enzymes that use NPP as sole substrate will proceed through an initial 1,6-cylisation of the nerolidyl cation (Davis and Croteau, 2000). Reactions mechanisms that proceed from FPP formed three distinct clades, indicated by the red triangles, with each clade showing a group for 1,10- and 1,11-cyclisations. Acyclic sesquiterpenes and those that require 1,11-cyclisation showed commonality in clade 1 that is distinct from the 1,10-cylisation group. Clade 2 showed three distinct groups with a unique subgroup consisting of both 1,10 and 1,11-cylisation enzymes. The third clade had a number of enzymes that could not be definitively placed into a cyclisation group but, as with clade 2, showed clear separation between the 1,10 and 1,11 cyclisation mechanisms.

Stellenbosch University https://scholar.sun.ac.za

Although the VviTPS-b subfamily utilizes a single substrate for monoterpene biosynthesis, enzymes could still be grouped into distinct reaction mechanisms, illustrated in Figure 3.5, where the cyclic reaction mechanism is referred to as TPS-b Type I while the acyclic mechanism is referred to as TPS-b Type II. Type II enzymes however, formed three distinct clades, of which two are for the single product enzymes associated with linalool and ocimene biosynthesis, VvPNRLin and VvGwBOci/VvCSbOci, respectively (Martin et al., 2010). The third Type II clade is represented by a single functional enzyme (VvCSbOciM) that produces 98% acyclic monoterpenes, (E)-beta-ocimene and myrcene, and minor amount of the cyclic monoterpene pinene (Martin et al., 2010). This clade is closely related to a clade of multiproduct Type I enzymes.

It was previously shown that the VviTPS-g family is expanded in grapevine, forming three distinct clades that separate according to the product profiles of *in vitro* characterised TPS-g enzymes (Martin et al., 2010). Those that accept only GPP as substrate to form geraniol formed a distinct clade with the multi-substrate enzymes forming the other two clades, as shown in Figure 3.6. Annotation of this subfamily to a large extent resolved the current lack in chromosome mapping of this family. Despite the various improvements of the reference genome, chr. 10 remained difficult to assemble, with inbreeding of PN40024 not being able to reduce the extent of heterozygosity. The lack of a sufficiently resolution for this chromosome therefore resulted in highly discontiguous mapping of diploid contigs to PN40024 chr. 10.

Stellenbosch University https://scholar.sun.ac.za

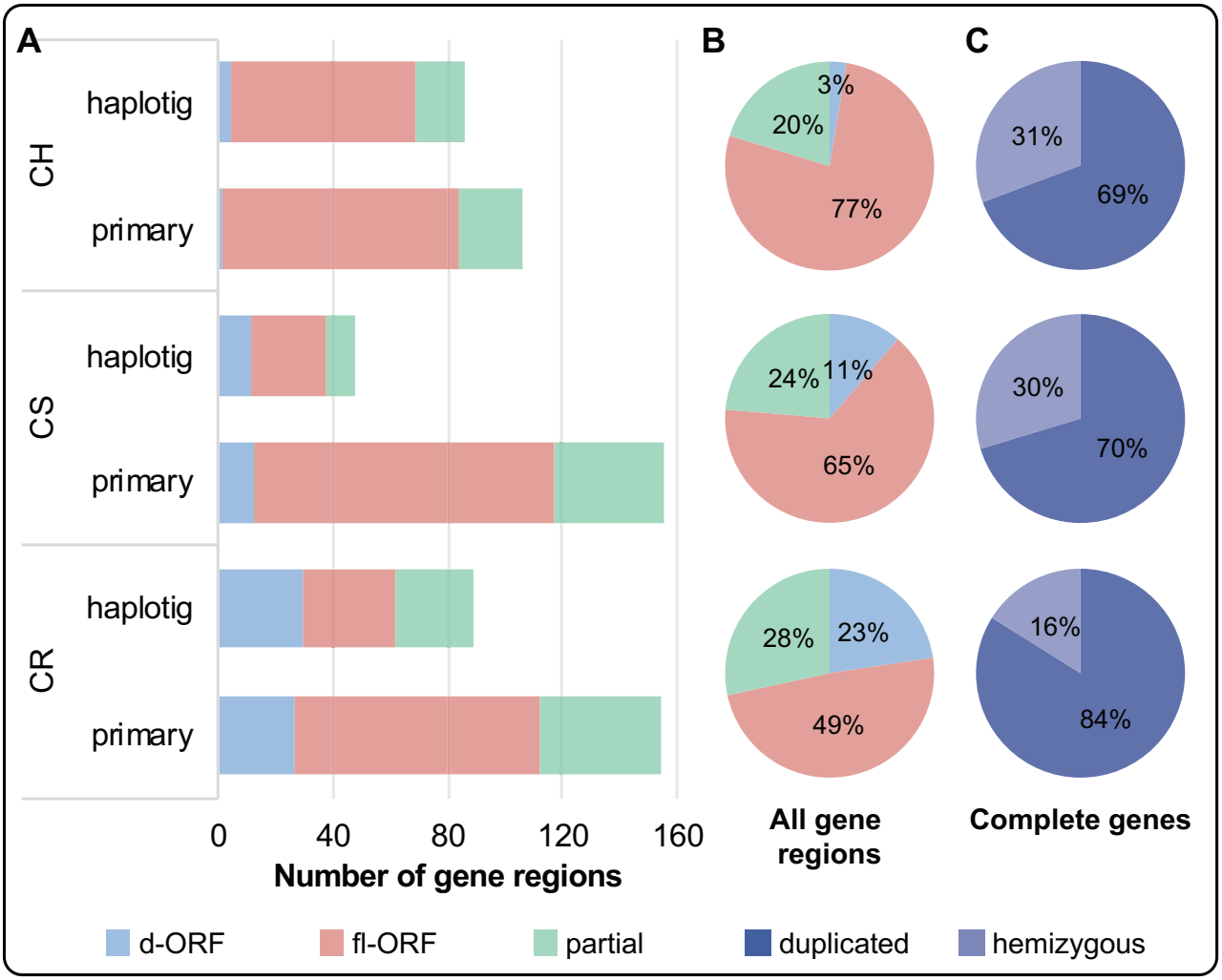

Figure 3.2 (A) The total number of *VviTPS*-like gene regions on the primary contigs and haplotigs is shown for the draft diploid genomes of Cabernet Sauvignon (CS), Carménère (CR) and Chardonnay (CH). **(A)** The total number of *VviTPS*-like gene regions were further classified by the type of open reading frame (ORF): disrupted (d-ORF) contain frameshifts and/or premature stop codons that render the gene non-functional; full-length (fl-ORF) are predicted to be functional; and partial genes that have four less exons (i.e. pseudogenes). The combined percentage distribution of these ORFs across the haplotypes is shown in **(B).** The percentage of complete gene regions, the sum of fl-ORF and d-ORF, that are duplicated (degree of similarity (I') > 80%) or hemizygous is shown in **(C)**.

Stellenbosch University https://scholar.sun.ac.za

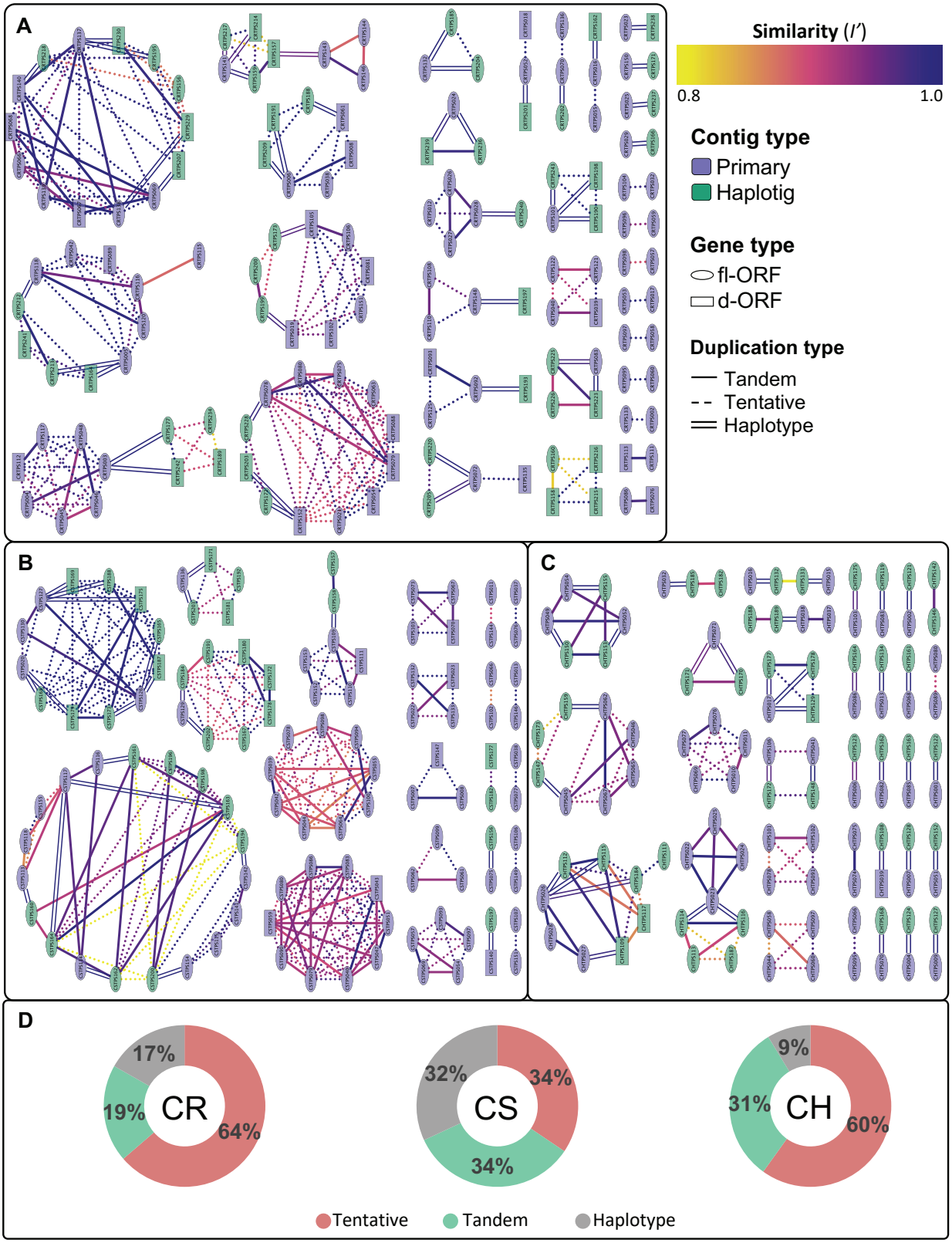

Figure 3.3 Cytoscape network illustrating the connectedness and degree of similarity (I') for duplicated complete genes for **(A)** Carménère (CR), **(B)** Cabernet Sauvignon (CS) and **(C)** Chardonnay (CH). Nodes represent *VviTPS* genes and are connected by edges, signifying homology of I' > 80%. Complete genes are grouped into those with a full-length or disrupted open reading frames (fl-ORF or d-ORF) for the cultivar-specific haplotypes. The type of edge interactions are further categorised as tandem duplicates if the gene is present on the same contig; haplotype duplications are on primary contigs and haplotigs that localise to the same chromosome and were inferred from RaGOO assemblies to PN40024; with tentative duplications showing genes with high homology that cannot be defined by the two previous groupings. The total percentage contribution of these groupings is shown in **(D)**.

Stellenbosch University https://scholar.sun.ac.za

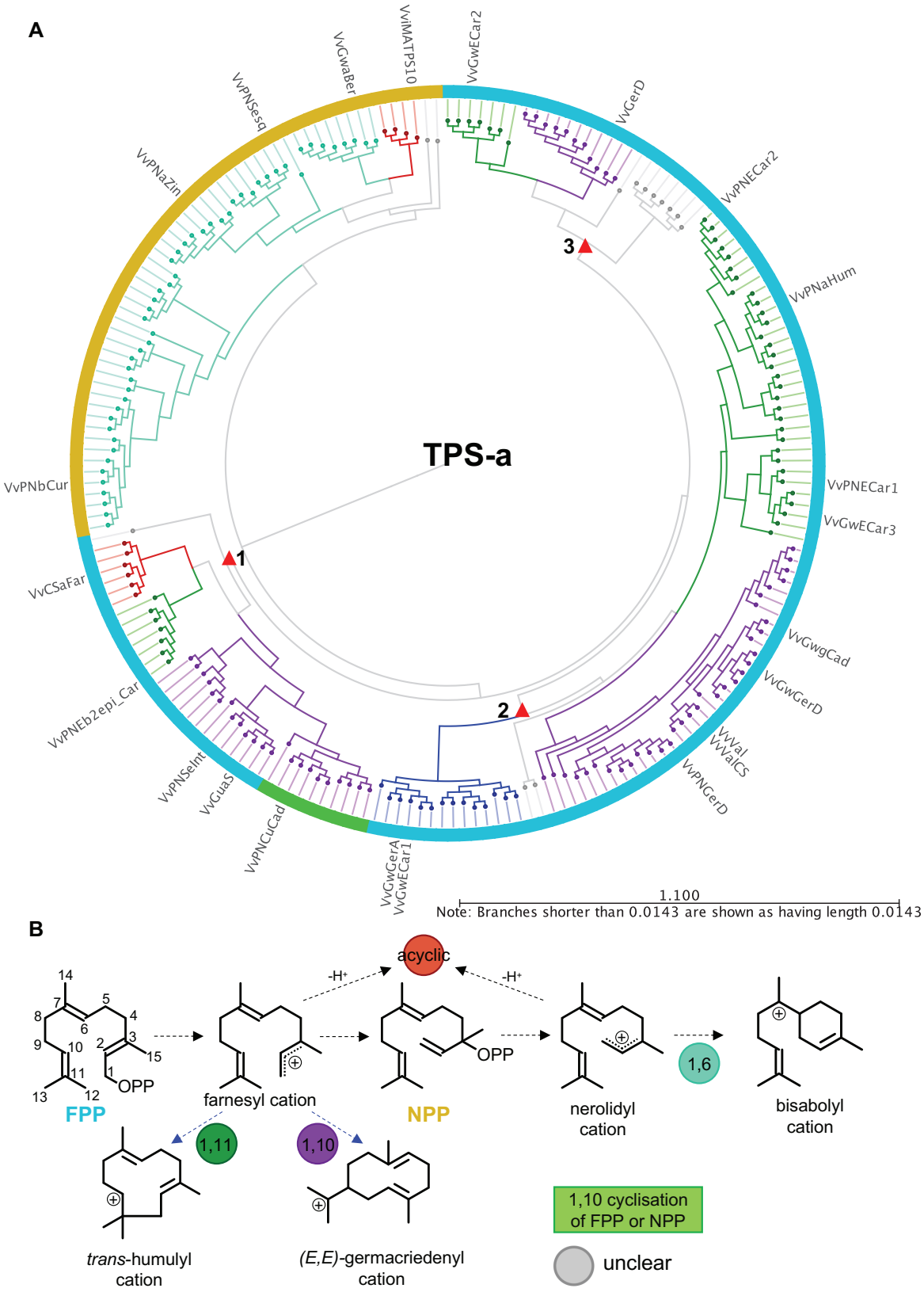

Figure 3.4. (A) Active site phylogeny of putative VviTPS-a proteins from the diploid genomes and functionally characterised enzymes (outer labels) were used to predict the initial substrate (outer coloured ring) and cyclisation mechanism (branch colour). The red triangles indicate subclades for enzymes that utilise farnesyl diphosphate (FPP) as initial substrate, shown by the blue outer ring. The yellow outer ring represents enzymes that utilize nerolidyl diphosphate (NPP) as initial substrate. Both FPP and NPP can be ionised or protonated to form the initial carbocation intermediates indicated in **(B)**. Branches in **(A)** are coloured according to these initial carbocation cyclisation mechanisms and the subsequent carbocation formed, shown in **(B)**. Deprotonation of the initial carbocation result in the formation of acyclic sesquiterpenes, as shown in **(B)**.

Stellenbosch University https://scholar.sun.ac.za

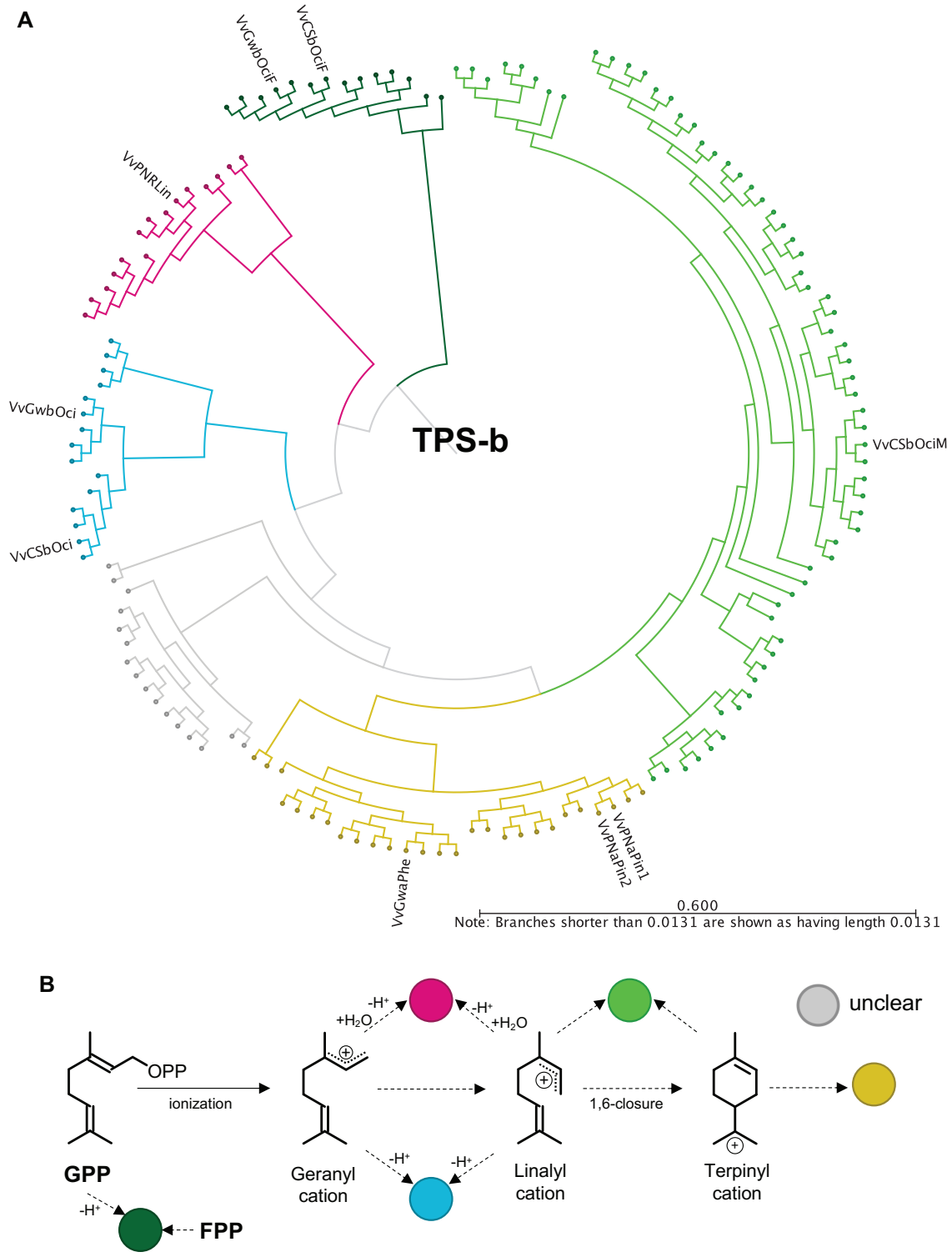

Figure 3.5. (A) Active site phylogeny of putative VviTPS-b proteins from the diploid genomes and functionally characterised enzymes (outer labels) were used to predict the carbocation cascades that proceed from geranyl diphosphate (GPP), as shown in **(B)**. The dark green branches represent enzymes that can utilize both GPP and farnesyl diphosphate (FPP), functioning as single product enzymes to form the monoterpene ocimene or the sesquiterpene (*E,E*)-alpha-farnesene, respectively. The pink branch is represented by a single product enzyme that proceeds through either the geranyl and/or linalyl cations, with a concerted protonation and deprotonation cascade to form the acyclic monoterpene alcohol limonene. Deprotonation of the geranyl and/or linalyl cations will result in the formation of the acyclic monoterpene ocimene, shown by the blue branch. The light green branch is represented by a multiproduct monoterpene synthase with a cascade proceeding from the linalyl cation to produce acyclic terpenes with a 1,6 ring-closure required to form the terpinyl carbocation and subsequent cyclic terpenes. The yellow branch represents multiproduct enzymes that produce only cyclic monoterpenes, with the grey branch representing a clade without a functional enzyme.

Stellenbosch University https://scholar.sun.ac.za

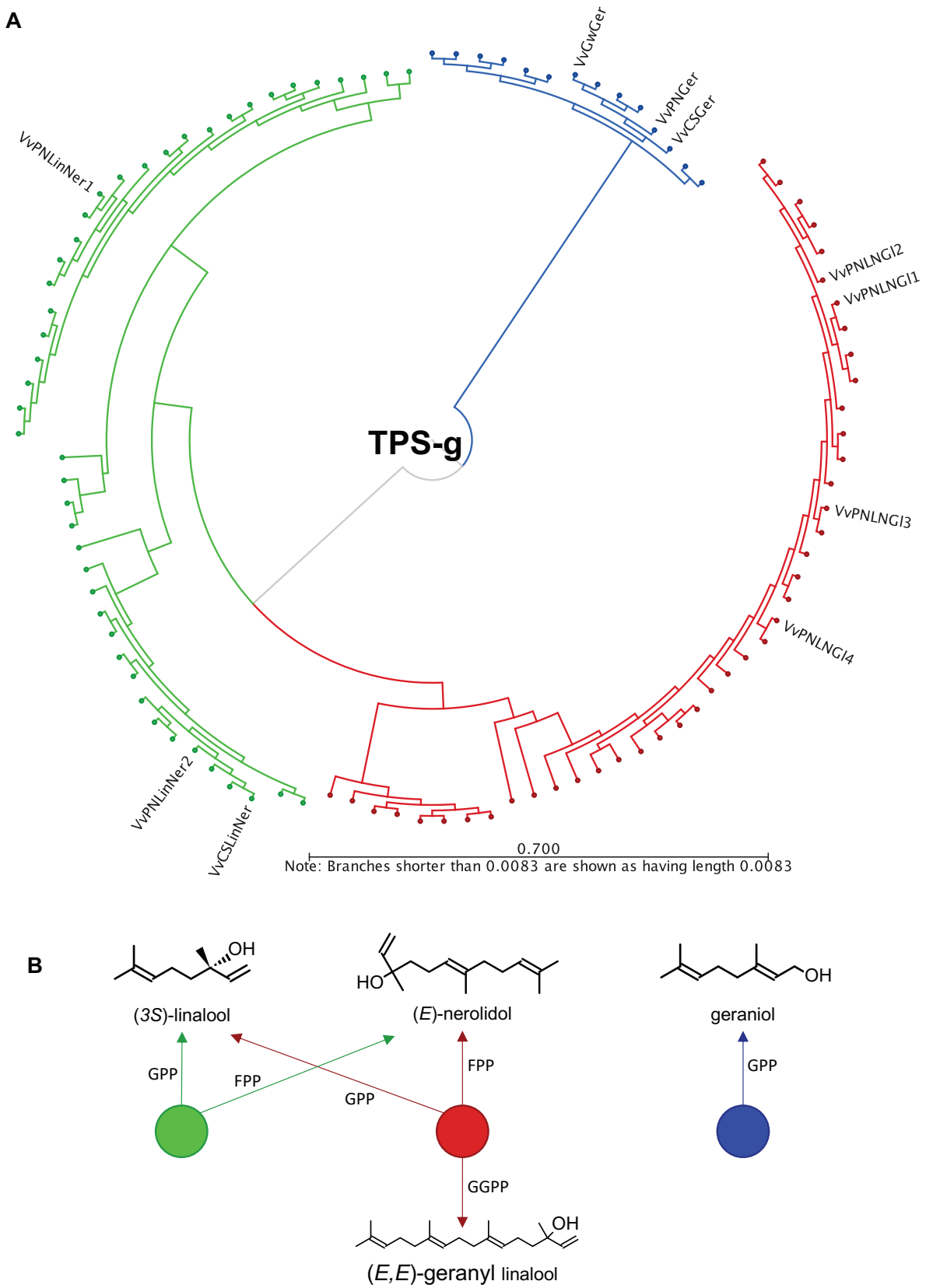

Figure 3.6. (A) Full length amino acid sequence phylogeny of putative VviTPS-g proteins identified for the diploid genomes and functionally characterised VviTPS-g members (outer label). Branches reflect the enzyme product profile and substrate specificity shown in **(B)**. The green clade consists of single product enzymes that utilize both geranyl diphosphate (GPP) and farnesyl diphosphate (FPP) to produce the acyclic monoterpene alcohol (3S)-linalool or acyclic sesquiterpene alcohol (E)-nerolidol, respectively. The red clade represents enzymes that in addition to the mechanism shown in green, accept geranyl geranyl diphosphate (GGPP) to produce the acyclic diterpene alcohol (*E,E*)-geranyl linalool. The blue clade represents enzymes that only accept GPP to form the acyclic monoterpene alcohol geraniol.

Stellenbosch University https://scholar.sun.ac.za

CS had 28 VviTPS-g members that mapped to chr. 10, of which 14 were located in a 262 kb region of the primary contig VvCabSauv08_v1_Primary000201F. Seven of the genes in this cluster were predicted to be functional and are highly connected to genes from seven different haplotigs, all mapping to chr. 10 (Figure 3.3b). Furthermore, the *VviTPS* gene order of the primary contig was dissimilar to the haplotigs with large size differences for the intergenic regions, indicating a high level of heterozygosity for this chromosome. CR had a similar sized TPS-g family on chr. 10, localising to two different primary contigs with almost no contiguity to PN-chr. 10. It was evident from the haplotig to primary contig mappings that chr. 10 is also highly heterozygous for CR. CH, was the exception with 17 of the chr. 10 TPS-g members found on a single contig, connected as tandem duplications in Figure 3.3c, suggesting that it is more homozygous for TPS-g members on chr. 10. As with the other two genomes, this region was highly discontiguous to the reference genome (Appendix C).

To understand the complexity of the *VviTPS* family an integrated view of all the components that influence the family is required. The network in Figure 3.7 shows the *VviTPS* containing chromosomes, gene duplications and putative proteins for the diploid genomes. Contig nodes were excluded from the visualisation, but can be accessed in the interactive network, available online, as described in Appendix D. The three major *VviTPS*-containing chromosomes, namely chr. 13, -18 and -19 show extensive duplications on the respective chromosome with few shared between chromosomes. Although the remaining chromosomes, excluding chr. 10, have few *VviTPS* genes, it is evident that they are extensively connected between chromosomes, specifically the multi-substrate TPS-g family of chr. 10.

An all-against-all clustering of diploid genome putative proteins and functionally characterised proteins is shown in Figure 3.8. The network consists of 533 proteins of which 44 are functionally characterised (Drew et al., 2015; Lücker et al., 2004; Martin et al., 2009, 2010; Smit et al., 2019), sized and shaded in Figure 3.8. To date no VviTPS-c or -e members have been characterised, therefore the three predicted PN40024 members from the respective subfamilies were included as representatives (Martin et al., 2010). The 533 proteins could be clustered into 111 representative sequences (Appendix D), indicted by the triangular nodes. Of the representative sequences, 24 VviTPS-a, 16 VviTPS-b and 7 VviTPS-g sequences were not connected to any other sequence indicating that they are unique.

Stellenbosch University https://scholar.sun.ac.za

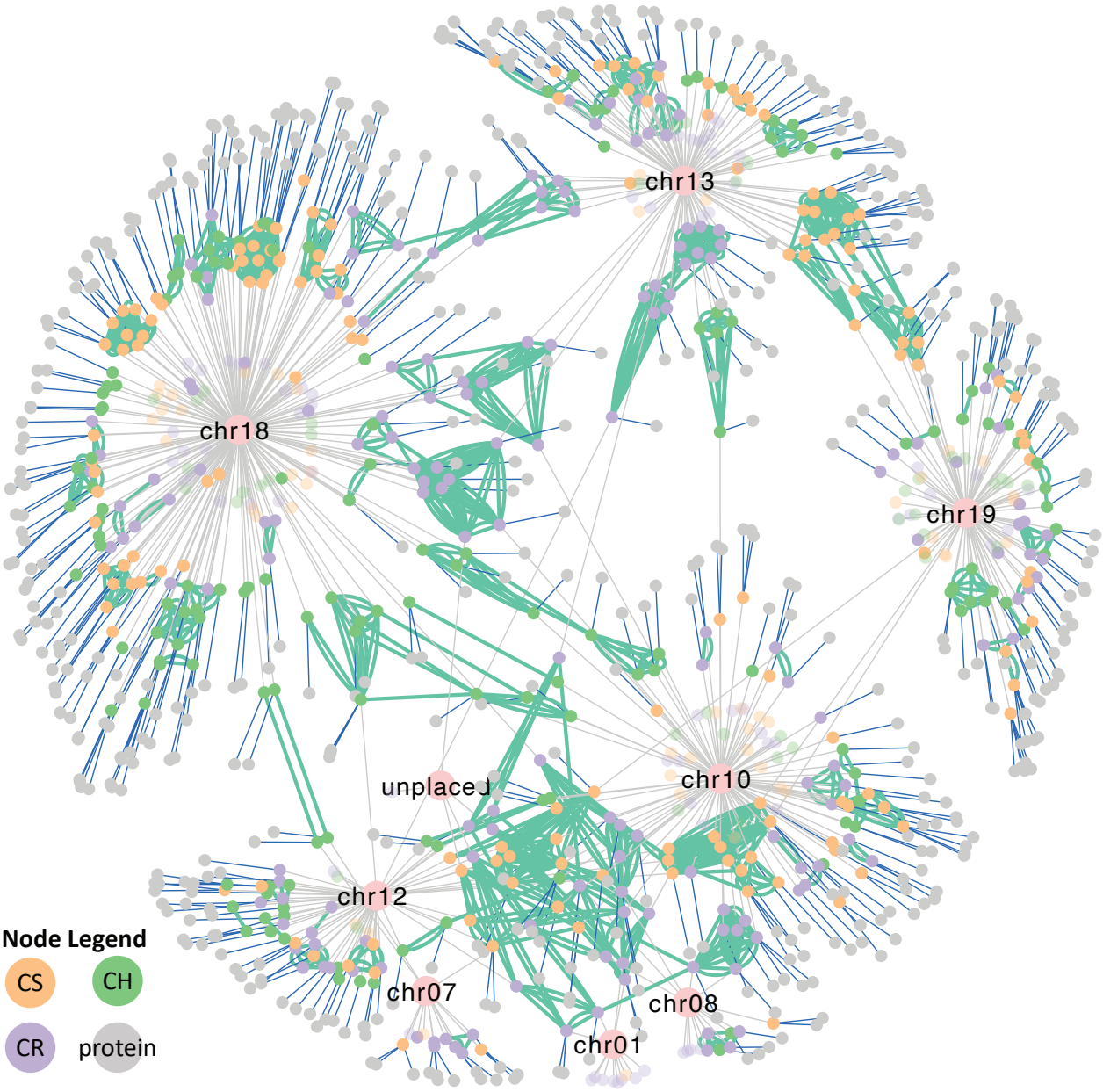

Figure 3.7. Cytoscape network representing all metadata represented in Figures 1-6 of the respective *VviTPS* gene families for Cabernet Sauvignon (CS), Carménère (CR), and Chardonnay (CH). Nodes represent *VviTPS* genes, coloured according to cultivar with the lighter shade showing partial genes. Grey edges show the chromosome mapping (pink nodes), with green edges showing genes with high homology (degree of similarity (I') > 80%). Green edges connecting nodes of the same colour reflect within cultivar homologs (i.e. duplicated genes) while edge connections between different coloured nodes show cross-cultivar homologs. Blue edges show putative proteins associated with a gene (i.e. predicted to be functional). An interactive version of this network to access the metadata is available online.

Stellenbosch University https://scholar.sun.ac.za

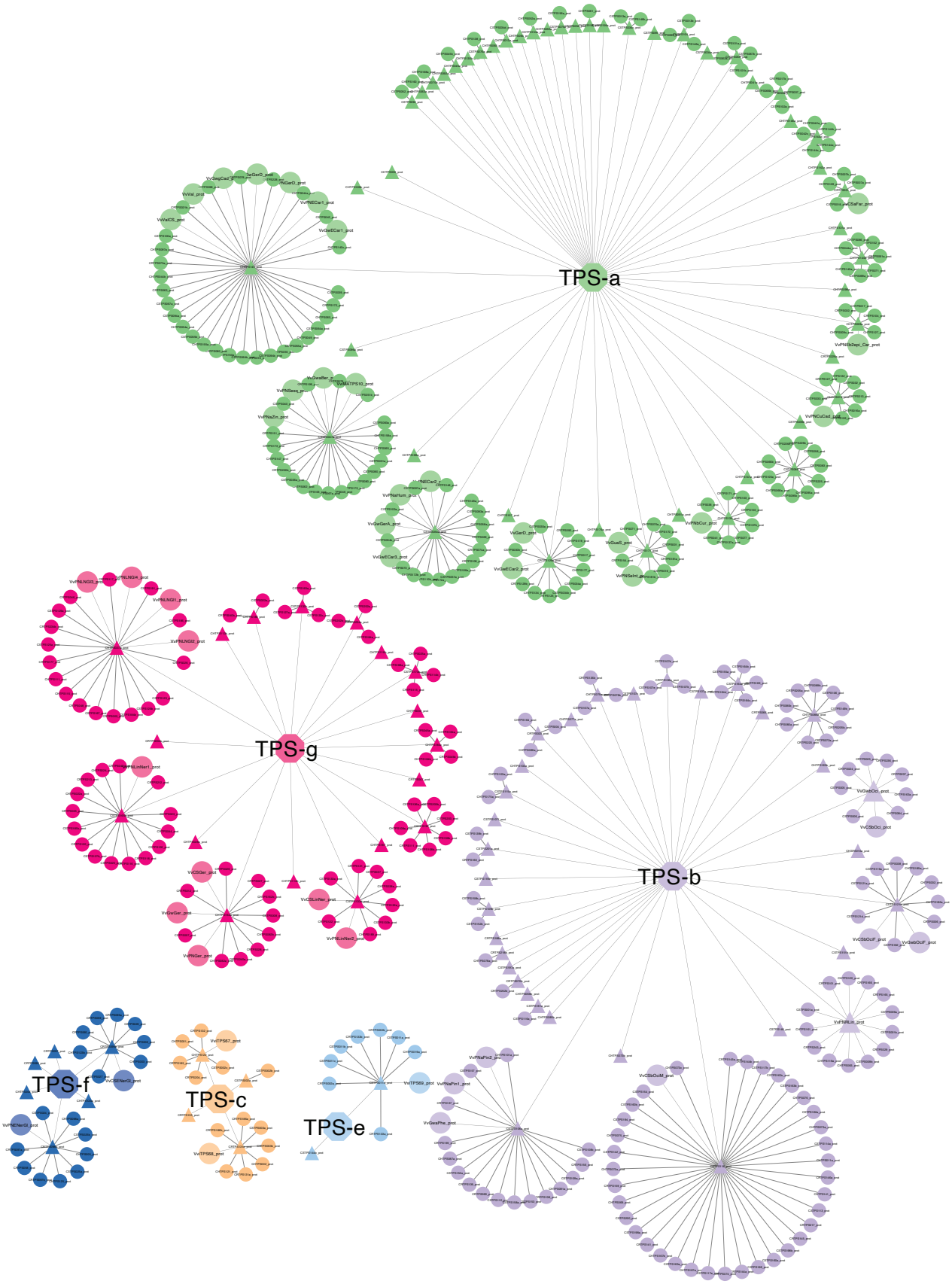

Figure 3.8. Representative proteins (triangles) show clustering of conserved proteins per *VviTPS* subfamily (central nodes with subfamily name). An absence of connections indicate that the sequence is unique. Enlarged circular nodes are functionally characterised enzymes. The representative sequences serve as target sequences through BLASTp in order to query any TPS of interest to identify mechanistically conserved enzyme clusters. To access the metadata through BLASTp refer to the network available online and the accompanying help document, as described in Appendix D.

Stellenbosch University https://scholar.sun.ac.za

3.4.4 The VviTPS compendium: A collection of interactive networks

The data generated in this study are intended to be accessed and mined interactively. Networks can be accessed and downloaded through NDEx (Pratt et al., 2015) using the link provided in Appendix D. Appendix D furthermore contains the representative sequences, illustrated in Figure 3.8, and allow for the BLAST lookup for genes of interest using the "align two or more sequences" function of protein BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi). All nodes and edges in the respective networks are clickable and represent a compendium of the data generated in this study. By interacting with the nodes and edges, a user can find the nearest functionally characterised protein, which includes metadata for NCBI accessions and nearest reference gene model, predicted by Martin et al. (2010), as well as subfamily specific reaction mechanisms. It is therefore possible to query any new gene of interest against the current *VviTPS* gene family for the three diploid genomes and the PN40024 reference genome. A help document to guide users through this, is made available with the networks in Appendix D. The curated genomic, coding and protein sequences represented in the VviTPS compendium are available as FASTA files in Appendix E.

3.5 Discussion

The reference genome highlighted the extensive duplications and functional diversification of the *VviTPS* family (Martin et al., 2010). Although it was stated that paralogous genes are spread across the genome; identifying homologs were not possible. A typical search to find paralogs entails a BLAST search to find a gene of interest, followed by locating it on the genome, usually through a genome browser. Although specialised gene families have been annotated for grapevine (Martin et al., 2010; Vannozzi et al., 2012), delayed incorporation of these annotations into the reference annotation (Canaguier et al., 2017) and limited visibility of these curations often result in an outdated annotation, most commonly 12x.v0, being used (Grimplet and Cramer, 2019). The web interface of EnsemblPlants (Overduin and Ph, 2011) presents the most complete set of tools to analyse the grapevine genome, but still relies in the 12x.v0 assembly and annotation, limiting its use for the *VviTPS* family. The approach presented in this study was akin to that of a pangenome but utilizes a network for data visualisation rather than a genome browser.

Pangenomes typically focus on the differences and similarities between species, however, the genotypes presented here are expected to be highly similar due to it being closely related cultivars of the same species (Appendix B, Figure 1). Although partial gene duplications were annotated (see VviTPS compendium network), their evolutionary importance was not explored further. For the same reason the transposable elements proximal to *VviTPS* genes were excluded. Both of these aspects will become more relevant once the diploid genomes are assembled to chromosomes, allowing for in-depth analysis of collinearity and synteny. Nevertheless, the current unassembled genomes allow for a comparative analysis of the *VviTPS* family. The absolute position of *VviTPS* genes on the diploid genomes was therefore not a focus of this study, but rather how genes are related and how their putative function will impact the genetic potential of a genotype.

Stellenbosch University https://scholar.sun.ac.za

Function inference, is however, not purely based on sequence similarity due to the complexity of the carbocation cascades involved in enzyme functions. The availability of various TPS crystal structures (Caruthers et al., 2000; Gennadios et al., 2009; Lesburg, 1997; Li et al., 2013; Rynkiewicz et al., 2001; Shishova et al., 2007; Starks, 1997; Williams et al., 1998b) and functionality characterised enzymes, combined with quantum mechanical modelling (Bülow et al., 2000; Davis and Croteau, 2000; Durairaj et al., 2019; Gao et al., 2012; Hong and Tantillo, 2014; Miller and Allemann, 2012; O'Brien et al., 2016; Wedler et al., 2015), have contributed to elucidating how these cascades proceed, to produce the thousands of naturally occurring terpene structures (Buckingham et al., 2015; Osbourn and Lanzotti, 2009).

Sequence identity and protein structure homology to experimentally characterised enzymes can therefore be utilized to predict TPS reaction mechanisms (Degenhardt et al., 2009; Durairaj et al., 2019; Smit et al., 2019). However, this approach requires an extensive understanding of TPS reaction mechanisms, which is especially relevant when considering that the presence of a transcript does not necessarily correspond to a functional enzyme; and that enzyme mechanics can differ (within and between genotypes) due to mutations (Drew et al., 2015; Smit et al., 2019). The compendium generated in this study therefore provides a multi-genotype view of the *VviTPS* family, consisting of both gene annotation and functional predictions. The benefits of long-read sequencing, allowing for haplotype resolution, despite being unassembled, must be emphasised as it overcomes erroneous assembly of highly similar and duplicated gene regions that could not be resolved through short-read sequencing. The VviTPS compendium therefore provides a wealth of information to study this family in grapevine.

3.5.1 The grapevine TPS-g family

The VviTPS-g subfamily is of particular interest from a VOC perspective, seeing that this family is responsible for the biosynthesis of terpene alcohols. Functional groupings in Figure 3.6 shows that dual substrate (GPP and FPP) enzymes capable of producing both linalool and nerolidol are overrepresented in all cultivars. Although a large clade of enzymes are predicted to use GGPP as well, resulting in (*E,E*)-geranyl linalool biosynthesis, the ability to use all three substrates *in planta* has not been reported. Subcellular compartmentalisation of precursor pools (IPP and DMAPP) and regulation of prenyl substrate biosynthesis is tightly regulated, resulting in compartment-specific biosynthesis of terpenes (Heinig et al., 2013; Wu et al., 2006). Substrate specificity is thought to be affected by the active site, resulting in differential affinities to GPP, FPP and GGPP when enzymes are studied *in vitro* (Arimura et al., 2007; Pazouki et al., 2015). This was also shown for *PlTPS2* from *Phaseolus lunatus* (lima bean), however, *in planta* expression of this gene resulted in (*E,E*)-geranyl linalool and hemiterpene accumulation (Brillada et al., 2013). It is thus likely that the tri-substrate VviTPS-g clade is involved in (*E,E*)-geranyl linalool biosynthesis rather than (*3S*)-linalool and/or (*E*)-nerolidol biosynthesis.

Stellenbosch University https://scholar.sun.ac.za

The clade for geraniol biosynthesis had only two putatively functional proteins for CS, with CH and CR having 4 and 5, respectively (Figure 3.6). During winemaking, geraniol is readily metabolized by yeast during fermentation to form important wine odorants that, along with nerolidol and linalool derivatives, make up the core constituents of aromatic wines, often described as having a Muscat or "floral" aroma (Emanuelli et al., 2010; King and Dickinson, 2000). These transformations are facilitated by specific yeast genera, for example reduction of the terpenoid or cleavage of glycosyl groups. The available substrate (cultivar-specific terpenoids) and vinification style will therefore directly influence the extent of floral aroma catalysis (Carrau et al., 2005; Cramer et al., 2014).

Furthermore, (*E*)-nerolidol and (*E,E*)-geranyl linalool are known precursors for the homoterpenes (E)-4,8-dimethyl-1,3,7-nonatriene (DMNT) and (E,E)- 4,8,12-trimethyltrideca-1,3,7,11-tetraene (TMTT), respectively. DMNT, is especially important from an ecological perspective due to it being emitted by various grapevine organs, with flower and leaf emissions linked to the attraction of the grapevine berry moth, *Lobesia botrana,* a major grapevine pest (Tasin et al., 2007). Recent efforts to alter the chemical emission profile of grapevine focussed on overexpressing an (*E*)-beta-farnesene synthase, decreasing *L. botrana* attraction to grapevine (Salvagnin et al., 2018). The numerous TPS-g members annotated here therefore provide alternative targets to alter (*E,E*)-geranyl linalool, and by extension DMNT, biosynthesis.

3.5.2 The VviTPS-a and -b subfamilies: an expanded group with specialised reaction mechanisms

The TPS-a and -b subfamilies are hypothesised to have evolved from diterpene synthases where the loss of the γ domain or transit peptide, coupled with changes in the active site, lead to neofunctionalization (Köksal et al., 2011b, 2011a; Pazouki and Niinemets, 2016). This likely allowed for spatial-temporal regulation and specialisation with vestigial functions explaining the ability to use multiple substrates *in vitro* (Pazouki and Niinemets, 2016).

Sesquiterpene synthases (VviTPS-a) represent the largest grapevine subfamily and are of special interest due to their ability to produce either a single terpene or a multitude of compounds. The diversity in sesquiterpene is largely due to the extra double bond in FPP, compared to GPP, with subsequent isomerisation to NPP resulting in further diversity. Currently accepted reaction mechanisms of plant sesquiterpene synthases (Durairaj et al., 2019) resulted in VviTPS-a members grouping according to which of these are used initial substrate. Premature quenching of the cyclisation reaction, regardless of whether FPP or NPP is the initial substrate, results in the formation of acyclic sesquiterpenes, with two small but distinct clades suggesting that there may be a distinction in substrate affinity. The isomerisation step is rate-limiting (Cane et al., 1997; Miller and Allemann, 2012) which could explain why fewer enzymes are in the NPP clade, suggesting a possible specialised *in planta* function. It was previously shown that PN40024 had distinct clades for 1,10 and 1,11-cylisations of FPP (Martin et al., 2010; Smit et al., 2019), however, the increased number of putative VviTPS-a proteins from the three diploid genomes added greater complexity to

Stellenbosch University https://scholar.sun.ac.za

the conservation of enzyme mechanisms. Three distinct clades were identified (Figure 3.6) with the functional enzymes of clades 2 and 3 sharing the same product profiles with a clear distinction between 1,10- and 1,11-cylisations. The 1,10-cylisations will proceed through the (*E,E*)-germacradienyl cation to either germacrene A or D as reactive intermediates. From germacrene A, an alkyl migration of eudesmyl cation will be necessary to explain the mechanism for enzymes in clade 2, while a lack of this migration and the presence of selinene-type synthases allowing for clade 1 enzymes (Calvert et al., 2002; Caruthers et al., 2000; Christianson, 2017). Due to these subtle complexities in enzyme mechanics, VviTPS-a functional predictions is limited to initial substrate and first cyclisation (Figure 3.4).

Although multiple reaction mechanism was identified for clades within the TPS-b subfamily, the overarching differences were between TPS-b Type I and II mechanisms. It was, however, noted that the single product Type II enzymes formed unique clades. This was also reported by Martin et al. (2010) where the two reaction types were bifurcated by sequences from other plants instead of a group of enzymes with no clear function, shown in Figure 3.5. The clades in Figure 3.5, indicate a conserved set of Type II enzymes that seemingly evolved to multi-product Type I enzymes. The clade of proteins associated with enzymes that accept FPP *in vitro* seems to be conserved, with its phylogenetic position supporting the specialisation hypothesis (Pazouki and Niinemets, 2016).

3.6 Conclusion

The availability of new genomic resources allowed for a comparative analysis of the *VviTPS* family, expanding on what the PN40024 genome offered. The resolution of haplotypes allowed for the identification of putative alleles with greater sequence contiguity, due to long-read sequencing, allowing for a comprehensive, and more complete annotation of this expanded gene family. Phylogenomic similarity and functional predictions greatly benefited from having expanded genotypic variation. This allowed for greater subfamily-specific functional predictions while addressing specific limitations on the reference genome. The VviTPS compendium is not intended to be a static resource with the incorporation of inter-varietal and -species variations at genomic and single base-pair levels expected to improve the accuracy of functional predictions. The recent release of a phased *V. riparia* genome (Girollet et al., 2019) and nucleotide variation data from 472 *Vitis* species (Liang et al., 2019), specifically hold great promise for elucidating the impact that domestication and breeding had *VviTPS* evolution, expansion and functionalisation. The utilisation of networks to show relatedness of *VviTPS* genes at the genomic, coding and protein sequence levels within and between cultivars provides a valuable and interactive resource. This resource is intended to provide a starting platform from which genotypic variation can be explored and expanded on to characterise the *VviTPS* family further.

Stellenbosch University https://scholar.sun.ac.za

3.7 References

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., et al. (2019). Fast and accurate reference-guided scaffolding of draft genomes. *bioRxiv*, 519637. doi:10.1101/519637.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–402.

Arimura, G., Garms, S., Maffei, M., Bossi, S., Schulze, B., Leitner, M., et al. (2007). Herbivore-induced terpenoid emission in Medicago truncatula: concerted action of jasmonate, ethylene and calcium signaling. *Planta* 227, 453–464. doi:10.1007/s00425-007-0631-y.

Bick, J. A., and Lange, B. M. (2003). Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: Unidirectional transport of intermediates across the chloroplast envelope membrane. *Arch. Biochem. Biophys.* 415, 146–154. doi:10.1016/S0003-9861(03)00233-9.

Bloch, K., Chaykin, S., Phillips, A. ., and De Waard, A. (1959). Mevalonic acid pyrophospahte and isopentenylpyrphosphate. *J. Biol. Chemsitry* 234, 2595–2604.

Bohlmann, J., Meyer-Gauen, G., and Croteau, R. (1998). Plant terpenoid synthases: Molecular biology and phylogenetic analysis. *Proc. Natl. Acad. Sci. U. S. A.* 95, 4126–33. doi:10.1073/pnas.95.8.4126.

Brillada, C., Nishihara, M., Shimoda, T., Garms, S., Boland, W., Maffei, M. E., et al. (2013). Metabolic engineering of the C16 homoterpene TMTT in Lotus japonicus through overexpression of (E,E)-geranyllinalool synthase attracts generalist and specialist predators in different manners. *New Phytol.* 200, 1200–1211. doi:10.1111/nph.12442.

Buckingham, J., Cooper, C. M., and Purchase, R. (2015). *Natural Products Desk Reference*. Boca Raton: CRC Press.

Bülow, N., König, W., and Konig, W. (2000). The role of germacrene D as a precursor in sesquiterpene biosynthesis: investigations of acid catalyzed, photochemically and thermally induced rearrangements. *Phytochemistry* 55, 141–68. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11065290.

Calvert, M. J., Ashton, P. R., and Allemann, R. K. (2002). Germacrene A is a product of the aristolochene synthase-mediated conversion of farnesylpyrophosphate to aristolochene. *J. Am. Chem. Soc.* 124, 11636–11641. doi:10.1021/ja020762p.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421. doi:10.1186/1471-2105-10-421.

Canaguier, A., Grimplet, J., Gaspero, G. Di, Scalabrin, S., Duchêne, E., and Choisne, N. (2017). A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3). *Genomics Data* 14, 56–62. doi:10.1016/J.GDATA.2017.09.002.

Cane, D. E. (1990). Enzymic formation of sesquiterpenes. *Chem. Rev.* 90, 1089–1103. doi:10.1021/cr00105a002.

Cane, D. E., Chiu, H. T., Liang, P. H., and Anderson, K. S. (1997). Pre-steady-state kinetic analysis of the trichodiene synthase reaction pathway. *Biochemistry* 36, 8332–8339. doi:10.1021/bi963018o.

Carrau, F. M., Medina, K., Boido, E., Farina, L., Gaggero, C., Dellacassa, E., et al. (2005). De novo synthesis of monoterpenes by Saccharomyces cerevisiae wine yeasts. *FEMS Microbiol. Lett.* 243, 107–15. doi:10.1016/j.femsle.2004.11.050.

Caruthers, J. M., Kang, I., Rynkiewicz, M. J., Cane, D. E., and Christianson, D. W. (2000). Crystal Structure Determination of Aristolochene Synthase from the Blue Cheese Mold , Penicillium roqueforti *. 275, 25533–25539. doi:10.1074/jbc.M000433200.

Chin, C., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* 13, 1050–1054. doi:10.1038/nmeth.4035.

Christianson, D. W. (2006). Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev.* 106, 3412–3442. doi:10.1021/cr050286w.

Christianson, D. W. (2017). Structural and chemical biology of terpenoid cyclases. *Chem. Rev.* 117. doi:10.1021/acs.chemrev.7b00287.

Cramer, G. R., Ghan, R., Schlauch, K. a, Tillett, R. L., Heymann, H., Ferrarini, A., et al. (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. doi:10.1186/s12870-014-0370-8.

D'Onofrio, C., Matarese, F., and Cuzzola, A. (2017). Study of the terpene profile at harvest and during berry development of Vitis vinifera L. aromatic varieties Aleatico, Brachetto, Malvasia di Candia aromatica and Moscato bianco. *J. Sci. Food Agric.* 97, 2898–2907. doi:10.1002/jsfa.8126.

Davis, E. M., and Croteau, R. (2000). Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes. *Top. Curr. Chem.* 209, 53–95.

Degenhardt, J., Köllner, T. G., and Gershenzon, J. (2009). Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* 70, 1621–1637.

Stellenbosch University https://scholar.sun.ac.za

doi:10.1016/j.phytochem.2009.07.030.
Dèrozier, S., Samson, F., Tamby, J.-P., Guichard, C., Brunaud, V., Grevet, P., et al. (2011). Exploration of plant genomes in the FLAGdb++ environment. *Plant Methods* 7, 8. doi:10.1186/1746-4811-7-8.
Drew, D. P., Andersen, T. B., Sweetman, C., Møller, B. L., Ford, C., and Simonsen, H. T. (2015). Two key polymorphisms in a newly discovered allele of the Vitis vinifera TPS24 gene are responsible for the production of the rotundone precursor α-guaiene. *J. Exp. Bot.* 67, 799–808. doi:10.1093/jxb/erv491.
Dueholm, B., Drew, D. P., Sweetman, C., and Simonsen, H. T. (2019). In planta and in silico characterization of five sesquiterpene synthases from Vitis vinifera (cv. Shiraz) berries. *Planta* 249, 59–70. doi:10.1007/s00425-018-2986-7.
Durairaj, J., Di Girolamo, A., Bouwmeester, H. J., de Ridder, D., Beekwilder, J., and van Dijk, A. D. (2019). An analysis of characterized plant sesquiterpene synthases. *Phytochemistry* 158, 157–165. doi:10.1016/j.phytochem.2018.10.020.
Edgar, R. C. (2004a). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 1–19. doi:10.1186/1471-2105-5-113.
Edgar, R. C. (2004b). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acid Res.* 32, 1792–1797. doi:10.1093/nar/gkh340.
Emanuelli, F., Battilana, J., Costantini, L., Le Cunff, L., Boursiquot, J.-M., This, P., et al. (2010). A candidate gene association study on muscat flavor in grapevine (Vitis vinifera L.). *BMC Plant Biol.* 10, 241. doi:10.1186/1471-2229-10-241.
Figueroa-Balderas, R., Minio, A., Morales-Cruz, A., Vondras, A. M., and Cantu, D. (2019). “Strategies for Sequencing and Assembling Grapevine Genomes,” in *The Grape Genome*, eds. D. Cantu and M. A. Walker (Cham: Springer International Publishing), 77–88. doi:10.1007/978-3-030-18601-2_5.
Gao, Y., Honzatko, R. B., and Peters, R. J. (2012). Terpenoid synthase structures: a so far incomplete view of complex catalysis. *Nat. Prod. Rep.* 29, 1153–75. doi:10.1039/c2np20059g.
Gennadios, H. A., Gonzalez, V., Di Costanzo, L., Li, A., Yu, F., Miller, D. J., et al. (2009). Crystal structure of (+)-δ-cadinene synthase from Gossypium arboreum and evolutionary divergence of metal binding motifs for catalysis. *Biochemistry* 48, 6175–6183. doi:10.1021/bi900483b.
Girollet, N., Rubio, B., and Bert, P.-F. (2019). De novo phased assembly of the Vitis riparia grape genome. *Sci. Data* 6, 1–8. doi:10.1038/s41597-019-0133-3.
Grimplet, J., and Cramer, G. R. (2019). “The Grapevine Genome Annotation,” in *The Grape Genome*, eds. D. Cantu and M. A. Walker (Cham: Springer International Publishing), 89–101. doi:10.1007/978-3-030-18601-2_6.
Heinig, U., Gutensohn, M., Dudareva, N., and Aharoni, A. (2013). The challenges of cellular compartmentalization in plant metabolic engineering. *Curr. Opin. Biotechnol.* 24, 239–246. doi:10.1016/j.copbio.2012.11.006.
Hemmerlin, A., Hoeffler, J. F., Meyer, O., Tritsch, D., Kagan, I. A., Grosdemange-Billiard, C., et al. (2003). Cross-talk between the cytosolic mevalonate and the plastidial methylerythritol phosphate pathways in tobacco bright yellow-2 cells. *J. Biol. Chem.* 278, 26666–26676. doi:10.1074/jbc.M302526200.
Hjelmeland, A. K., Zweigenbaum, J., and Ebeler, S. E. (2015). Profiling monoterpenol glycoconjugation in Vitis vinifera L. cv. Muscat of Alexandria using a novel putative compound database approach, high resolution mass spectrometry and collision induced dissociation fragmentation analysis. *Anal. Chim. Acta* 887, 138–147. doi:10.1016/j.aca.2015.06.026.
Hong, Y. J., and Tantillo, D. J. (2014). Branching out from the bisabolyl cation. unifying mechanistic pathways to barbatene, bazzanene, chamigrene, chamipinene, cumacrene, cuprenene, dunniene, isobazzanene, iso-γ-bisabolene, isochamigrene, laurene, microbiotene, sesquithujene, sesquisabinene, t. *J. Am. Chem. Soc.* 136, 2450–2463. doi:10.1021/ja4106489.
Hyatt, D. C., Youn, B., Zhao, Y., Santhamma, B., Coates, R. M., Croteau, R. B., et al. (2007). Structure of limonene synthase, a simple model for terpenoid cyclase catalysis. *Proc. Natl. Acad. Sci. U. S. A.* 104, 5360–5365. doi:10.1073/pnas.0700915104.
Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449, 463–7. doi:10.1038/nature06148.
Jukes, T. H., and Cantor, C. R. (1969). “Evolution of Protein Molecules,” in *Mammalian Protein Metabolism*, ed. H. N. Munro (Elsevier), 21–132. doi:10.1016/B978-1-4832-3211-9.50009-7.
King, A., and Dickinson, J. R. (2000). Biotransformation of monoterpene alcohols by Saccharomyces cerevisiae, Torulaspora delbrueckii and Kluyveromyces lactis. *Yeast* 16, 499–506. doi:10.1002/(SICI)1097-0061(200004)16:6<499::AID-YEA548>3.0.CO;2-E.
Köksal, M., Hu, H., Coates, R. M., Peters, R. J., and Christianson, D. W. (2011a). Structure and mechanism of the diterpene cyclase ent-copalyl diphosphate synthase. *Nat. Chem. Biol.* 7, 431–433. doi:10.1038/nchembio.578.
Köksal, M., Jin, Y., Coates, R. M., Croteau, R., and Christianson, D. W. (2011b). Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature* 469, 116–122. doi:10.1038/nature09628.

Stellenbosch University https://scholar.sun.ac.za

Larsson, J. (2019). eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. Available at: https://cran.r-project.org/package=eulerr.
Lesburg, C. A. (1997). Crystal Structure of Pentalenene Synthase: Mechanistic Insights on Terpenoid Cyclization Reactions in Biology. *Science (80-.).* 277, 1820–1824. doi:10.1126/science.277.5333.1820.
Li, J.-X., Fang, X., Zhao, Q., Ruan, J.-X., Yang, C.-Q., Wang, L.-J., et al. (2013). Rational engineering of plasticity residues of sesquiterpene synthases from *Artemisia annua* : product specificity and catalytic efficiency. *Biochem. J.* 451, 417–426. doi:10.1042/BJ20130041.
Li, W. H., Gu, Z., Wang, H., and Nekrutenko, A. (2001). Evolutionary analyses of the human genome. *Nature* 409, 847–849. doi:10.1038/35057039.
Liang, Z., Duan, S., Sheng, J., Zhu, S., Ni, X., Shao, J., et al. (2019). Whole-genome resequencing of 472 Vitis accessions for grapevine diversity and demographic history. *Nat. Commun.*, 1–12. doi:10.1038/s41467-019-09135-8.
Lichtenthaler, H. K. (1999). The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50, 47–65. doi:10.1146/annurev.arplant.50.1.47.
Lücker, J., Bowen, P., and Bohlmann, J. (2004). Vitis vinifera terpenoid cyclases: functional identification of two sesquiterpene synthase cDNAs encoding (+)-valencene synthase and (−)-germacrene D synthase and expression of mono- and sesquiterpene synthases in grapevine flowers and berries. *Phytochemistry* 65, 2649–2659. doi:10.1016/j.phytochem.2004.08.017.
Martin, D., Aubourg, S., Schouwey, M., Daviet, L., Schalk, M., Toub, O., et al. (2010). Functional Annotation, Genome Organization and Phylogeny of the Grapevine (Vitis vinifera) Terpene Synthase Gene Family Based on Genome Assembly, FLcDNA Cloning, and Enzyme Assays. *BMC Plant Biol.* 10, 226. doi:10.1186/1471-2229-10-226.
Martin, D. M., Toub, O., Chiang, A., Lo, B. C., Ohse, S., Lund, S. T., et al. (2009). The bouquet of grapevine (Vitis vinifera L. cv. Cabernet Sauvignon) flowers arises from the biosynthesis of sesquiterpene volatiles in pollen grains. *Proc. Natl. Acad. Sci. U. S. A.* 106, 7245–7250. doi:10.1073/pnas.0901387106.
Martin, S., and Leggett, R. M. (2019). Alvis: a tool for contig and read ALignment VISualisation and chimera detection. *bioRxiv*, 663401. doi:10.1101/663401.
Maul, E., and Töpfer, R. (2015). Vitis International Variety Catalogue (V IVC): A cultivar database referenced by genetic profiles and morphology. *BIO Web Conf.* doi:10.1051/bioconf/20150501009.
Miller, D. J., and Allemann, R. K. (2012). Sesquiterpene synthases: Passive catalysts or active players? *Nat. Prod. Rep.* 29, 60–71. doi:10.1039/C1NP00060H.
Minio, A., Lin, J., Gaut, B. S., and Cantu, D. (2017). How Single Molecule Real-Time Sequencing and Haplotype Phasing Have Enabled Reference-Grade Diploid Genome Assembly of Wine Grapes. *Front. Plant Sci.* 8, 1–6. doi:10.3389/fpls.2017.00826.
Minio, A., Massonnet, M., Figueroa-Balderas, R., Castro, A., and Cantu, D. (2019). Diploid Genome Assembly of the Wine Grape Carménère. *G3: Genes|Genomes|Genetics* 9, g3.400030.2019. doi:10.1534/g3.119.400030.
Myles, S., Boyko, A. R., Owens, C. L., Brown, P. J., Grassi, F., Aradhya, M. K., et al. (2011). Genetic structure and domestication history of the grape. *Proc. Natl. Acad. Sci.* 108, 3530–3535. doi:doi.org/10.1073/pnas.1009363108.
O'Brien, T. E., Bertolani, S. J., Tantillo, D. J., and Siegel, J. B. (2016). Mechanistically informed predictions of binding modes for carbocation intermediates of a sesquiterpene synthase reaction. *Chem. Sci.* 7, 4009–4015. doi:10.1039/C6SC00635C.
Osbourn, A. E., and Lanzotti, V. (2009). *Plant-derived Natural Products*. , eds. A. E. Osbourn and V. Lanzotti New York, NY: Springer US doi:10.1007/978-0-387-85498-4.
Overduin, B., and Ph, D. (2011). Browsing Genes and Genomes with Ensembl. *Mol. Biol.*
Pazouki, L., Memari, H. R., Kännaste, A., Bichele, R., and Niinemets, Ü. (2015). Germacrene A synthase in yarrow (Achillea millefolium) is an enzyme with mixed substrate specificity: gene cloning, functional characterization and expression analysis. *Front. Plant Sci.* 6, 111. doi:10.3389/fpls.2015.00111.
Pazouki, L., and Niinemets, Ü. (2016). Multi-Substrate Terpene Synthases: Their Occurrence and Physiological Significance. *Front. Plant Sci.* 7, 1–16. doi:10.3389/fpls.2016.01019.
Pratt, D., Chen, J., Welker, D., Rivas, R., Pillich, R., Rynkov, V., et al. (2015). NDEx, the Network Data Exchange. *Cell Syst.* 1, 302–305. doi:10.1016/j.cels.2015.10.001.
R Core Team (2013). R: A Language and Environment for Statistical Computing. Available at: http://www.r-project.org/.
Roach, M. J., Johnson, D. L., Bohlmann, J., van Vuuren, H. J. J., Jones, S. J. M., Pretorius, I. S., et al. (2018). Population sequencing reveals clonal diversity and ancestral inbreeding in the grapevine cultivar Chardonnay. *PLoS Genet.*, 1–24. doi:https://doi.org/10.1371/journal.pgen.1007807.
Robinson, A. L., Boss, P. K., Solomon, P. S., Trengove, R. D., Heymann, H., and Ebeler, S. E. (2014). Origins of grape and wine aroma. Part 1. Chemical components and viticultural impacts. *Am. J. Enol. Vitic.* 65, 1–24. doi:10.5344/ajev.2013.12070.
Rohmer, M. (1999). The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in

Stellenbosch University https://scholar.sun.ac.za

bacteria, algae and higher plants. *Nat. Prod. Rep.* 16, 565–574. doi:10.1039/a709175c.
Rynkiewicz, M. J., Cane, D. E., and Christianson, D. W. (2001). Structure of trichodiene synthase from Fusarium sporotrichioides provides mechanistic inferences on the terpene cyclization cascade. *Proc. Natl. Acad. Sci.* 98, 13543–13548. doi:10.1073/pnas.231313098.
Rynkiewicz, M. J., Cane, D. E., and Christianson, D. W. (2002). X-ray crystal structures of D100E trichodiene synthase and its pyrophosphate complex reveal the basis for terpene product diversity. *Biochemistry* 41, 1732–1741. doi:10.1021/bi011960g.
Salvagnin, U., Malnoy, M., Thöming, G., Tasin, M., Carlin, S., Martens, S., et al. (2018). Adjusting the scent ratio: using genetically modified Vitis vinifera plants to manipulate European grapevine moth behaviour. *Plant Biotechnol. J.* 16, 264–271. doi:10.1111/pbi.12767.
Schwab, W., Williams, D. C., Davis, E. M., and Croteau, R. (2001). Mechanism of monoterpene cyclization: Stereochemical aspects of the transformation of noncyclizable substrate analogs by recombinant (-)-limonene synthase, (+)-bornyl diphosphate synthase, and (-)-pinene synthase. *Arch. Biochem. Biophys.* 392, 123–136. doi:10.1006/abbi.2001.2442.
Schwab, W., and Wüst, M. (2015). Understanding the constitutive and induced biosynthesis of mono- and sesquiterpenes in grapes (Vitis vinifera) – A key to unlocking the biochemical secrets of unique grape aroma profiles. *J. Agric. Food Chem.* 63, 10591–10603. doi:10.1021/acs.jafc.5b04398.
Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi:10.1101/gr.1239303.
Shishova, E. Y., Di Costanzo, L., Cane, D. E., and Christianson, D. W. (2007). X-ray crystal structure of aristolochene synthase from Aspergillus terreus and evolution of templates for the cyclization of farnesyl diphosphate. *Biochemistry* 46, 1941–1951. doi:10.1021/bi0622524.
Slater, G. S. C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6, 1–11. doi:10.1186/1471-2105-6-31.
Smit, S. J., Vivier, M. A., and Young, P. R. (2019). Linking Terpene Synthases to Sesquiterpene Metabolism in Grapevine Flowers. *Front. Plant Sci.* 10, 1–18. doi:10.3389/fpls.2019.00177.
Starks, C. M. (1997). Structural Basis for Cyclic Terpene Biosynthesis by Tobacco 5-Epi-Aristolochene Synthase. *Science (80-.).* 277, 1815–1820. doi:10.1126/science.277.5333.1815.
Steinegger, M., and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* 35, 1026–1028. doi:10.1038/nbt.3988.
Swiegers, J. H., Bartowsky, E. J., Henschke, P. A., and Pretorius, I. S. (2005). Yeast and bacterial modulation of wine aroma and flavour. *Aust. J. Grape Wine Res.* 11, 139–173. doi:10.1111/j.1755-0238.2005.tb00285.x.
Tasin, M., Bäckman, A.-C., Coracini, M., Casado, D., Ioriatti, C., and Witzgall, P. (2007). Synergism and redundancy in a plant volatile blend attracting grapevine moth females. *Phytochemistry* 68, 203–9. doi:10.1016/j.phytochem.2006.10.015.
This, P., Lacombe, T., and Thomas, M. R. (2006). Historical origins and genetic diversity of wine grapes. *Trends Genet.* 22, 511–519. doi:10.1016/j.tig.2006.07.008.
Vannozzi, A., Dry, I. B., Fasoli, M., Zenoni, S., and Lucchin, M. (2012). Genome-wide analysis of the grapevine stilbene synthase multigenic family: genomic organization and expression profiles upon biotic and abiotic stresses. *BMC Plant Biol.* 12, 130. doi:10.1186/1471-2229-12-130.
Velasco, R., Zharkikh, A., Troggio, M., Cartwright, D. a, Cestaro, A., Pruss, D., et al. (2007). A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS One*, e1326. doi:10.1371/journal.pone.0001326.
Wedler, H., Pemberton, R., and Tantillo, D. (2015). Carbocations and the complex flavor and bouquet of wine: mechanistic aspects of terpene biosynthesis in wine grapes. *Molecules* 20, 10781–10792. doi:10.3390/molecules200610781.
Williams, D. C., McGarvey, D. J., Katahira, E. J., and Croteau, R. (1998a). Truncation of limonene synthase preprotein provides a fully active “pseudomature” form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair. *Biochemistry* 37, 12213–12220. doi:10.1021/bi980854k.
Williams, D. C., McGarvey, D. J., Katahira, E. J., and Croteau, R. (1998b). Truncation of limonene synthase preprotein provides a fully active “pseudomature” form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair. *Biochemistry* 37, 12213–12220. doi:10.1021/bi980854k.
Wu, S., Schalk, M., Clark, A., Miles, R. B., Coates, R., and Chappell, J. (2006). Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants. *Nat. Biotechnol.* 24, 1441–1447. doi:10.1038/nbt1251.
Xu, J., Xu, J., Ai, Y., Farid, R. A., Tong, L., and Yang, D. (2018). Mutational analysis and dynamic simulation of S-limonene synthase reveal the importance of Y573: Insight into the cyclization mechanism in monoterpene synthases. *Arch. Biochem. Biophys.* 638, 27–34. doi:10.1016/j.abb.2017.12.007.

Stellenbosch University https://scholar.sun.ac.za

Chapter 4

Seeing the forest through the (phylogenetic) trees: functional characterisation of *VviTPS* paralogues and orthologues.

This chapter will be submitted for peer review, once Chapter 3 has been accepted.

Stellenbosch University https://scholar.sun.ac.za

Seeing the forest through the (phylogenetic) trees: functional characterisation of *VviTPS* paralogues and orthologues.

4.1 Introduction

The *Vitis vinifera* reference genome is an invaluable resource for the identification of genes involved in metabolic processes of agronomic interest. Th genetic basis for various traits associated with flavour and aroma have been elucidated since the genome became available, as recently reviewed in (Lin et al., 2019). A major contributor to flavour and aroma profiles are the terpenes, a chemically diverse class of metabolites, mainly associated with floral and Muscat (linalool, geraniol, nerol, α-terpineol, and hotrienol) or pepper (rotundone) aromas inferred by mono -and sesquiterpenes, respectively. The near-homozygous reference genome PN40024 contains 152 *VviTPS*-like loci, of which 69 are predicted to be functional. Thirty gene models were subsequently linked to a functional enzyme, but were isolated from different genotypes (Martin et al., 2010).

Genotype-specific structural variations (SV) are known to have a significant impact on the functionality of a VviTPS (Drew et al., 2015; Dueholm et al., 2019; Smit et al., 2019). The homozygosity (~93%) of PN40024, however, does not reflect the structural complexity associated with the commercial diploid cultivars Cabernet Sauvignon (CS), Chardonnay (CH) or Carménère (CR) (Chin et al., 2016; Minio et al., 2019a; Roach et al., 2018). The extensive duplication of *VviTPS* genes, furthermore, result in numerous genes having similar function with various mechanisms potentially involved in these duplication events (Carretero-Paulet et al., 2015; Del Pozo and Ramirez-Parra, 2015; Martin et al., 2010; Smit et al., 2019).

The proximity (closeness) of related *VviTPS*-subfamily members on a chromosome are indicative of tandem duplication events, whereas remnants of transposable elements suggest that transposon-mediated duplication allowed for genome-wide movement of *VviTPS* genes. Segmental duplications that are highly homologous (94% and >10 kb in size) contribute to 17% of the PN40024 genome, resulting in large repetitive genomic regions (Giannuzzi et al., 2011). Domestication driven introgression between germplasms has furthermore resulted in highly complex genomes that show greater heterozygosity in the domesticated varieties than the wild parent(s) (Aradhya et al., 2003; Laucou et al., 2018; Salmaso et al., 2004; Zhou et al., 2019).

Tandem TPS duplications often form multi-gene clusters of varying size, but similar catalytic mechanisms. In tomato (*Solanum lycopersicum*), for example, a sesqui-TPS cluster of six genes consists of two pseudogenes and characterised genes with divergent catalytic active sites (Falara et al., 2011). In rice (*Oryza spp.),* the evolution of *TPS* gene clusters involved in insect defence showed species-specific cluster differences that went through a combination of duplication and subsequent functionalisation events. These gene clusters furthermore showed species specific expression and emission patterns in response to insect infestation (Chen et al., 2019). A 690 kb region of PN40024 chr18 contains 44/152 *VviTPS*-like loci, 20 encoding for putative sesquiterpene

Stellenbosch University https://scholar.sun.ac.za

synthases (*VviTPS*01-19 and -30). Eleven of these gene have been functionally characterised (Martin et al., 2010; Smit et al., 2019). The annotation of the *VviTPS* families for CH, CS and CR, however, revealed that there are potential structural differences in the genome space for the chr18 cluster (Chapter 3). To date the relatedness of *VviTPS* paralogs in this cluster has not been explored.

Of particular interest is *VviTPS10*, a gene model linked to two distinct enzyme functions with the VvGwaBer nucleotide sequence being concordant to the gene model. VviMATPS10, however, showed various nonsynonymous mutations that resulted in a novel function (Smit et al., 2019). The extent of these sequence differences presented two possibilities: VviTPS10 has two alleles with unique functions, or a highly conserved paralog inferring the second function. It is important to note the current limitations of how the *VviTPS* family is annotated. Various version changes of the reference annotation have been merged into the latest VCost.v3 annotation, resulting in a new unique locus ID. Historical IDs from previous versions as well as the VviTPS ID are maintained for congruency between the different versions. To address possible confusion the International Grape Genome Program (IGGP) released a gene nomenclature guideline which is recommended to guide future gene annotations (Grimplet et al., 2014). However, these guidelines are not fully amenable for annotating paralogous genes and structural differences between allelic regions (i.e. alleles and/or paralogues on heterozygous regions).

The unique TPS loci ID used for the respective diploid genomes in Chapter 3, combined with a network view to show paralogy provides a solution for this. In this manner the locus ID is static to each cultivar with edge connectivity reflecting either sequence paralogy or conserved enzyme function. However, it is still limited in its ability to show chromosomal positions and proximity of duplicates. This should however only be a temporary problem until the diploid genomes are assembled.

In this study we compared genes that cluster on chr18 based on the annotations in the VviTPS compendium. The VviTPS10 paralogues were targeted in nine *V. vinifera* genotypes (cultivars) for sequence and functional characterisation in order to investigate the extent of paralogy and possible mutations that could impact functionality.

4.2 Materials and methods

4.2.1 Identification of paralogs using diploid grapevine genomes

VviTPS04 and *-10* gene models are highly homologous on the PN40024 genome, positioned 262 kb apart on chr18. *VviTPS10* was previously shown to have multiple putative paralogs on the CS genome (Smit et al., 2019). The analysis of paralogs was extended to the CH genome, and the CS genome re-evaluated using the respective curated gene models for CH and CS (discussed in Chapter 3). CR was not initially included due to genome data not being available at the onset of this study. The *VviTPS04*- and *-10* gDNA sequences were retrieved from FLAGdb++ (Dèrozier et al., 2011) and queried with BLAST (Camacho et al., 2009) to identify their homologous gDNA sequences

Stellenbosch University https://scholar.sun.ac.za

in CH and CS. These sequences were aligned with MUSCLE (Edgar, 2004b) followed by phylogenetic tree construction with UPGMA as linkage method (Edgar, 2004a; Sokal, 1958), Jukes-Cantor as distance measure (Jukes and Cantor, 1969) and 100 bootstrap replicates, using CLC Main Workbench 7 (CLC Bio-Qiagen, Denmark).

4.2.2 Isolation and cloning of *VviTPS04 and -10* paralogs

Primers were designed to target putative paralogs: Forward Primer A (*VviTPS10* paralogs): 5'-ATGGCCTTAATTCTCGCTACCAGCAACGGG-3', Forward primer B (*VviTPS04* paralogs): 5'-ATGTTGAGTGCTCGAGTGAGTTTACACATGGCC-3') and a single conserved reverse primer: 5'-TCATATTGGCACAGGGTCTATGAGCAGCATTGAAATATT-3'.

RNA extraction from flowers of nine *Vitis vinifera* genotypes was performed using the Spectrum Plant Total RNA Kit (Sigma-Aldrich, United States) with on-column DNase I (Thermo Fisher Scientific, United States) treatment. RNA was checked for genomic DNA (gDNA) contamination through PCR, followed by reverse strand synthesis (cDNA) using SuperScript VILO Master Mix (Thermo Fisher Scientific, United States). Genomic DNA (gDNA) was isolated from the same genotypes according to the method described by Reid et al. (2006). Flower genotypes used for gene isolation are described in Chapter 2 with the cultivars abbreviations maintained, namely: Chardonnay (CH), Chenin Blanc (CB), Muscat of Alexandria (MA), Pinot noir (PN), Pinotage (PI), Sauvignon Blanc (SB), Shiraz (SH), Viognier (VG), and Weisser Riesling (WR).

PCR reactions using cDNA or gDNA as template was performed using TaKARA ExTaq proofreading polymerase, as per the product specifications (Separations, RSA). PCR reaction conditions were as follow: 2 min of initial denaturation, followed by 35 cycles of 30 sec denaturation, 30 sec annealing and 2 min 30 sec extension with final extension incubation of 7min. cDNA amplicons were purified after electrophoretic separation using the Zymoclean Gel DNA Recovery Kit (Inqaba Biotech, South Africa), TA cloned, screened for positives through colony PCR and plasmids isolated from overnight cultures. Plasmid DNA was diluted to 1/50 and used as template for two-step Gateway PCR with primers modified accordingly to build in *attB* sites as described in the Gateway Technology Manual (Thermo Fisher Scientific, United States). Gateway cloning was performed according to the aforementioned manual using the appropriate clonases to generate entry clones with pDONR-Zeocin (Thermo Fisher Scientific, United States) and expression clones with pEAQ-HT-DEST1 (Peyret and Lomonossoff, 2013; Sainsbury et al., 2009). The respective clones were transformed into chemically competent DH5α *Escherichia coli* using appropriate antibiotics and screened for positives through colony PCR. Entry clones were bidirectionally sequenced according to the standard methods of the Central Analytical Facility, Stellenbosch University, South Africa using M13 sequencing primers and a walking primer (5'-CTCTTCTATTGTTGGTATGTATTTTC-3').

4.2.3 Sequence analysis of isolated paralogs

Stellenbosch University https://scholar.sun.ac.za

Sequence analyses were performed using CLC Main Workbench 7 (CLC Bio-Qiagen, Denmark). The VviTPS compendium (Chapter 3) was queried to identify the cluster of proteins most similar to the sequenced isolates. MUSCLE alignment and phylogenetic tree construction were performed as described earlier. Data was extracted from the VviTPS compendium as described in Chapter 3. The *I'* score (Li et al., 2001) was calculated from BLASTn (Camacho et al., 2009) scores of isolates queried against all diploid complete genes (mRNA sequences).

4.2.4 *Agrobacterium*-mediated transient expression

Expression clones for isolates with predicted full-length open reading frames (fl-ORF) were transformed via electroporation into the *Agrobacterium tumefaciens* GV3101::pMP90 strain as described by Bach et al. (2014) and plated onto selective media. *Agrobacterium*-mediated transient expression in *Nicotiana benthamiana* was performed, as described by Smit et al. (2019), with three leaves from a single plant pooled after 4 days for the subsequent analysis. A mock infiltration using MMA buffer and non-infiltrated wild type leaves served as controls. Headspace solid phase micro-extraction (HS-SPME) and gas chromatography-mass spectrometry (GC-MS) was performed as described by Williams and Buica (2019), with the following modifications: adsorption of analytes was done at 30°C for 20 min and an initial oven temperature of 100°C. Sesquiterpenes were identified using the authentic standards (*E*)-β-farnesene and (*E*)-β-caryophyllene from Sigma-Aldrich, United Sates, and/or the Wiley 275 mass spectral library. For extracted ion chromatograms (EIC) we used the cumulative mass (m/z) response for 161, 189 and 204.

4.3 Results

4.3.1 Isolated VviTPS04 and -10-like paralogs

The forward strand of VviTPS04-like genes had a start codon 27 bp upstream of the start codon for *VviTPS10*-like loci with primers designed to target this difference, as shown in Figure 4.1a. It should be noted that the 27 bp differences results in two start codons being in-frame. The downstream region from the predicted VviTPS10 start codon is also highly conserved between all gene models from CS and CH, shown in Figure 4.1b. Primers were therefore validated through PCR amplification using gDNA from nine genotypes, as shown in Figure 4.1c-d. The expected band of ±2 260 bp was present in all genotypes with certain genotypes showing a second band at ±1 700 bp for primer set B (Figure 4.1d). A second band was notably absent for CH using either primer combination. For primer set A we saw a 2 000 bp amplicon in PN with this band faintly visible in 6 other genotypes (Figure 4.1c). Primer set A resulted in a unique second amplicon for CS. These results indicate that despite our efforts to design locus specific primers, multiple amplicons could not be avoided.

Isolation of the genes from grapevine flower cDNA resulted in the cloning for 6/9 genotypes using primer set A and 7/9 genotypes using primer set B. No isolates were amplified from WR. CS resulted in two isolates for both primer sets. cDNA amplicon sizes were, however, different for the genotypes (Figure 4.2a). Sequencing confirmed that two different forms were isolated and that the

Stellenbosch University https://scholar.sun.ac.za

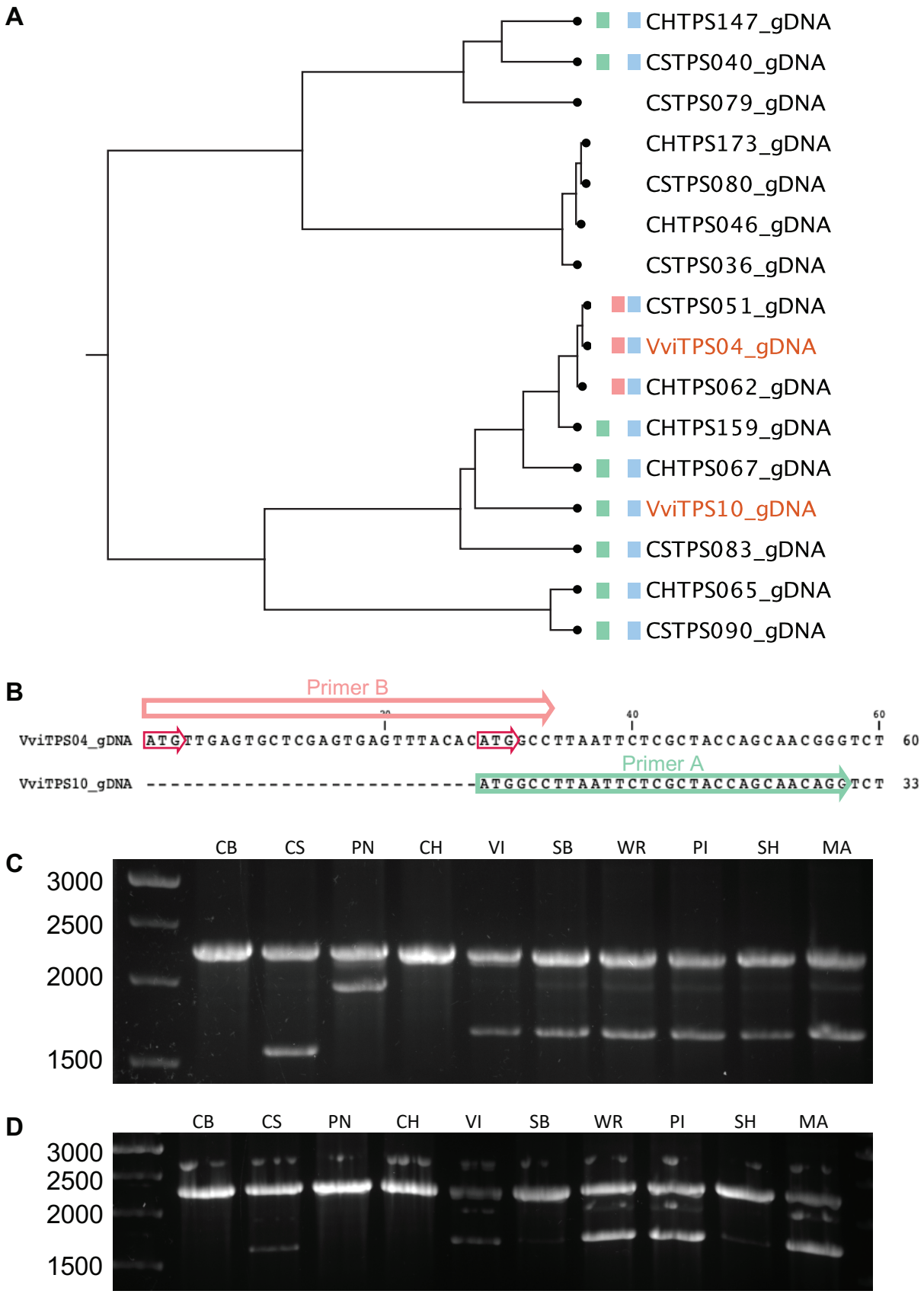

Figure 4.1 – **(A)** gDNA phylogeny of putative loci from CH and CS showing primer binding site conservation for the forward primers, show in **(B)**, and a single conserved reverse primer (blue square). gDNA PCR of 9 *V. vinifera* genotypes using the conserved reverse primer in combination with primer A, shown in **(C)**, or primer B, shown in **(D)**. The following abbreviations were used for the respective cultivars: Cabernet Sauvignon (CS), Chardonnay (CH), Chenin Blanc (CB), Muscat of Alexandria (MA), Pinot noir (PN), Sauvignon Blanc (SB), Shiraz (SH) and Viognier (VI).

Stellenbosch University https://scholar.sun.ac.za

27 bp differences was sufficient for the isolation of genotypic variants, evident in the pairwise sequence alignments in Appendix F. Of the 15 isolates, only six had fl-ORFs, while the remaining 9 variants were rendered non-functional due to premature stop codons, frame shifts and/or intron retention (Figure 4.2b). The multiple sequence alignment (MSA) for Figure 4.3b is available in Appendix F.

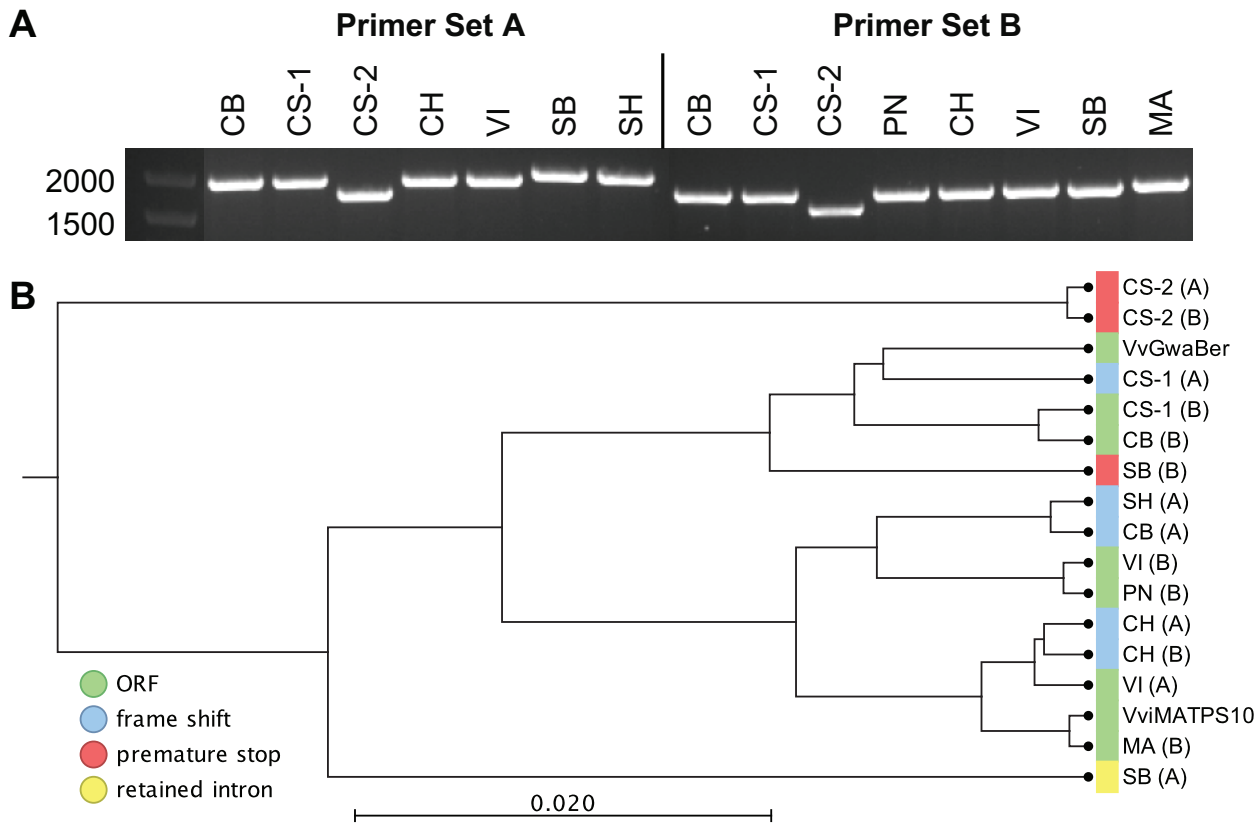

Figure 4.2. cDNA isolates from different genotypes using the two different primer combinations is shown in **(A).** Phylogenetic positions relative to VvGWaBer and VviMATPS10 nucleotide sequences is shown in **(B)**. The letter code in brackets refers to the primer set used for isolation with the type of disruption when non-functional indicated.

4.3.2 Transient expression of fl-ORFs of VviTPS10 in tobacco

Transient expression in *N. benthamiana* showed that all fl-ORF VviTPS10s produced volatile sesquiterpenes. Isolate VI-A was the only functional isolate, other than MA-B (VviTPS10), that produces (*E*)-β-farnesene as a single product. The VI genotype was the only cultivar that resulted in a functional isolate using both primers, with VI-B producing β-caryophyllene and an unidentified sesquiterpene (Figure 4.3). The unidentified sesquiterpene had a retention time very near to that of (*E*)-β-farnesene, with its *m/z* pattern confirming that it is not (*E*)-β-farnesene. No definitive match in either the Wiley 275 or NIST libraries could be found due the signal-to-noise ratio of this peak being below the limit of detection (LOD). All other genotypes produced β-caryophyllene and the unidentified sesquiterpene.

Stellenbosch University https://scholar.sun.ac.za

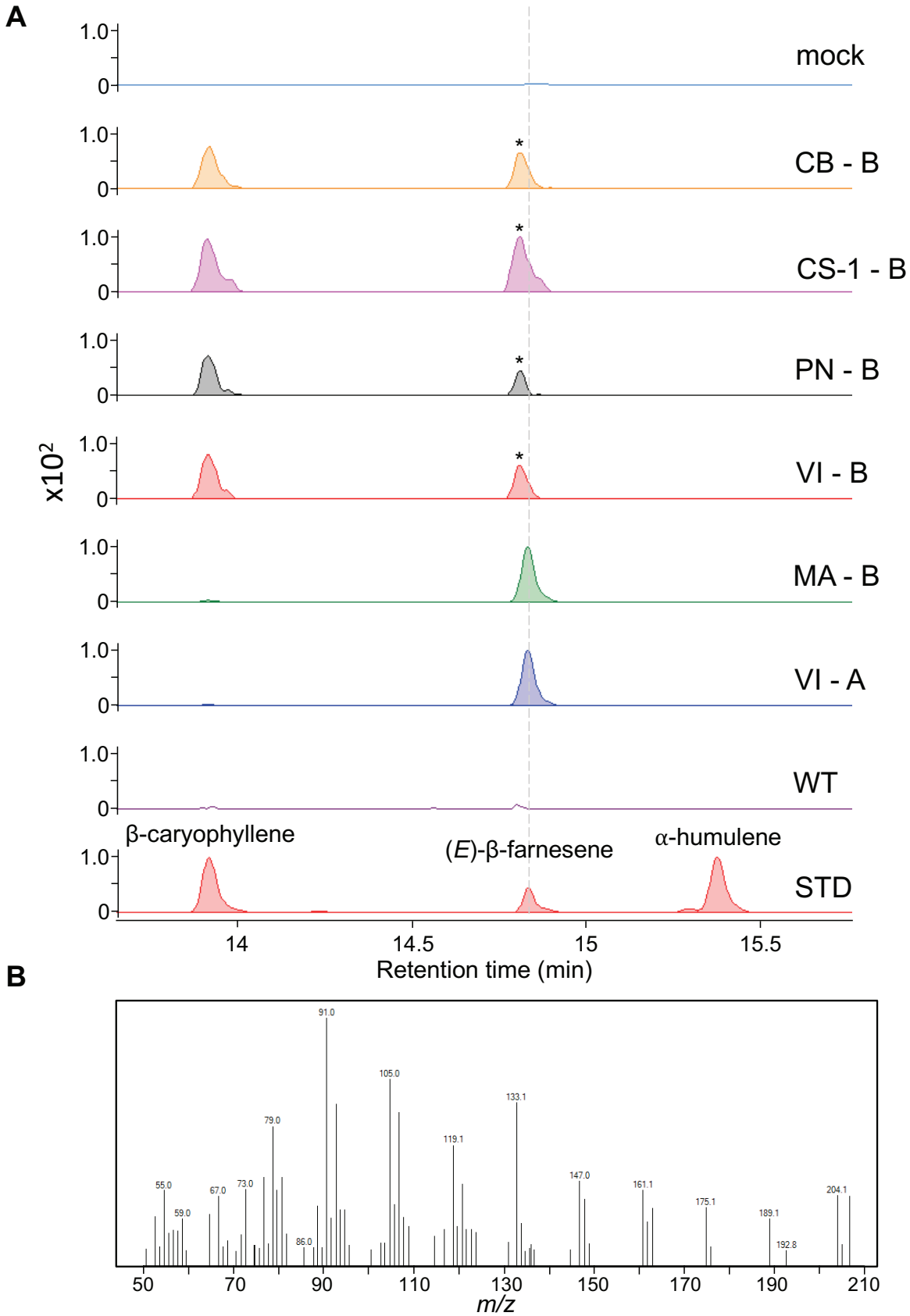

Figure 4.3. (A) Cumulative EIC using the masses 161, 189 and 204 for *N. benthamiana* transiently expressing genes of interest. Cultivar-specific genes are shown with peaks identified using authentic standards (STD). Wild-type (WT) and mock infiltrations served as controls. The *m/z* spectra for the peak indicated by the asterisk is shown in **(B)** with the dashed grey line showing that its retention time is different to that of (E)-β-farnesene. The following cultivar abbreviations were used: Chenin Blanc (CB), Cabernet Sauvignon (CS), Muscat of Alexandria (MA), Pinot noir (PN), and Viognier (VI).

Stellenbosch University https://scholar.sun.ac.za

4.3.3 Sequence-function relationships of isolated paralogs

Querying the VviTPS compendium (Chapter 3) for VviTPS10 paralogs revealed that they are part of a cluster consisting if 21 putative proteins that connect to four functionally characterized proteins (Figure 4.4a). Amino acid sequence similarity of the active sites to that of characterised enzymes showed that proteins in this cluster are predicted to use NPP as initial substrate with all, except VviMATPS10, predicted to have an initial 1,6-cyclisation, as described in Chapter 3. Subsequent analysis of the associated sequences (Figure 4.4a), showed that these proteins are encoded by a set of genotype-specific duplicated genes. These duplications form highly connected networks, illustrating the extent of paralogy, with the most extensive network connectivity seen in CS. Complete gene models associated with the gene duplicates, CSTPS083 and CSTPS090, were the closest paralogs of the functional isolates MA-B, CS-1-B and CB-B (Figure 4.4b).

Genomic regions where the duplication clusters (Figure 4.4a) is located were compared, with the gene order and structure (i.e. the VviTPS landscape) showing extensive differences within and between genotypes (Figure 4.5). CSTPS083 and CSTPS090 are 216 kb apart, separated by three complete genes and four partial genes (Figure 4.5a). CSTPS086, positioned in between CSTPS083 and CSTPS090, forms part of the gene duplication network in Figure 4.5a, along with two other genes that are positioned more than 314 kb upstream of CSTPS083. The rest of the genes in the duplication network are found on the second contig (Figure 4.6b). CRTPS019 and CRTPS199 were identified as potential gene duplications in CR, with the former not predicted to encode for a functional gene. The functional isolates PN-B and VI-B had the highest I' score with the CRTPS019 gene model (97.31 and 97.08, respectively) while VI-A was linked to CRTPS199 with a score of 98.99 (Figure 4.5b). The I' scores (Li et al., 2001) quantifies the extent of paralogy with phylogenetically similar genes (Figure 4.3) linking to distinct proteins based on this score (Figure 4.5). This allowed for inference of the protein and gene structures (Chapter 3) most similar to the isolates.

Stellenbosch University https://scholar.sun.ac.za

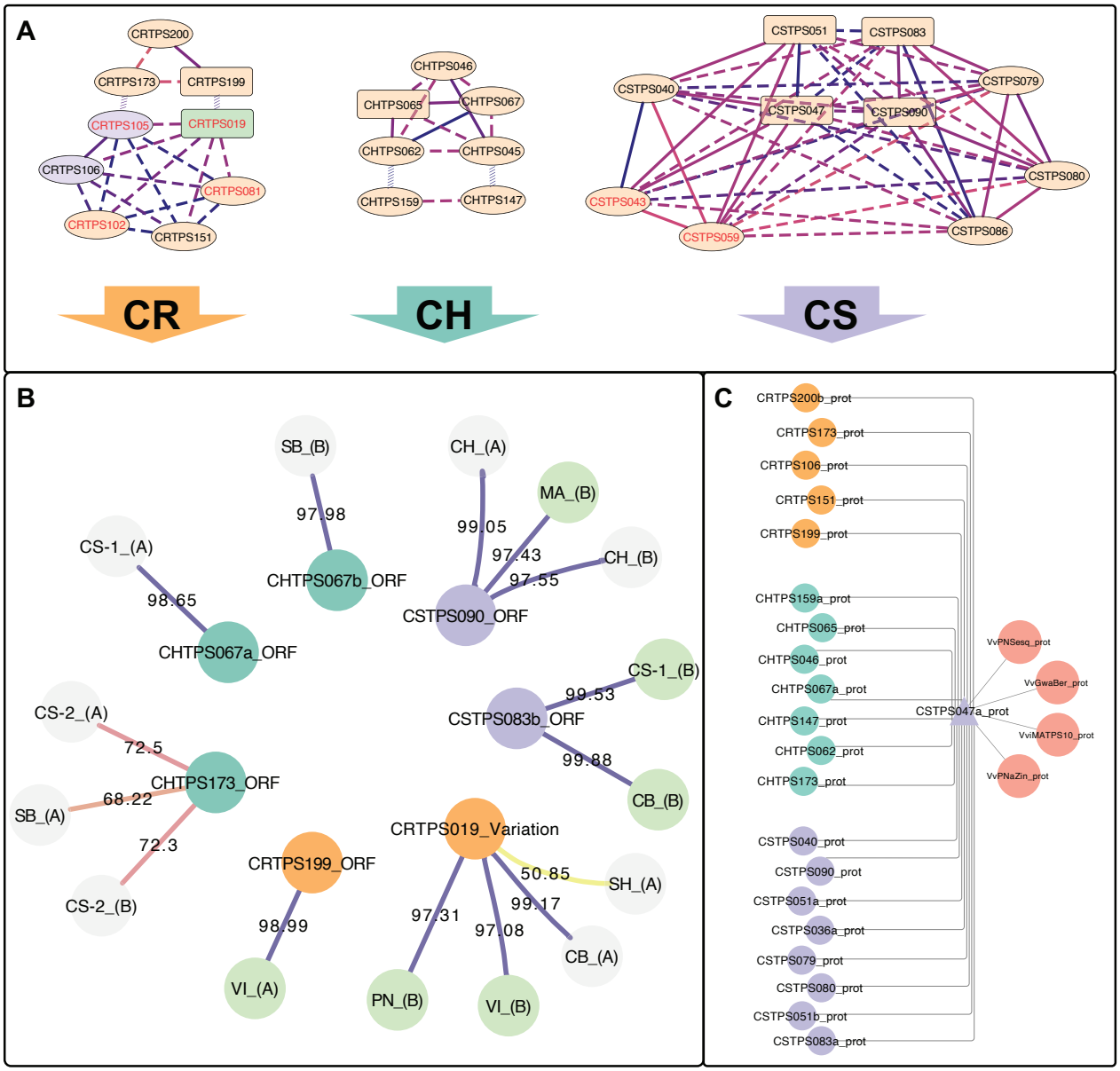

Figure 4.4. VviTPS compendium data for targeted paralogs. Genotype specific duplication networks is shown in **(A).** The best scoring complete gene match of the isolated genes are shown in **(B)** with edge values representing the I' score. **(C)** shows the protein cluster associated with the isolates. The following cultivar abbreviations were used: Chenin Blanc (CB), Chardonnay (CH), Carménère (CR), Cabernet Sauvignon (CS), Muscat of Alexandria (MA), Pinot noir (PN) and Viognier (VI).

Stellenbosch University https://scholar.sun.ac.za

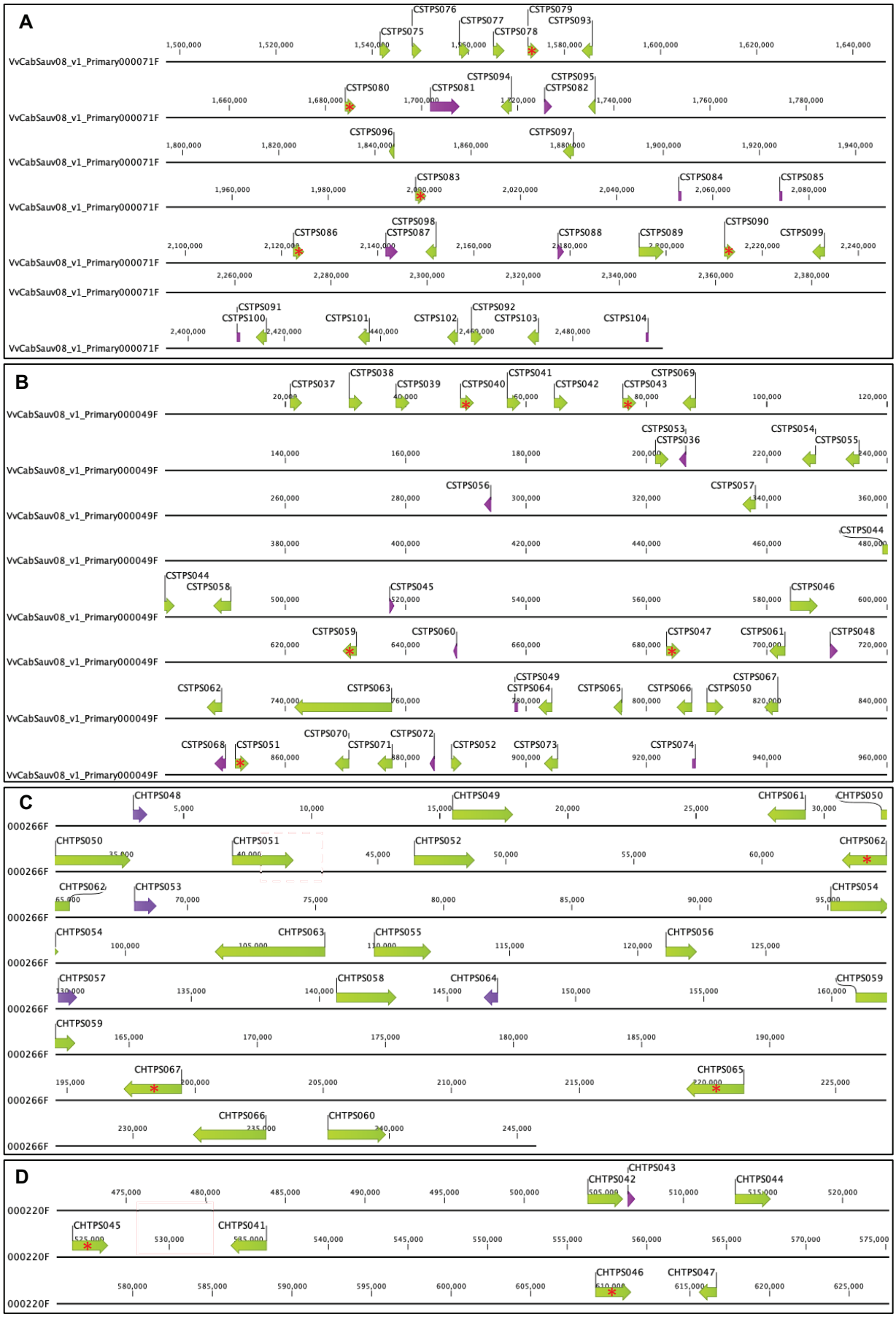

Figure 4.5 Cabernet Sauvignon (CS), shown in **(A)** and **(B)**, and Chardonnay (CH), shown in **(C)** and **(D)**, contigs associated with chr. 18 gene clusters. The red asterisks indicate that these genes form part of the gene duplication clusters in Figure 4.4a. Complete genes are shown as grown arrows with purple arrows showing partial genes.

Stellenbosch University https://scholar.sun.ac.za

The two CR gene models were, furthermore, located on a primary contig and haplotig, respectively, that, unlike the CH and CS paralogs, map to two different chromosomes (chr. 13 and 18). Although five of the isolates were connected to CH gene models, none of these isolates were predicted functional (Figure 4.4b). Closer inspection of the active site amino acids of derived protein sequences revealed that VI-A and MA-B, (*E*)-β-farnesene synthases, had near identical active sites with the only difference to MA-B being an amino acid deletion and a single nonsynonymous mutation, shown in Figure 4.6. Furthermore, pairwise comparisons of the full-length proteins for VI-A, VI-B and MA-A revealed that these proteins are more similar between cultivars than within a cultivar (Appendix F). The active sites of CB-B and CS-1-B were very similar to that of VvGwaBer, with only two amino acid differences (Figure 4.6), however, *in planta* expression (Figure 4.3) did not result in the synthesis of any compounds similar to those observed *in vitro* for VvGwaBer (Martin et al., 2010).

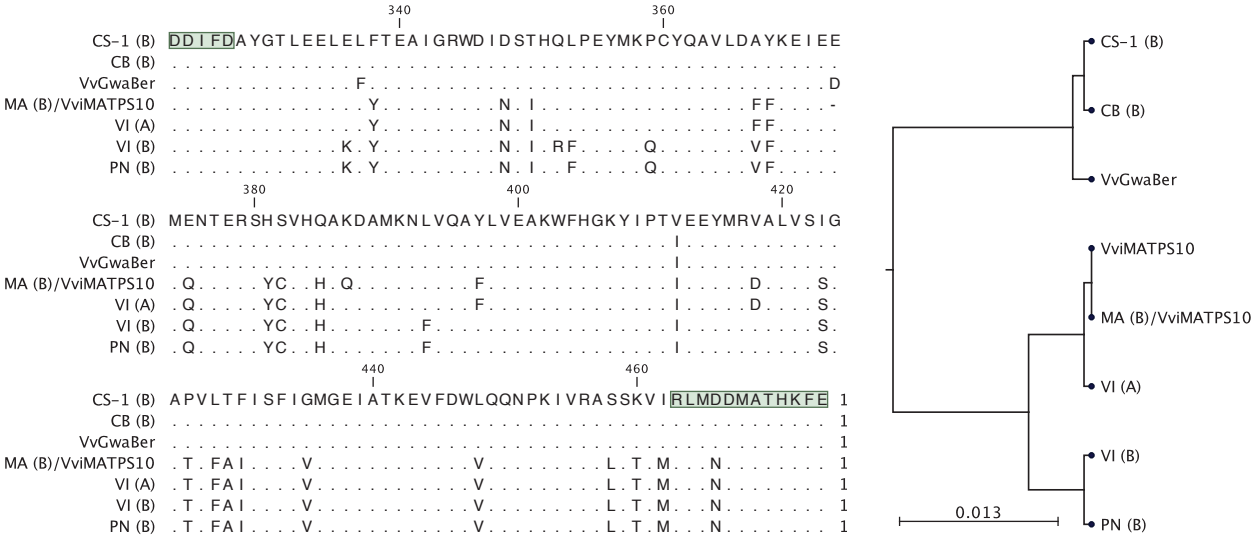

Figure 4.6 Active site alignment of the six functional isolates to VviMATPS10 and VvGwaBer. The position of the DDxxD and NSE/DTE active site motifs are highlighted in green. The following cultivar abbreviations were used: Chenin Blanc (CB), Cabernet Sauvignon (CS), Muscat of Alexandria (MA), Pinot noir (PN) and Viognier (VI).

4.4 Discussion

4.4.1 Genotype specific structural variation for isolated *VviTPS* paralogs

Genotypic variations in *VviTPS* genes were suggested to have a significant impact on the genetic potential of a given cultivar (Chapter 2). By targeting two paralogs in nine genotypes the extent of structural differences and sequence variations became evident (Figures 4.2). Previous efforts focussed only on enzyme function with only two studies reporting on the isolation of new and/or cultivar specific VviTPS genes (Drew et al., 2015; Dueholm et al., 2019). This was also shown for VviTPS24 where two point mutations were linked to an altered function in Shiraz (Drew et al., 2015). The use of the reference genome in an “applies-to-most” manner is therefore limiting when trying to explore both sequence and structural variations for *VviTPS* genes. The VviTPS compendium (Chapter 3) aimed to address some of these limitations with the results presented here providing functional validation for the VviTPS04 and -10 paralogs from multiple genotypes. Two paralogs were

Stellenbosch University https://scholar.sun.ac.za

isolated with extensive homology at the terminal ends; however, when compared to the diploid genomes, each presented a different genomic scenario for the isolated paralogs with the current lack of assembly limiting the extent of the conclusions that can be drawn. A key finding was that the presence of a transcript does not equate to a functional enzyme with only 6/15 isolates shown to be functional (Figures 4.2 and 4.3). Extensive SV between genotypes rendered the majority of isolates non-functional, however, two distinct active site forms were present. The isolation of a second genotype with an active site similar to that of VviMATPS10 provides evidence of a conserved active site cavity for (*E*)-β-farnesene biosynthesis (Figure 4.6). It is not yet clear which amino acid differences are key to this function. The paralog sequences therefore provide a biologically relevant sequence reference for possible future studies focussed on modelling the protein structure to identify key functional residues for site directed mutagenesis.

The sequence differences in the terminal end could be used to isolate two distinct loci, however, the VI and MA variants involved in (*E*)-β-farnesene biosynthesis were isolated with different primer pairs. From this it was concluded that the 27 bp difference used to distinguish paralogs (i.e. genome sequence differences) had no significant impact on the enzyme function. The genotype-specific expansion of paralogs for these two genes (Figures 4.4a and 4.5), furthermore, suggests that they are on a region that consists of multiple recent tandem duplications that went through genotype-specific neo-functionalization, likely exacerbated by domestication and continuous vegetative propagation (Keller, 2010; Zhou et al., 2017)

The CS region where these paralogs are found (Figure 4.5a-b) revealed that neither of the primary contigs had a gene order similar to that of PN40024. Assuming that phased sequencing of the two CS contigs in Figure 4.5a-ab accurately represent two separate genomic regions; one can conclude that these two VviTPS regions are ~950 kb segmental duplications. Alternatively, these two contigs could represent allelic regions on chr.18, despite the structural differences. The misidentification of allelic regions is a limitation of FALCON-UNZIP (the haplotype aware sequence assembler of PacBio sequences) where heterozygous alleles are assembled into primary contigs instead of haplotigs. New algorithms (for example FALCON-PHASE) have however been developed to address this (Kronenberg et al., 2018). Whether these are heterozygous alleles or segmental duplications does not change the fact that there are distinctly different genomic regions associated with the paralog clusters. The lack of assembled genomes is however temporary and will likely be addressed in the near future.

4.4.2 The impact of domestication on *VviTPS* expansion

It is theorised that the diversity in natural terpenes provide a fitness advantage with different models proposed for the origin of this diversity, reviewed in Pichersky and Raguso (2018). These models, however, do not account for the influence of domestication within a species, instead they focus on species-wide evolution. The majority of commercial grapevine cultivars planted currently have a

Stellenbosch University https://scholar.sun.ac.za

shared ancestry and are maintained primarily through vegetative propagation to preserve the desired phenotypic traits (Myles et al., 2011).

Grapevine genomic structure has been shown to be heavily impacted by somatic variation (Carbonell-Bejerano et al., 2017; Keller, 2010). Of particular interest is the occurrence of chromosome shattering (chromothripsis), where large parts of chromosomes are deleted, resulting in the hemizygous loss of large genomic regions, or incorrect joining of such regions, resulting in clustered rearrangements. This phenomenon was linked to berry colour development were the hemizygous loss of a genomic region, coupled with specific mutations in regulatory genes of anthocyanin biosynthetic genes, resulted in white grapes (Carbonell-Bejerano et al., 2017). Various studies have quantified the extent of SV and/or SNVs between grapevine genotypes (Di Genova et al., 2014; Giannuzzi et al., 2011; Xu et al., 2016) with a comparison between wine and table grapes showing up to 8% of the genome being affected (Cardone et al., 2016). The *VviTPS* family has likely been subjected to multiple genome level events which include segmental duplications (Giannuzzi et al., 2011) and transposon-mediated duplications, evident by presence of transposon-like remnants around *VviTPS* genes (Martin et al., 2010).

The *VviTPS* genomic landscape of chr. 18 (Figure 4.5) suggests that large-scale genome rearrangements (Carbonell-Bejerano et al., 2017), compounded by introgression (between species) and admixture (within species) (Zhou et al., 2017, 2018, 2019), resulted in heterozygous clustered duplications of *VviTPS* genes. The paralog/duplication networks in Figure 4.4a illustrates how these large-scale genomic events resulted in numerous genes of similar function being localised in a clustered manner. When one considers these clusters in the context of adaptive introgression, where large genomic regions are introduced from a different species, it is highly likely that geographic expansion resulted in the incorporation of alleles adapted to specific environments (Purugganan, 2019; Zhou et al., 2019). These alleles have been subjected to extensive mutational changes due to vegetative propagation, allowing for genotype-specific functionality (Figure 4.2). It is therefore highly likely that a combination of duplications, chromothripsis, extensive introgression and SNV events resulted in genotype specific *VviTPS* clusters, with the primary assemblies of CH and CS genomes, shown in Figure 4.5, supporting this hypothesis. In this study two paralogs were explored and, although limited, provide valuable insight into how the genomic landscape can shape a cultivar's potential to produce terpenes. Zhou et al. (2017) suggested that clonal propagation results in the accumulation of deleterious mutations on the recessive allele which could explain the SNV observed that rendered nine of the 15 *VviTPS* isolates non-functional. Clonal differences within a single cultivar therefore adds even further complexity to mutational events that impact gene function (Olmo, 1980; Roach et al., 2018).

Once assembled, the diploid genomes should provide new scaffolds for the mapping and mining of SV and SNVs (Di Genova et al., 2014; Laucou et al., 2018; Liang et al., 2019; Mercenaro et al., 2017; Myles et al., 2010; Tabidze et al., 2017; Zhou et al., 2017) to study allelic differences

Stellenbosch University https://scholar.sun.ac.za

that impact VviTPS functionality. Whole-genome sequencing (WGS) data in particular will provide valuable insight into the SNP events that shaped genotypic variations (Liang et al., 2019; Myles et al., 2010). The subtle impact of SNPs, which can render a gene non-functional or alter its catalytic mechanisms will therefore always necessitate functional characterisation. An alternative, more targeted approach than using a genome to identify targets, would be the use of PacBio SMRT sequencing of full-length cDNA transcripts (IsoSeq), as it allows for genome-independent sequencing of transcripts, addressing both structural and single base pair variation (Minio et al., 2019b).

4.4.3 VviTPS04 and -10 functions: an ecophysiological perspective

Continuous vegetative propagation ensures that desired traits are maintained, but simultaneously provides an unchanging host for pathogens and pests. In order to maintain fruit quality and production levels, European growers apply between 12-30 fungicide and 1-10 pesticide treatments per season, depending on the region where the vines are cultivated (Pertot et al., 2017). VviTPS04 and -10 paralogs are of particular ecophysiological interest due their role in the biosynthesis (*E*)-β-farnesene and (*E*)-β-caryophyllene, both known kairomones for the grapevine berry moth *Lobesia botrana* (Salvagnin et al., 2018; Tasin et al., 2006, 2011). This moth causes widespread production losses for grapevine planted in the Mediterranean region with recent reports of the polyphagous relative *L. vanilana* causing damage to grapes in a geographically localized region in the Western Cape of South Africa (Addison, 2017). Due to the commercial losses caused by the grape berry moth alternative methods to disrupt the kairomone ques through genetic engineering of grapevine was explored by overexpression of Aaβ-FS, an (*E*)-β-farnesene ortholog from *Artemisia annua* (Picaud et al., 2005; Salvagnin et al., 2018). The results from this study was promising as it decreased moth attraction to grapevine. The isolation and functional characterisation of a grapevine (*E*)-β-farnesene synthase (Smit et al., 2019) therefore presents a new target for silencing and/or overexpression studies aimed at disrupting *L. botrana* kairomones. (*E*)-β-caryophyllene is, however, a second major grapevine sesquiterpene that, along with (*E*)-β-farnesene and the homoterpene (*E*)-4,8-dimethyl-1,3,7-nonatriene (DMNT), form the core volatiles that attract the grapevine berry moth (Salvagnin et al., 2018; Tasin et al., 2006, 2011). Eight genes, including the one from this study, have so far been linked to (*E*)-β-caryophyllene biosynthesis in grapevine (Martin et al., 2010; Smit et al., 2019), which will make it difficult to create a null background for this sesquiterpene.

4.4 Conclusion

Genome-wide and gene-specific variation resulted in the expansion of the *VviTPS* family resulting in tremendous genetic diversity. The paralogs characterised in this chapter provided valuable insight into how SV and SNP impact gene function within and between genotypes. Whole genome resequencing data coupled with the increasing number of diploid *Vitis* spp. gene sequences will undoubtedly provide valuable new genetic targets to study *VviTPS* gene variations. However, the

Stellenbosch University https://scholar.sun.ac.za

complex biomechanical aspects of a TPS will require functional characterisation of such targets due to the subtle influence a single point mutation can have on the enzyme's function. The availability of diploid genomes allowed for the identification of paralogs within a heterozygous and highly duplicated region with the results presented here providing validation of what was predicted in the VviTPS compendium (Chapter 3). By utilizing these resources to consider both the extent of paralogy and heterozygosity it should now be possible to more accurately target *VviTPS* genes to establish structure-function relationships.

4.4 References

Addison, P. (2017). The ecology of Lobesia vanillana in wine grape vineyards, with emphasis on management options.

Aradhya, M. K., Dangl, G. S., Prins, B. H., Boursiquot, J.-M., Walker, M. A., Meredith, C. P., et al. (2003). Genetic structure and differentiation in cultivated grape, Vitis vinifera L. *Genet. Res.* 81, 179–92. doi:10.1017/S0016672303006177.

Bach, S. S., Bassard, J.-É., Andersen-Ranberg, J., Møldrup, M. E., Simonsen, H. T., and Hamberger, B. (2014). "High-throughput testing of terpenoid biosynthesis candidate genes using transient expression in Nicotiana benthamiana," in *Plant Isoprenoids: Methods and Protocols*, ed. M. Rodríguez-Concepción (New York: Springer Science), 245–255. doi:10.1007/978-1-4939-0606-2_18.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421. doi:10.1186/1471-2105-10-421.

Carbonell-Bejerano, P., Royo, C., Torres-Pérez, R., Grimplet, J., Fernandez, L., Franco-Zorrilla, J. M., et al. (2017). Catastrophic Unbalanced Genome Rearrangements Cause Somatic Loss of Berry Color in Grapevine. *Plant Physiol.* 175, 786–801. doi:10.1104/pp.17.00715.

Cardone, M. F., D'Addabbo, P., Alkan, C., Bergamini, C., Catacchio, C. R., Anaclerio, F., et al. (2016). Inter-varietal structural variation in grapevine genomes. *Plant J.* 88, 648–661. doi:10.1111/tpj.13274.

Carretero-Paulet, L., Librado, P., Chang, T. H., Ibarra-Laclette, E., Herrera-Estrella, L., Rozas, J., et al. (2015). High gene family turnover rates and gene space adaptation in the compact genome of the carnivorous plant utricularia gibba. *Mol. Biol. Evol.* 32, 1284–1295. doi:10.1093/molbev/msv020.

Chen, H., Köllner, T. G., Li, G., Wei, G., Chen, X., Zeng, D., et al. (2019). Combinatorial Evolution of a Terpene Synthase Gene Cluster Explains Terpene Variations in Oryza. *Plant Physiol.*, pp.00948.2019. doi:10.1104/pp.19.00948.

Chin, C., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* 13, 1050–1054. doi:10.1038/nmeth.4035.

Del Pozo, J. C., and Ramirez-Parra, E. (2015). Whole genome duplications in plants: An overview from Arabidopsis. *J. Exp. Bot.* 66, 6991–7003. doi:10.1093/jxb/erv432.

Dèrozier, S., Samson, F., Tamby, J.-P., Guichard, C., Brunaud, V., Grevet, P., et al. (2011). Exploration of plant genomes in the FLAGdb++ environment. *Plant Methods* 7, 8. doi:10.1186/1746-4811-7-8.

Di Genova, A., Almeida, A. M., Muñoz-Espinoza, C., Vizoso, P., Travisany, D., Moraga, C., et al. (2014). Whole genome comparison between table and wine grapes reveals a comprehensive catalog of structural variants. *BMC Plant Biol.* 14. doi:10.1186/1471-2229-14-7.

Drew, D. P., Andersen, T. B., Sweetman, C., Møller, B. L., Ford, C., and Simonsen, H. T. (2015). Two key polymorphisms in a newly discovered allele of the Vitis vinifera TPS24 gene are responsible for the production of the rotundone precursor α-guaiene. *J. Exp. Bot.* 67, 799–808. doi:10.1093/jxb/erv491.

Dueholm, B., Drew, D. P., Sweetman, C., and Simonsen, H. T. (2019). In planta and in silico characterization of five sesquiterpene synthases from Vitis vinifera (cv. Shiraz) berries. *Planta* 249, 59–70. doi:10.1007/s00425-018-2986-7.

Edgar, R. C. (2004a). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 1–19. doi:10.1186/1471-2105-5-113.

Edgar, R. C. (2004b). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acid Res.* 32, 1792–1797. doi:10.1093/nar/gkh340.

Falara, V., Akhtar, T. A., Nguyen, T. T. H., Spyropoulou, E. A., Bleeker, P. M., Schauvinhold, I., et al. (2011). The Tomato Terpene Synthase Gene Family. *Plant Physiol.* 157, 770–789. doi:10.1104/pp.111.179648.

Giannuzzi, G., D'Addabbo, P., Gasparro, M., Martinelli, M., Carelli, F. N., Antonacci, D., et al. (2011). Analysis of high-identity segmental duplications in the grapevine genome. *BMC Genomics* 12, 436. doi:10.1186/1471-2164-12-436.

Grimplet, J., Adam-Blondon, A.-F., Bert, P.-F., Bitz, O., Cantu, D., Davies, C., et al. (2014). The grapevine

Stellenbosch University https://scholar.sun.ac.za

gene nomenclature system. *BMC Genomics* 15, 1077. doi:10.1186/1471-2164-15-1077.
Jukes, T. H., and Cantor, C. R. (1969). “Evolution of Protein Molecules,” in *Mammalian Protein Metabolism*, ed. H. N. Munro (Elsevier), 21–132. doi:10.1016/B978-1-4832-3211-9.50009-7.
Keller, M. (2010). “Botany and Anatomy,” in *The Science of Grapevines: Anatomy and Physiology* (Elsevier), 1–47. doi:10.1016/B978-0-12-374881-2.00001-5.
Kronenberg, Z. N., Rhie, A., Koren, S., Concepcion, G. T., Peluso, P., Munson, K. M., et al. (2018). Extended haplotype phasing of de novo genome assemblies with FALCON-Phase. *bioRxiv*, 1–27.
Laucou, V., Launay, A., Bacilieri, R., Lacombe, T., Adam-Blondon, A. F., Bérard, A., et al. (2018). Extended diversity analysis of cultivated grapevine Vitis vinifera with 10K genome-wide SNPs. *PLoS One* 13, 1–27. doi:10.1371/journal.pone.0192540.
Li, W. H., Gu, Z., Wang, H., and Nekrutenko, A. (2001). Evolutionary analyses of the human genome. *Nature* 409, 847–849. doi:10.1038/35057039.
Liang, Z., Duan, S., Sheng, J., Zhu, S., Ni, X., Shao, J., et al. (2019). Whole-genome resequencing of 472 Vitis accessions for grapevine diversity and demographic history. *Nat. Commun.*, 1–12. doi:10.1038/s41467-019-09135-8.
Lin, J., Massonnet, M., and Cantu, D. (2019). The genetic basis of grape and wine aroma. *Hortic. Res.* 6. doi:10.1038/s41438-019-0163-1.
Martin, D., Aubourg, S., Schouwey, M., Daviet, L., Schalk, M., Toub, O., et al. (2010). Functional Annotation, Genome Organization and Phylogeny of the Grapevine (Vitis vinifera) Terpene Synthase Gene Family Based on Genome Assembly, FLcDNA Cloning, and Enzyme Assays. *BMC Plant Biol.* 10, 226. doi:10.1186/1471-2229-10-226.
Mercenaro, L., Nieddu, G., Porceddu, A., Pezzotti, M., and Camiolo, S. (2017). Sequence polymorphisms and structural variations among four grapevine (Vitis vinifera L.) cultivars representing sardinian agriculture. *Front. Plant Sci.* 8, 1–13. doi:10.3389/fpls.2017.01279.
Minio, A., Massonnet, M., Figueroa-Balderas, R., Castro, A., and Cantu, D. (2019a). Diploid Genome Assembly of the Wine Grape Carménère. *G3: Genes|Genomes|Genetics* 9, g3.400030.2019. doi:10.1534/g3.119.400030.
Minio, A., Massonnet, M., Figueroa-Balderas, R., Vondras, A. M., Blanco-Ulate, B., and Cantu, D. (2019b). Iso-Seq Allows Genome-Independent Transcriptome Profiling of Grape Berry Development. *G3 Genes, Genomes, Genet.* 9, g3.201008.2018. doi:10.1534/g3.118.201008.
Myles, S., Boyko, A. R., Owens, C. L., Brown, P. J., Grassi, F., Aradhya, M. K., et al. (2011). Genetic structure and domestication history of the grape. *Proc. Natl. Acad. Sci.* 108, 3530–3535. doi:doi.org/10.1073/pnas.1009363108.
Myles, S., Chia, J. M., Hurwitz, B., Simon, C., Zhong, G. Y., Buckler, E., et al. (2010). Rapid genomic characterization of the genus Vitis. *PLoS One* 5. doi:10.1371/journal.pone.0008219.
Olmo, H. (1980). Selecting and breeding new grape varieties. *Calif. Agric.* 34, 23–24.
Pertot, I., Caffi, T., Rossi, V., Mugnai, L., Hoffmann, C., Grando, M. S., et al. (2017). A critical review of plant protection tools for reducing pesticide use on grapevine and new perspectives for the implementation of IPM in viticulture. *Crop Prot.* 97, 70–84. doi:10.1016/j.cropro.2016.11.025.
Peyret, H., and Lomonossoff, G. P. (2013). The pEAQ vector series: The easy and quick way to produce recombinant proteins in plants. *Plant Mol. Biol.* 83, 51–58. doi:10.1007/s11103-013-0036-1.
Picaud, S., Brodelius, M., and Brodelius, P. E. (2005). Expression, purification and characterization of recombinant (E)-β-farnesene synthase from Artemisia annua. *Phytochemistry* 66, 961–967. doi:10.1016/j.phytochem.2005.03.027.
Pichersky, E., and Raguso, R. A. (2018). Why do plants produce so many terpenoid compounds? *New Phytol.* 220, 692–702. doi:10.1111/nph.14178.
Purugganan, M. D. (2019). Evolutionary Insights into the Nature of Plant Domestication. *Curr. Biol.* 29, R705–R714. doi:10.1016/j.cub.2019.05.053.
Reid, K. E., Olsson, N., Schlosser, J., Peng, F., and Lund, S. T. (2006). An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. *BMC Plant Biol.* 6, 27. doi:10.1186/1471-2229-6-27.
Roach, M. J., Johnson, D. L., Bohlmann, J., van Vuuren, H. J. J., Jones, S. J. M., Pretorius, I. S., et al. (2018). Population sequencing reveals clonal diversity and ancestral inbreeding in the grapevine cultivar Chardonnay. *PLoS Genet.*, 1–24. doi:https://doi.org/10.1371/journal.pgen.1007807.
Sainsbury, F., Thuenemann, E. C., and Lomonossoff, G. P. (2009). pEAQ: Versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* 7, 682–693. doi:10.1111/j.1467-7652.2009.00434.x.
Salmaso, M., Faes, G., Segala, C., Stefanini, M., Salakhutdinov, I., Zyprian, E., et al. (2004). Genome diversity and gene haplotypes in the grapevine (Vitis vinifera L.), as revealed by single nucleotide polymorphisms. *Mol. Breed.* 14, 385–395. doi:10.1007/s11032-004-0261-z.
Salvagnin, U., Malnoy, M., Thöming, G., Tasin, M., Carlin, S., Martens, S., et al. (2018). Adjusting the scent ratio: using genetically modified Vitis vinifera plants to manipulate European grapevine moth behaviour. *Plant Biotechnol. J.* 16, 264–271. doi:10.1111/pbi.12767.

Stellenbosch University https://scholar.sun.ac.za

Smit, S. J., Vivier, M. A., and Young, P. R. (2019). Linking Terpene Synthases to Sesquiterpene Metabolism in Grapevine Flowers. *Front. Plant Sci.* 10, 1–18. doi:10.3389/fpls.2019.00177.

Sokal, R. R. (1958). A statistical method for evaluating systematic relationships. *Univ. Kansas, Sci. Bull.* 38, 1409–1438.

Tabidze, V., Pipia, I., Gogniashvili, M., Kunelauri, N., Ujmajuridze, L., Pirtskhalava, M., et al. (2017). Whole genome comparative analysis of four Georgian grape cultivars. *Mol. Genet. Genomics* 292, 1377–1389. doi:10.1007/s00438-017-1353-x.

Tasin, M., Bäckman, A. C., Bengtsson, M., Ioriatti, C., and Witzgall, P. (2006). Essential host plant cues in the grapevine moth. *Naturwissenschaften* 93, 141–144. doi:10.1007/s00114-005-0077-7.

Tasin, M., Betta, E., Carlin, S., Gasperi, F., Mattivi, F., and Pertot, I. (2011). Volatiles that encode host-plant quality in the grapevine moth. *Phytochemistry* 72, 1999–2005. doi:10.1016/j.phytochem.2011.06.006.

Williams, C., and Buica, A. (2019). Comparison of an offline-SPE-GC-MS and online-HS-SPME-GC-MS method for the analysis of volatile terpenoids in wine. 24.

Xu, Y., Gao, Z., Tao, J., Jiang, W., Zhang, S., Wang, Q., et al. (2016). Genome-wide detection of SNP and SV variations to reveal early ripening-related genes in grape. *PLoS One* 11, 1–26. doi:10.1371/journal.pone.0147749.

Zhou, Y., Massonnet, M., Sanjak, J. S., Cantu, D., and Gaut, B. S. (2017). Evolutionary genomics of grape (Vitis vinifera ssp. vinifera) domestication. *Proc. Natl. Acad. Sci.* 114, 201709257. doi:10.1073/pnas.1709257114.

Zhou, Y., Muyle, A., and Gaut, B. S. (2019). “Evolutionary Genomics and the Domestication of Grapes,” in *The Grape Genome*, eds. D. Cantu and M. A. Walker (Cham: Springer International Publishing), 39–55. doi:10.1007/978-3-030-18601-2_3.

Zhou, Y. S., Minio, A., Massonnet, M., Solares, E. A., Lyu, Y., Beridze, T., et al. (2018). Structural variants, clonal propagation, and genome evolution in grapevine (Vitis vinifera). *bioRxiv*, 508119. doi:10.1101/508119.

Stellenbosch University https://scholar.sun.ac.za

Chapter 5

General Discussion and Conclusions

Stellenbosch University https://scholar.sun.ac.za

General Discussion and Conclusions

5.1 Research Approach and Overview

In order to investigate the structure-function relationships of grapevine *VviTPS* genes involved in flower terpene biosynthesis, a combination of computational biology, chemical profiling and molecular biology techniques were employed. The first aim was to evaluate the role of *VviTPS* genes in flower volatile emissions which resulted in: (i) detailed functional characterisation of *VviTPS* genes involved in flower terpene biosynthesis (Chapters 2 and 4) and; (ii) chemotyping (chemical profiling) of flower terpenes for commercially important grapevine cultivars (Chapter 2).

The second aim was to evaluate the extent of genotypic variations for the *VviTPS* family using available genomic resources. A variety of computational biology approaches were applied resulting in: (i) remapping of microarray probes to *VviTPS* genes to facilitate expression profiling using the grapevine gene atlas (Chapter 2); and (ii) identification and characterisation of genome-wide structural differences of the *VviTPS* family (Chapter 3). The latter contribution was possible due to the recent release of multiple new grapevine genome sequences using long-read PacBio technology. It was subsequently possible to identify heterozygous *VviTPS*-like gene regions (Chapter 3) and explore the extent of paralogy for targeted gene clusters (Chapter 4).

5.2 General discussion

5.2.1 Chemotypic differences in grapevine flowers can be linked to the function of specific *VviTPS* genes

Existing transcriptomic data for grapevine flowers is very limited with the grapevine gene atlas presenting the most comprehensive dataset, encompassing different stages in flower phenology as well as flower sub-organelles (Fasoli et al., 2012). The disjoint between the Nimblegen microarray and Martin et al. (2010) gene models was addressed through *in silico* remapping of the probes, allowing for a more accurate analysis of *VviTPS* expression in the gene atlas (Chapter 2). However, extensive probe ambiguities were identified for the 69 putatively functional *VviTPS* models; only 133 of 276 probes bound unambiguously with the remaining 143 probes having ambiguous affinity for up to six genes. It was, however, possible to identify the global *VviTPS* expression patterns within the gene atlas with a subset of the *VviTPS* genes (those with unique probes) analysed for their flower specific expression patterns. From this result it was expected to see both mono- and sesquiterpene volatiles in grapevine flowers. Screening of nine different flower chemotypes revealed a unique bouquet for each cultivar, with only sesquiterpenes being emitted (Chapter 2), according to our analysis that focused only on the unbound terpenes.

Subsequent in-depth chemotyping of three cultivars at different flower phenological stages resulted in the quantification of a unique sesquiterpene volatile for Shiraz ((+)-valencene), Sauvignon Blanc ((*E,E*)-α-farnesene) and Muscat of Alexandria ((E)-β-farnesene). The emission of a unique

Stellenbosch University https://scholar.sun.ac.za

major sesquiterpene for each cultivar was contrasting to existing flower chemotypes that reported (+)-valencene as the major flower sesquiterpene for thirteen of sixteen cultivars, with two cultivars emiting (+)-valencene as the second major volatile after β-caryophyllene (Buchbauer et al., 1994a, 1994b, 1995; Martin et al., 2009), while one cultivar was dominated by monoterpenes rather than sesquiterpenes (Matarese et al., 2014). The majority of these flower chemotypes are, however, for lesser known Austrian cultivars with only five of the profiled cultivars being non-Austrian varieties. The HS-SPME-GC-MS methodology applied in this study provided excellent chromatographic resolution for sesquiterpenes but likely at the cost of lower abundance volatile monoterpenes. It is also possible that monoterpenes are glycosidically bound which would require either, heat, derivatization or enzymatic hydrolysis to cleave these bonds. Future efforts to quantify the bound fraction of terpenes will add to the volatile sesquiterpene chemotypes observed in this study.

The disjoint between the gene atlas expression patterns and observed flower volatiles could to some extent be explained by the unique transcriptome of Corvina (a.k.a. Corniola), the cultivar used for the gene atlas, keeping in mind that the Nimblegen array that was used for the atlas was designed on the near-homozygous PN40024 genome (Fasoli et al., 2012). Transcriptomic analysis using RNAseq revealed that Corvina has extensive sequence and structural variations compared to PN40024, resulting in novel isoforms as well as unique (private) genes not detected on the reference genome (Venturini et al., 2013). Myles et al. (2011) furthermore showed, through SNP profiling, that Corvina is an offspring of Muscat of Alexandria, which makes it genetically dissimilar to Pinot Noir. The unique flower chemotypes and discordant gene atlas expression patterns highlight the limitation of using reference-genome based platforms to study genotypic differences for specialised gene families.

Characterisation of the *VviTPS10* gene in nine genotypes resulted in the isolation of a novel cultivar-specific variant (VviMATPS10) that is involved in (*E*)-β-farnesene biosynthesis (Chapters 2 and 4). It was furthermore found that the sequence variations are extensive between the genotypes with only 13/30 gene isolates being functional. Sequence variations were found to result in nonsynonymous mutations that disrupt the coding sequence (premature stop or frameshifts) for 57% of the gene isolates. The presence of a transcript therefore does not necessarily equate to a functional protein, a factor that must be considered for expressions studies. The microarray remapping presented in Chapter 2 and the genotype-specific *VviTPS* gene models in Chapter 3, provide new resources for transcriptomic analyses. Although transcriptomic data for grapevine flowers is very limited, the wealth in grape berry transcriptomes, present numerous datasets for *post hoc* analyses using the aforementioned resources. Such analyses will undoubtedly advance the currant understanding of VviTPS spatial-temporal expression patterns.

The active site of VviMATPS10 was compared to that of the Gewürztraminer VviTPS10 variant VvGwaBer, characterised as a multi-product sesquiterpene synthase with α-bergamotene as the major volatile (Martin et al., 2010). Extensive amino acid differences (37 of 50) were observed

Stellenbosch University https://scholar.sun.ac.za

between these two variants (Chapter 2). The exclusive biosynthesis of (*E*)-β-farnesene was also found in a Viognier variant of VviTPS10, with its active site being near identical to that of VviMATPS10 (Chapter 4). Furthermore, none of the cultivar variants isolated resulted in α-bergamotene when expressed *in planta,* instead producing β-caryophyllene. The active sites of all isolated VviTPS10 gene variants were dissimilar to that of VvGwaBer. It is thus possible that the Gewürztraminer variant has a unique function.

The cultivar variants isolated in this study provides biologically relevant sequence differences with characterised functional differences that can be applied for targeted mutagenesis to elucidate which amino acid residues are critical for the divergent enzyme functions. Köllner et al., (2009) suggested that a single amino acid residue in TPS10 of maize dictates the extent of (*E*)-β-farnesene to α-bergamotene ratios for different teosinte species. However, in grapevine three different enzyme functions have now been linked to VviTPS10, two of these generated in this study. It is therefore unlikely that a single amino acid will act as the catalytic switch. Directed mutagenesis of the *A. annua* humulene synthase has revealed extensive amino acid plasticity within the active site with numerous residues involved in divergent catalytic mechanisms (Li et al., 2013). The biologically relevant catalytic differences observed for VviTPS10 therefore requires further study. Directed mutagenesis should in future be coupled with protein modelling and ligand docking to ascertain how the active site interacts with FPP/NPP to produce (*E*)-β-farnesene (VviMATPS10 function) and identify which amino acid residues are required for cyclisation of NPP to produce α-bergamotene (VvGwaBer function).

5.2.2 The ecophysiological importance of grapevine terpene volatiles

(*E*)-β-farnesene is one of the few grapevine terpenes that have been shown to a have an important ecophysiological function. Its role as a kairomone, along with DMNT and β-caryophyllene, for the grapevine berry moth (*L. botrana*) has been well established (Anfora et al., 2009; Tasin et al., 2005, 2007, 2009). This moth causes extensive damage to grapevine and is currently controlled through extensive chemical applications which is not only costly but also damaging to the environment. Salvagnin et al., (2018) showed that the overexpression of the *A. annua* (*E*)-β-farnesene synthase in grapevine disrupted kairomone ratios, reducing the moth's attraction to grapevine. Identifying and characterising the genotypic and phenotypic differences for VviTPS10 (Chapters 2 and 4), therefore, provides valuable insight into the innate capacities of specific cultivars to emit (*E*)-β-farnesene.

The larvae of *L. botrana* have furthermore been identified as vectors for *B. cinerea* with the larvae-mediated transport of conidia resulting in increased bunch rot (Keller et al., 2003). Grapevine flower volatiles have been shown to restrict *B. cinerea* sporulation by forcing the conidia into a quiescent state until more favourable conditions arise for sporulation (Haile et al., 2016). Sesquiterpene synthase expression was shown to be upregulated in response to *B. cinerea* infection, suggesting a possible role in defence against this fungus (Mehari et al., 2017). This

Stellenbosch University https://scholar.sun.ac.za

possible trophic interaction between grapevine flowers, *L. botrana* and *B. cinerea* requires further investigation and could provide new insights for mitigating the damage caused by both moths and fungal pathogens.

5.2.3 Grapevine genomes: A step towards a pangenomic view of the *VviTPS* family

The PN40024 grapevine reference genome has been the genomic workhorse for just over a decade (Grimplet and Cramer, 2019; Jaillon et al., 2007). Various invaluable insights were facilitated by the availability of a reference genome. However, the limitations of a near-homozygous genome have become increasingly evident, especially when studying expanded gene families involved in secondary/specialised metabolism. Extensive duplications within the *VviTPS* family, coupled with a propensity for mutations due to vegetative propagation, limits the accuracy of short-read sequencing technology (genomic and transcriptomic) for analysing a highly duplicated gene family. The accuracy and completeness of the genome assembly, with regards to the *VviTPS* family, is therefore greatly limited by this with nearly a third of the gene models being mapped to chr. 00 (unassembled pseudo-molecule) (Canaguier et al., 2017). Transcriptomic approaches using paired-end RNAseq has mitigated some of genome limitations through *de novo* assembling of transcripts, however, the majority of studies still utilised V1 of the gene annotation.

The availability of new grapevine genomes provides extensive genomic resources that can address some of the PN40024 limitations. An initial evaluation of the gene annotation that accompanied the Cabernet Sauvignon diploid genome, revealed that *VviTPS* gene structures were often erroneous in areas were duplications were prevalent, akin to what was observed for the v0 and v1 annotations of PN40024 (Chin et al., 2016; Minio et al., 2017, 2019b). The Cabernet Sauvignon VviTPS family was therefore manually curated. Subsequent to this was the release of two additional diploid genomes, Chardonnay and Carménère (Minio et al., 2019a; Roach et al., 2018). The *VviTPS* family of these two genomes were also curated, allowing for a comparative analysis of the *VviTPS* “landscape” for three distinct *VviTPS* gene families. The complexity of the genotype-specific gene families was evident once the data was visualised as gene networks (Chapter 3). The diploid draft assemblies and network visualisation of the *VviTPS* families provided the first insights for genome wide structural differences for a complex and expanded specialised gene family. A particularly significant contribution was the identification of paralogous gene clusters for each of the genotypes. Finding homologous genes on the reference genome was a tedious task, however, the VviTPS compendium (Chapter 3) provides a more simple and interactive resource to find *VviTPS* orthologs.

The annotation of *VviTPS* genes was extended to include important structural features (motifs and domains) of plant TPSs. These features, as well as the amino acid composition of the enzyme active site are highly informative due to the number of sequenced and experimentally characterised plant TPSs currently available, as reviewed by Christianson, (2006, 2017). It was therefore possible to identify the gene subfamily and predict enzyme mechanisms based on sequence identity (Degenhardt et al., 2009; Wymore et al., 2011).

Stellenbosch University https://scholar.sun.ac.za

The active site in particular was used to predict the initial substrate that will be utilised as well as the first cyclisation mechanism. This approach has been previously applied for plant sesquiterpene synthases (Durairaj et al., 2019; Smit et al., 2019) with the VviTPS compendium extending this application to the TPS-b (monoterpene) and TPS-g (terpene alcohol) subfamilies. In its entirety the VviTPS compendium provides an extensive multi-genotype resource for functional analysis of *VviTPS* genes. Future studies will be able to use the curated gene models to target specific catalytic mechanisms, allowing for a more targeted approach when attempting to characterise *VviTPS* genes. Furthermore, the gene models provide extensive mapping targets for RNAseq data which will allow for more accurate characterisation of specific *VviTPS* transcriptional patterns. This resource is intended to be available online and will therefore be applicable to studies on other plant species, due to the conserved mechanisms of plant terpene synthases.

5.2.4 A model for *VviTPS* gene expansion

The results generated in the research chapters and the domestication history of grapevine (Chapter 1) can be contextualised with the model shown in Figure 5.1. Introgression shaped the initial grapevine *VviTPS* family through various large-scale genomic events that include duplications (tandem and segmental) (Giannuzzi et al., 2011; Martin et al., 2010) with chromothripsis (Carbonell-Bejerano et al., 2017), due to vegetative propagation resulting in large-scale chromosomal rearrangements where duplications are clustered and/or completely lost (Chapters 3 and 4). Structural variations were observed for the *VviTPS* family with the diploid genomes providing evidence for the impact of heterozygosity on this gene family (Chapter 3). The majority of the *VviTPS* family is localised to three chromosomes with TPS-a members (sesqui-TPS) largely found on chr. 18, TPS-b (mono-TPS) members in chr. 13 and the terpene alcohol synthases (TPS-g) on chr 10. Although these observations are currently limited by the unassembled nature of the different diploid genomes, the chromosome positions are largely concordant with the reference genome (Chapter 3). The duplication networks provided in Chapter 3, combined with the genomic structural differences discussed in Chapter 4 explain the paralogy observed within the *VviTPS* family. However, further analysis is required of complete chromosomal regions (synteny) in order to establish the extent and genotype-specific impact of structural differences. The model presented in Figure 5.1 is focussed on the *VviTPS* family but can be extended to other expanded gene families. For example, the grapevine cytochrome P450 (CYP450) family consists of 579 loci on the reference genome, with 279 complete genes (Ilc et al., 2018). CYP450 enzyme are known to modify terpenes, resulting in new bioactive metabolites (Ilc et al., 2016a, 2016b).

Domestication and more recent breeding efforts likely result in a “shuffling” of genes involved in specialised metabolism, depending on the extent of admixture and introgression (Vondras et al., 2019; Zhou et al., 2019). Sequencing of both the parents and the offspring will be required to establish how this affects the *VviTPS* family. The haplotype scenarios presented in figure 5.1 (i-vii) provides a model for the impact of structural variation on the *VviTPS* family. In addition to large structural variations (i.e. shuffling of genes), there is an increased propensity for mutations due to vegetative propagation (Zhou et al., 2017). The results in Chapters 2 and 4 provided insight into how these mutations (most often a loss of function) allow for *VviTPS* cultivar differences.

Stellenbosch University https://scholar.sun.ac.za

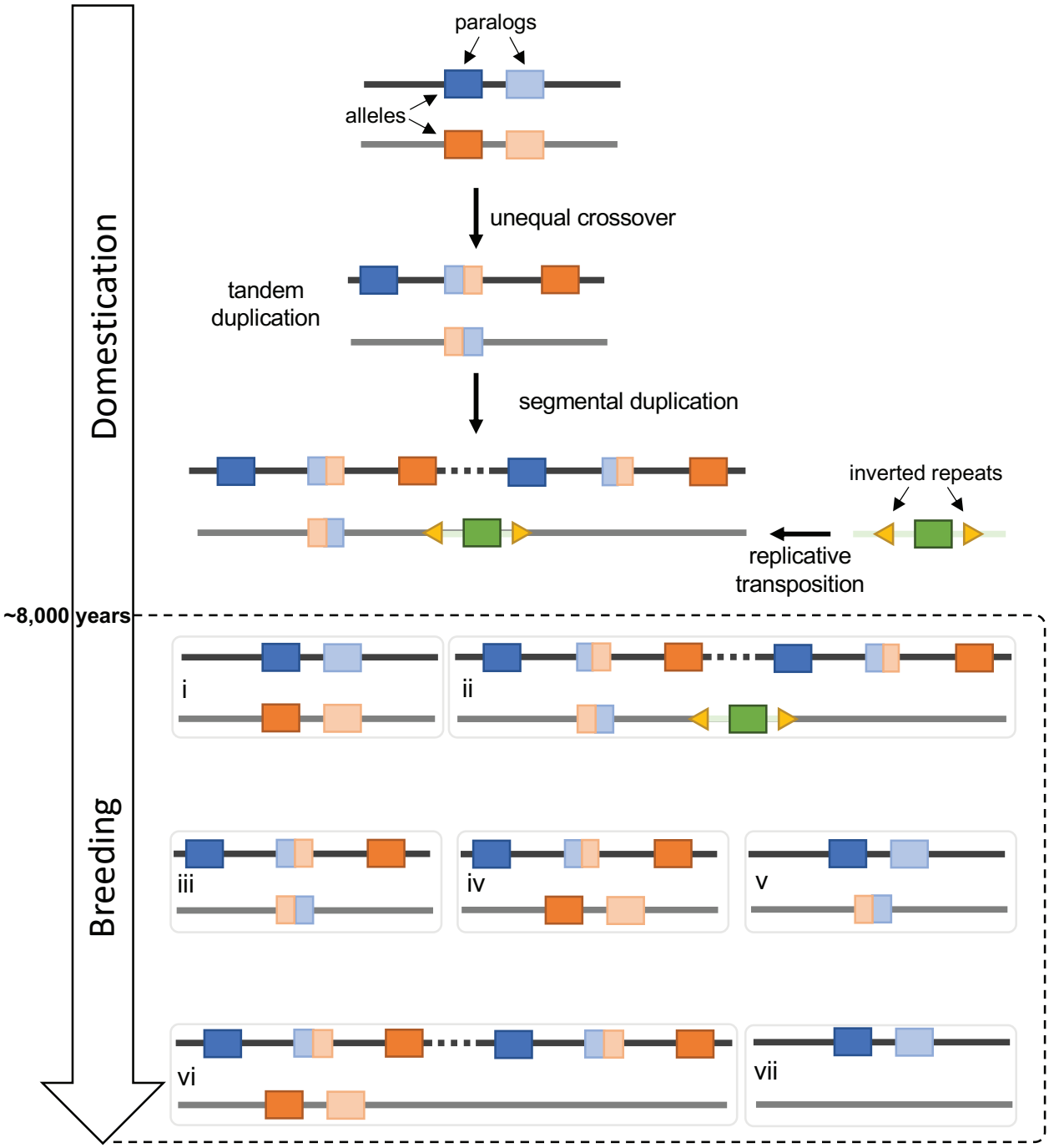

Figure 5.1 A model for how domestication events shaped the initial *VviTPS* gene family with current cultivation practices continuously generating new haplotype scenarios: (i) duplicated paralogs; (ii) segmental and transposon-mediated duplication (iii) unequal crossover; (iv-vi) combinatorial events due to admixture and (vii) hemizygous deletion.

5.3 Conclusions

The ecophysiological importance of terpenes in plant defence and stress mitigation is of particular relevance to grapevine. The expansion of this gene family, extensive genotypic differences and thousands of cultivars provide numerous genetic backgrounds that can be explored to identify traits for crop improvement. Annotation of the *VviTPS* family in three diploid genomes is therefore an important first step towards deciphering the complex “pedigree” of grapevine cultivars and how it affects terpene biosynthesis. The results generated in this study provides a roadmap for the

Stellenbosch University https://scholar.sun.ac.za

fundamental understanding and prediction of VviTPS functions, providing a platform for future efforts. The rapid generation of large-scale omic data necessitates an interactive and easy to understand view of complex gene families with a concerted effort required to make the data accessible. Future efforts to understand VviTPS sequence-function relationships will require the integration of genomic, transcriptomic and proteomic data as well as enzyme characterisation. In its entirety, this study advanced our understanding of the *VviTPS* family and provides important fundamental insights for future studies.

5.4 References

Anfora, G., Tasin, M., de Cristofaro, A., Ioriatti, C., and Lucchi, A. (2009). Synthetic grape volatiles attract mated Lobesia botrana females in laboratory and field bioassays. *J. Chem. Ecol.* 35, 1054–1062. doi:10.1007/s10886-009-9686-5.

Buchbauer, G., Jirovetz, L., Wasicky, M., Herlitschka, A., and Nikiforov, A. (1994a). Aroma von WeiBweinblüten: Korrelation sensorischer Daten mit Headspace-Inhaltsstoffen. *Zeitschrift für Leb. Und-forsch.* 199, 1–4. doi:http://sci-hub.tw/10.1007/BF01192941.

Buchbauer, G., Jirovetz, L., Wasicky, M., and Nikiforov, A. (1994b). Headspace Analysis of Vitis vinifera (Vitaceae) Flowers. *J. Essent. Oil Res.* 6, 311–314. doi:10.1080/10412905.1994.9698383.

Buchbauer, G., Jirovetzl, L., Wasieky, M., and Nikiforov, A. (1995). Aroma von Rotweinblüten: Korrelation sensorischer Daten mit Headspace-Inhaltsstoffen. *Zeitschrift für Leb. Und-forsch.* 200, 443–446.

Canaguier, A., Grimplet, J., Gaspero, G. Di, Scalabrin, S., Duchêne, E., and Choisne, N. (2017). A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3). *Genomics Data* 14, 56–62. doi:10.1016/J.GDATA.2017.09.002.

Carbonell-Bejerano, P., Royo, C., Torres-Pérez, R., Grimplet, J., Fernandez, L., Franco-Zorrilla, J. M., et al. (2017). Catastrophic unbalanced genome rearrangements cause somatic loss of berry color in grapevine. *Plant Physiol.* 175, 786–801. doi:10.1104/pp.17.00715.

Chin, C., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* 13, 1050–1054. doi:10.1038/nmeth.4035.

Christianson, D. W. (2006). Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev.* 106, 3412–3442. doi:10.1021/cr050286w.

Christianson, D. W. (2017). Structural and chemical biology of terpenoid cyclases. *Chem. Rev.* 117. doi:10.1021/acs.chemrev.7b00287.

Degenhardt, J., Köllner, T. G., and Gershenzon, J. (2009). Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* 70, 1621–1637. doi:10.1016/j.phytochem.2009.07.030.

Durairaj, J., Di Girolamo, A., Bouwmeester, H. J., de Ridder, D., Beekwilder, J., and van Dijk, A. D. (2019). An analysis of characterized plant sesquiterpene synthases. *Phytochemistry* 158, 157–165. doi:10.1016/j.phytochem.2018.10.020.

Fasoli, M., Dal Santo, S., Zenoni, S., Tornielli, G. B., Farina, L., Zamboni, A., et al. (2012). The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. *Plant Cell* 24, 3489–505. doi:10.1105/tpc.112.100230.

Giannuzzi, G., D'Addabbo, P., Gasparro, M., Martinelli, M., Carelli, F. N., Antonacci, D., et al. (2011). Analysis of high-identity segmental duplications in the grapevine genome. *BMC Genomics* 12, 436. doi:10.1186/1471-2164-12-436.

Grimplet, J., and Cramer, G. R. (2019). "The Grapevine Genome Annotation," in *The Grape Genome*, eds. D. Cantu and M. A. Walker (Cham: Springer International Publishing), 89–101. doi:10.1007/978-3-030-18601-2_6.

Haile, Z. M., Sonego, P., Engelen, K., Vrhovsek, U., Tudzynski, P., Baraldi, E., et al. (2016). Characterizing the interaction between *Botrytis cinerea* and grapevine inflorescences. *Acta Hortic.*, 29–36. doi:10.17660/ActaHortic.2016.1144.4.

Ilc, T., Arista, G., Tavares, R., Navrot, N., Duchêne, E., Velt, A., et al. (2018). Annotation, classification, genomic organization and expression of the Vitis vinifera CYPome. *PLoS One* 13, 1–22. doi:10.1371/journal.pone.0199902.

Ilc, T., Parage, C., Boachon, B., Navrot, N., and Werck-Reichhart, D. (2016a). Monoterpenol Oxidative Metabolism: Role in Plant Adaptation and Potential Applications. *Front. Plant Sci.* 7, 1–16. doi:10.3389/fpls.2016.00509.

Ilc, T., Werck-Reichhart, D., and Navrot, N. (2016b). Meta-analysis of the core aroma components of grape

Stellenbosch University https://scholar.sun.ac.za

and wine aroma. *Front. Plant Sci.* 7, 1–15. doi:10.3389/fpls.2016.01472.
Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449, 463–7. doi:10.1038/nature06148.
Keller, M., Viret, O., and Cole, F. M. (2003). Botrytis cinerea Infection in Grape Flowers: Defense Reaction, Latency, and Disease Expression. *Phytopathology* 93, 316–322. doi:10.1094/PHYTO.2003.93.3.316.
Köllner, T. G., Gershenzon, J., and Degenhardt, J. (2009). Molecular and biochemical evolution of maize terpene synthase 10, an enzyme of indirect defense. *Phytochemistry* 70, 1139–1145. doi:10.1016/j.phytochem.2009.06.011.
Li, J.-X., Fang, X., Zhao, Q., Ruan, J.-X., Yang, C.-Q., Wang, L.-J., et al. (2013). Rational engineering of plasticity residues of sesquiterpene synthases from *Artemisia annua* : product specificity and catalytic efficiency. *Biochem. J.* 451, 417–426. doi:10.1042/BJ20130041.
Martin, D., Aubourg, S., Schouwey, M., Daviet, L., Schalk, M., Toub, O., et al. (2010). Functional Annotation, Genome Organization and Phylogeny of the Grapevine (Vitis vinifera) Terpene Synthase Gene Family Based on Genome Assembly, FLcDNA Cloning, and Enzyme Assays. *BMC Plant Biol.* 10, 226. doi:10.1186/1471-2229-10-226.
Martin, D. M., Toub, O., Chiang, A., Lo, B. C., Ohse, S., Lund, S. T., et al. (2009). The bouquet of grapevine (Vitis vinifera L. cv. Cabernet Sauvignon) flowers arises from the biosynthesis of sesquiterpene volatiles in pollen grains. *Proc. Natl. Acad. Sci. U. S. A.* 106, 7245–7250. doi:10.1073/pnas.0901387106.
Matarese, F., Cuzzola, A., Scalabrelli, G., and D'Onofrio, C. (2014). Expression of terpene synthase genes associated with the formation of volatiles in different organs of Vitis vinifera. *Phytochemistry*. doi:10.1016/j.phytochem.2014.06.007.
Mehari, Z. H., Pilati, S., Sonego, P., Malacarne, G., Vrhovsek, U., Engelen, K., et al. (2017). Molecular analysis of the early interaction between the grapevine flower and Botrytis cinerea reveals that prompt activation of specific host pathways leads to fungus quiescence. *Plant. Cell Environ.*, 1409–1428. doi:10.1111/pce.12937.
Minio, A., Lin, J., Gaut, B. S., and Cantu, D. (2017). How Single Molecule Real-Time Sequencing and Haplotype Phasing Have Enabled Reference-Grade Diploid Genome Assembly of Wine Grapes. *Front. Plant Sci.* 8, 1–6. doi:10.3389/fpls.2017.00826.
Minio, A., Massonnet, M., Figueroa-Balderas, R., Castro, A., and Cantu, D. (2019a). Diploid Genome Assembly of the Wine Grape Carménère. *G3: Genes|Genomes|Genetics* 9, g3.400030.2019. doi:10.1534/g3.119.400030.
Minio, A., Massonnet, M., Figueroa-Balderas, R., Vondras, A. M., Blanco-Ulate, B., and Cantu, D. (2019b). Iso-Seq Allows Genome-Independent Transcriptome Profiling of Grape Berry Development. *G3 Genes, Genomes, Genet.* 9, g3.201008.2018. doi:10.1534/g3.118.201008.
Roach, M. J., Johnson, D. L., Bohlmann, J., van Vuuren, H. J. J., Jones, S. J. M., Pretorius, I. S., et al. (2018). Population sequencing reveals clonal diversity and ancestral inbreeding in the grapevine cultivar Chardonnay. *PLoS Genet.*, 1–24. doi:https://doi.org/10.1371/journal.pgen.1007807.
Salvagnin, U., Malnoy, M., Thöming, G., Tasin, M., Carlin, S., Martens, S., et al. (2018). Adjusting the scent ratio: using genetically modified Vitis vinifera plants to manipulate European grapevine moth behaviour. *Plant Biotechnol. J.* 16, 264–271. doi:10.1111/pbi.12767.
Smit, S. J., Vivier, M. A., and Young, P. R. (2019). Linking Terpene Synthases to Sesquiterpene Metabolism in Grapevine Flowers. *Front. Plant Sci.* 10, 1–18. doi:10.3389/fpls.2019.00177.
Tasin, M., Anfora, G., Ioriatti, C., Carlin, S., De Cristofaro, A., Schmidt, S., et al. (2005). Antennal and behavioral responses of grapevine moth Lobesia botrana females to volatiles from grapevine. *J. Chem. Ecol.* 31, 77–87. doi:10.1007/s10886-005-0975-3.
Tasin, M., Bäckman, A.-C., Coracini, M., Casado, D., Ioriatti, C., and Witzgall, P. (2007). Synergism and redundancy in a plant volatile blend attracting grapevine moth females. *Phytochemistry* 68, 203–9. doi:10.1016/j.phytochem.2006.10.015.
Tasin, M., Bäckman, A. C., Anfora, G., Carlin, S., Ioriatti, C., and Witzgall, P. (2009). Attraction of female grapevine moth to common and specific olfactory cues from 2 host plants. *Chem. Senses* 35, 57–64. doi:10.1093/chemse/bjp082.
Venturini, L., Ferrarini, A., Zenoni, S., Tornielli, G. B., Fasoli, M., Dal Santo, S., et al. (2013). De novo transcriptome characterization of Vitis vinifera cv. Corvina unveils varietal diversity. *BMC Genomics* 14, 41. doi:10.1186/1471-2164-14-41.
Vondras, A. M., Minio, A., Blanco-Ulate, B., Figueroa-Balderas, R., Penn, M. A., Zhou, Y., et al. (2019). The genomic diversification of grapevine clones. *BMC Genomics* 20, 972. doi:10.1186/s12864-019-6211-2.
Wymore, T., Chen, B. Y., Nicholas, H. B., Ropelewski, A. J., and Brooks, C. L. (2011). A mechanism for evolving novel plant sesquiterpene synthase function. *Mol. Inform.* 30, 896–906. doi:10.1002/minf.201100087.
Zhou, Y., Massonnet, M., Sanjak, J. S., Cantu, D., and Gaut, B. S. (2017). Evolutionary genomics of grape (Vitis vinifera ssp. vinifera) domestication. *Proc. Natl. Acad. Sci.* 114, 201709257. doi:10.1073/pnas.1709257114.

Stellenbosch University https://scholar.sun.ac.za

Zhou, Y., Muyle, A., and Gaut, B. S. (2019). “Evolutionary Genomics and the Domestication of Grapes,” in *The Grape Genome*, eds. D. Cantu and M. A. Walker (Cham: Springer International Publishing), 39–55. doi:10.1007/978-3-030-18601-2_3.

Stellenbosch University https://scholar.sun.ac.za

Appendix A

Stellenbosch University https://scholar.sun.ac.za

Supplementary Data Sheet 1

```
# The shell script below is an example shows how Exonorate est2genome
mapping was performed
# The script was adapted from the wiki posted by alvaralmstedt on github at
https://github.com/alvaralmstedt/Tutorials/wiki/From-exonerate-to-igv:-A-
story-about-GFF

##########################################
# Step 1: Perform the exonerate analyses
##########################################

# to execute the shell script on a Mac, open your terminal and run:
# sh /path_to/example_script.sh

# you need to specify paths for the following
# --query
# --target
# output

echo "Starting exonerate e2g at:"
date
wait

# chunk 1
exonerate -m est2genome --percent 90 --maxintron 3000 --showtargetgff yes -
-showalignment no --ryo ">%qi length=%ql alnlen=%qal\n>%ti length=%tl
alnlen=%tal\n" --targetchunkid 1 --targetchunktotal 8 --query
path_to/query.fasta --target path_to/target.fasta >
path_to/output_e2g_chunk1.output &

# chunk n
# repeat this for the number of chunks (8 in the example above) you
specified and change the output chunk number to match the targetchunkid
number

wait
echo "All done with e2g at:"
date

##########################################
# Step 2: Concatenate output files
##########################################

cat * > output_e2g.gff
# This will concatenate all output chunks in the directory to one GFF file.
This file can be imported as is into CLC Main Workbench 7
```

Stellenbosch University https://scholar.sun.ac.za

Supplementary Data Sheet 2

```
# STEP 1: create your databases

mmseqs createdb <path/to/file> <DB name>
# specify the path to the fasta file you want to make a database of
# replace queryDB with a name for the database
# note that the DB will be created in the current working directory

# STEP 2: perform the clustering

# mmseqs cluster <DB> <DB_clu> tmp
# DB = database file
# DB_clu =  the output file
# tmp is the the tmp file where all the behind the scene stuff is dumped
# the following parameters were used in this study

mmseqs cluster <DB> <DB_clu> tmp --cov-mode 0 -c 0.85 -e 0.00001 --min-seq-
id 0.75 --cluster-mode 2

# To extract the clustering output as a text file
mmseqs createtsv <i:queryDB> [<i:targetDB>] <i:resultDB> <o:tsvFile>
[options]

# To extract the representative sequence for each cluster
mmseqs result2repseq <i:sequenceDB> <i:resultDB> <o:sequenceDb> [options]
mmseqs result2flat <i:queryDB> <i:targetDB> <i:resultDB> <o:fastaDB>
[options]
```

Stellenbosch University https://scholar.sun.ac.za

Appendix B

Stellenbosch University https://scholar.sun.ac.za

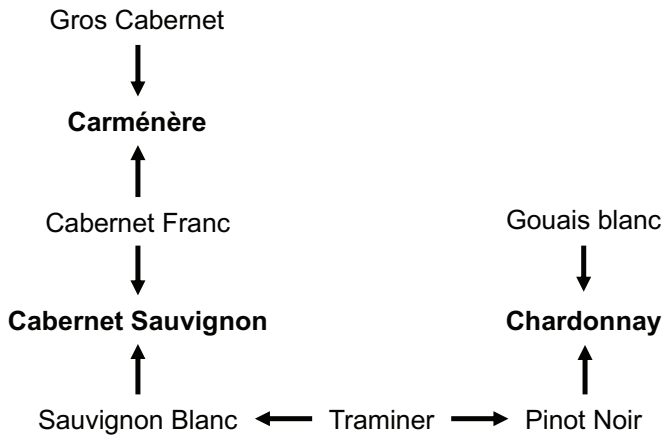

Figure 1. Relatedness (pedigree) of cultivars used for genomes discussed in this study.

Stellenbosch University https://scholar.sun.ac.za

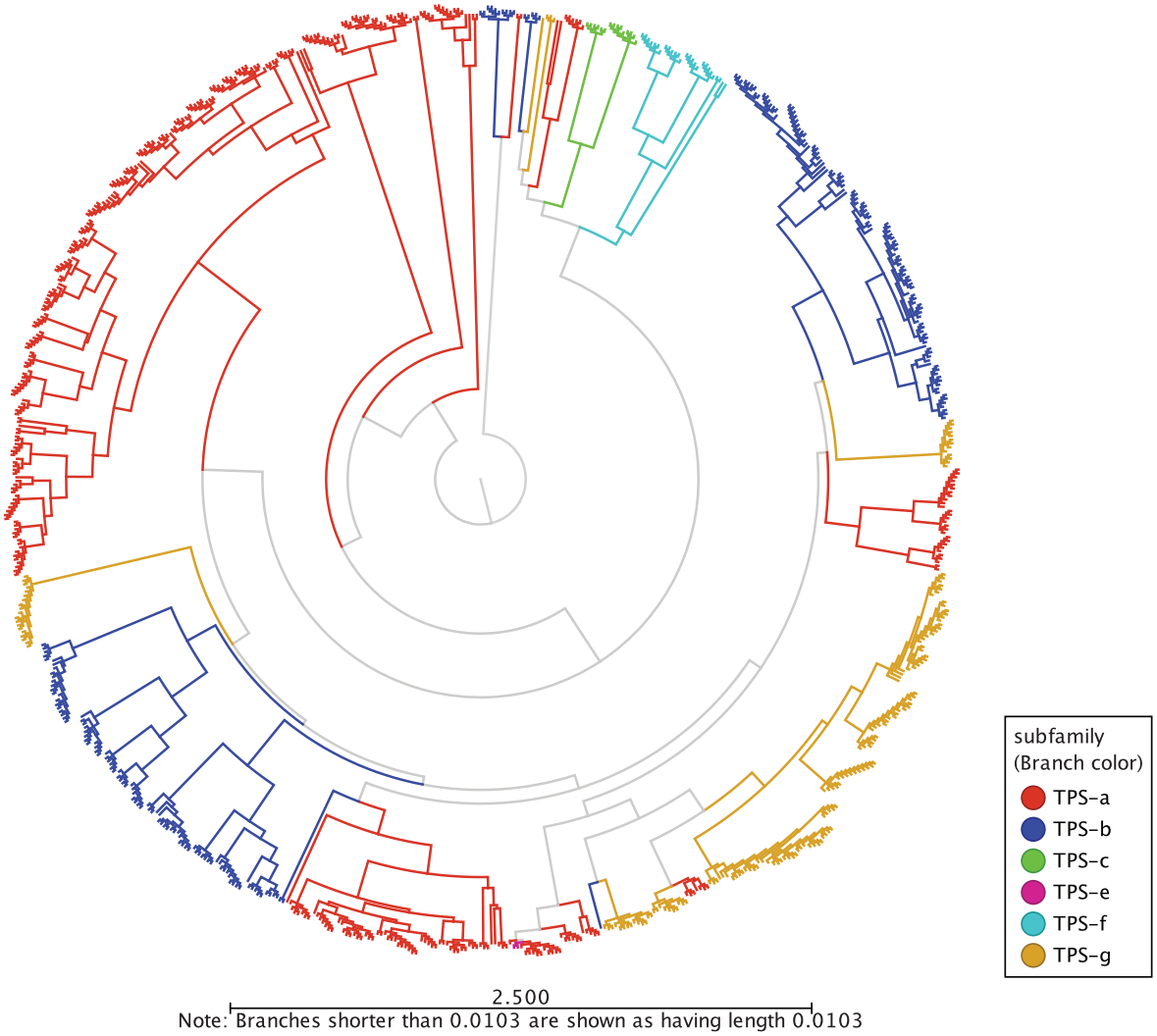

Figure 2. Phylogenetic tree of gDNA sequences. Subfamily groupings were inferred by the positions of PN40024 VviTPS gDNA sequences

Stellenbosch University https://scholar.sun.ac.za

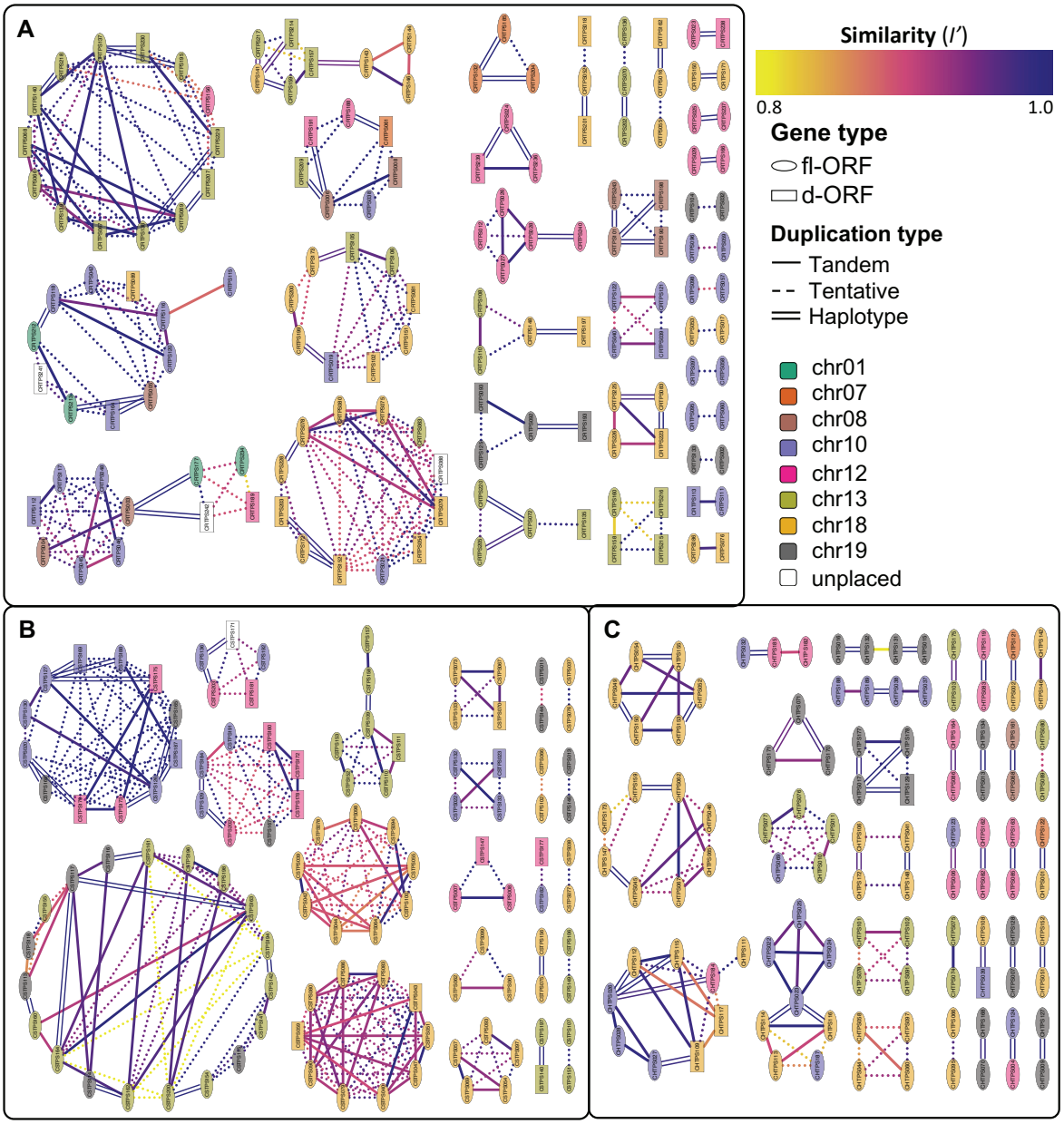

Figure 3. Gene duplication networks with nodes coloured according to chromosome.

Stellenbosch University https://scholar.sun.ac.za

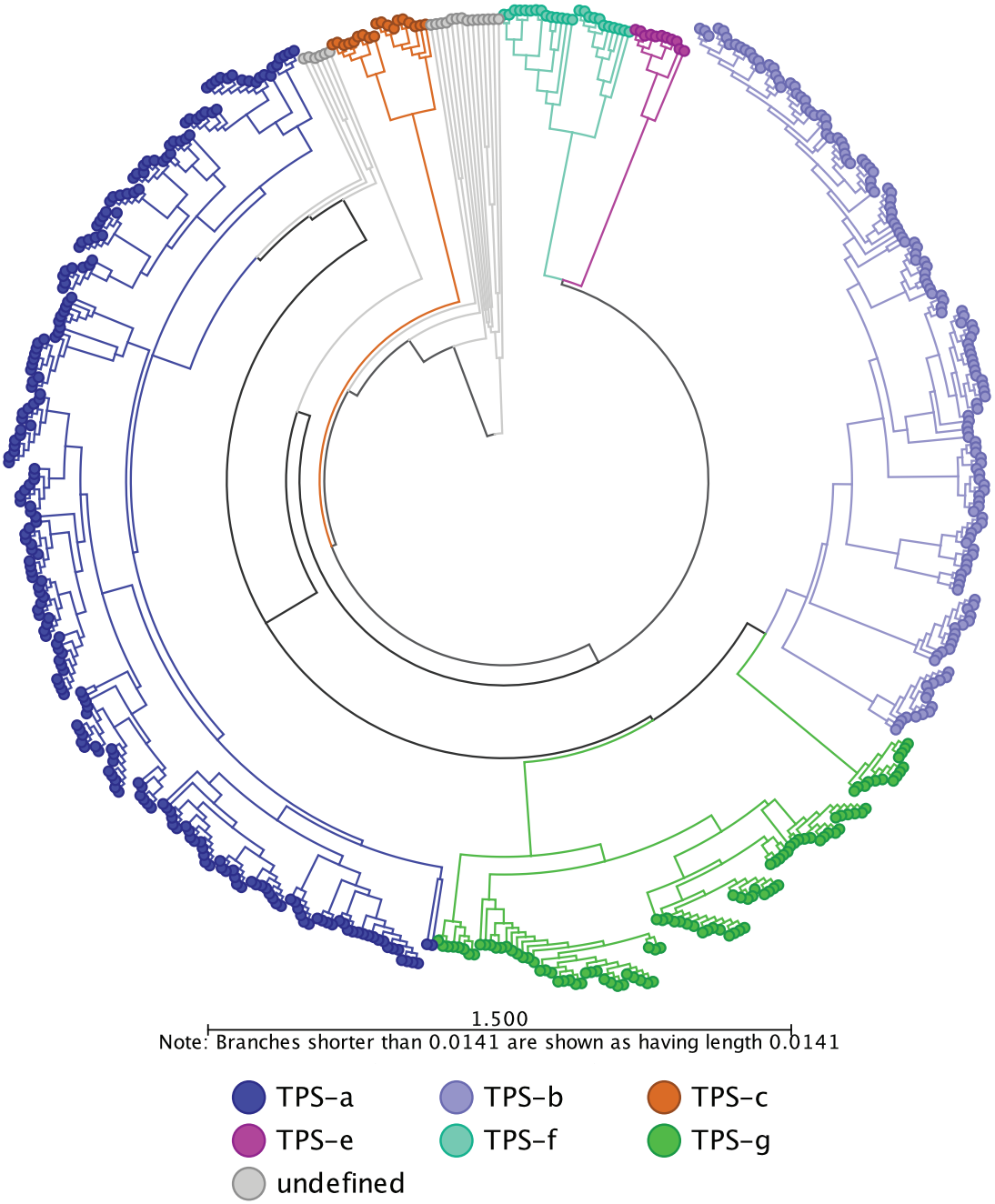

Figure 4. Phylogeny of all putative proteins with subfamilies inferred by phylogenetic position relative to functional VviTPS enzymes.

Stellenbosch University https://scholar.sun.ac.za

Table 1. Protein sequences used to infer enzyme mechanisms

Gene ID	Accession	Martin Gene Model	Reference
VvGwECar1	HM807373	VviTPS01	(Martin et al., 2010)
VvGwECar2	HM807374	VviTPS27	(Martin et al., 2010)
VvGwECar3	HM807375	VviTPS02	(Martin et al., 2010)
VvPNECar1	HM807402	VviTPS02	(Martin et al., 2010)
VvPNECar2	HM807403	VviTPS13	(Martin et al., 2010)
VvGwGerA	HQ326230	VviTPS01	(Martin et al., 2010)
VvGwaBer	HM807376	VviTPS10	(Martin et al., 2010)
VvGwGerD	HM807377	VviTPS07	(Martin et al., 2010)
VvPNGerD	HM807378	VviTPS15	(Martin et al., 2010)
VvCSaFar	HM807379	VviTPS20	(Martin et al., 2010)
VvGwgCad	HM807380	VviTPS08	(Martin et al., 2010)
VvPNbCur	HM807381	VviTPS30	(Martin et al., 2010)
VvPNSesq	HM807404	VviTPS12	(Martin et al., 2010)
VvPNaZin	HM807405	VviTPS14	(Martin et al., 2010)
VvPNSeInt	HM807406	VviTPS24	(Martin et al., 2010)
VvPNCuCad	HM807407	VviTPS26	(Martin et al., 2010)
VvPNaHum	HM807408	VviTPS11	(Martin et al., 2010)
VvPNEb2epi Car	HM807409	VviTPS21	(Martin et al., 2010)
VvGuaS	HM807406	N/A	(Drew et al., 2015)
VvVal	AY561843	N/A	(Lücker et al., 2004)
VvGerD	AY561842	N/A	(Lücker et al., 2004)
VvValCS	FJ696653	N/A	(Martin et al., 2009)
VvGwaPhe	HM807382	VviTPS45	(Martin et al., 2010)
VvPNaPin1	HM807383	VviTPS44	(Martin et al., 2010)
VvPNaPin2	HM807384	VviTPS44	(Martin et al., 2010)
VvGwbOci	HM807385	VviTPS34	(Martin et al., 2010)
VvCSbOci	HM807386	VviTPS35	(Martin et al., 2010)
VvCSbOciM	HM807387	VviTPS38	(Martin et al., 2010)
VvGwbOciF	HM807388	VviTPS47	(Martin et al., 2010)
VvCSbOciF	HM807389	VviTPS47	(Martin et al., 2010)
VvPNRLin	HM807390	VviTPS31	(Martin et al., 2010)
VvPNLinNer1	HM807391	VviTPS54	(Martin et al., 2010)
VvPNLinNer2	HM807392	VviTPS56	(Martin et al., 2010)
VvCSLinNer	HM807393	VviTPS56	(Martin et al., 2010)
VvPNLNGl1	HM807394	VviTPS57	(Martin et al., 2010)
VvPNLNGl2	HM807395	VviTPS63	(Martin et al., 2010)
VvPNLNGl3	HM807396	VviTPS58	(Martin et al., 2010)
VvPNLNGl4	HM807397	VviTPS61	(Martin et al., 2010)
VvGwGer	HM807398	VviTPS52	(Martin et al., 2010)
VvCSGer	HQ326231	VviTPS52	(Martin et al., 2010)
VvPNGer	HM807399	VviTPS52	(Martin et al., 2010)
VvCSENerGl	HM807400	N/A	(Martin et al., 2010)
VvPNENerGl	HM807401	N/A	(Martin et al., 2010)
VviMATPS10	QBL52481.1		(Smit et al., 2019)

Stellenbosch University https://scholar.sun.ac.za

Table 2. Average RaGOO scores per chromosome for the three cultivars assembled to PN40024 12x.v2

Chr.	Cv.	Location Score	Orientation Score	Grouping Score
chr01	CS	9%	98%	87%
	CR	15%	97%	53%
chr07	CS	9%	94%	74%
	CR	47%	95%	82%
	CH	9%	99%	93%
chr08	CS	22%	99%	93%
	CR	45%	92%	82%
	CH	4%	99%	97%
chr10	CS	5%	70%	65%
	CR	21%	66%	76%
	CH	3%	59%	74%
chr12	CS	25%	89%	57%
	CR	30%	92%	67%
	CH	4%	88%	68%
chr13	CS	19%	87%	83%
	CR	36%	91%	70%
	CH	16%	86%	74%
chr18	CS	5%	97%	79%
	CR	12%	86%	64%
	CH	14%	91%	76%
chr19	CS	12%	85%	64%
	CR	21%	97%	75%
	CH	11%	91%	77%
AVG		17%	89%	75%